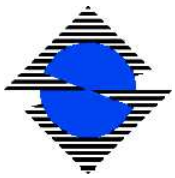


Nº D'ORDRE :

UNIVERSITÉ PARIS XI
UFR SCIENTIFIQUE D'ORSAY

THÈSE

présentée pour obtenir le

GRADE de DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS XI ORSAY

par

Yann OLLIVIER

Sujet :
Probabilités sur les espaces de configuration d'origine géométrique

Soutenue le 18 décembre 2003 devant la Commission d'examen

M. Thomas DELZANT
M. Étienne GHYS
M. Mikhael GROMOV, directeur
M. Richard KENYON
M. Pierre PANSU, co-directeur
M. Wendelin WERNER

Table sommaire

* * *

* * *

Au lecteur

LES TRAVAUX présentés dans ce volume sont fort divers, tant pour ce qui est des matières abordées que, disons-le sans ambages, du point de vue de l'intérêt des résultats qu'ils contiennent. Tel est long de quatre-vingt-dix pages, tel de cinq. Tel offre une preuve achevée et polie, tel se borne à indiquer comment la démonstration doit être construite ; tel enfin, n'ébauche que des pistes de recherches n'ayant pas abouti.

Par souci de rationalisation, on a choisi de placer en premier lieu les textes qui, semble-t-il, méritent le plus l'attention du lecteur. Par souci d'exhaustivité, on a fait figurer aussi les réflexions qui, reflets des débuts, n'ont pas été si fécondes qu'on l'aurait souhaité.

La variété des sujets ne doit néanmoins pas masquer l'idée directrice, révélée par l'intitulé suggéré par M. Gromov pour cette thèse : « Probabilités sur les espaces de configuration d'origine géométrique ». Tout du long, il s'agira donc d'espaces géométriques classiques abordés sous un angle mesuré. Deux ontologies s'opposent alors : ou bien, les parties de cet espace sont considérées comme plus ou moins existantes suivant le poids qui leur sera concédé ; sinon, la probabilité n'est qu'un moyen de naviguer dans un espace fixe mais inconnu. Également riches, ces deux approches font bon ménage dans les textes qui suivent, et l'on n'est pas tenu de choisir.

Enfin, le lecteur magnanime voudra bien pardonner les nombreuses fautes qui entachent inévitablement un tel recueil, et dont chaque relecture apporte son lot nouveau.

Paris, juillet 2003

Présentation

Présentation

Le point commun aux travaux recueillis ici est l'interaction entre géométrie et probabilités. Cette interaction se décline principalement en deux variantes. Dans l'une, on munit un espace géométrique d'une mesure de probabilité « naturelle » en rapport avec sa géométrie, et on se demande ensuite quelles sont les propriétés géométriques « typiques » dans cet espace, c'est-à-dire celles qui seront réalisées avec une probabilité très proche de 1. Dans l'autre, on utilise les probabilités pour accéder à certains éléments de l'espace, soit « typiques » (par exemple en effectuant une marche aléatoire dans l'espace), soit au contraire répondant à des propriétés très particulières (on montre un théorème d'existence en prouvant qu'une certaine construction aléatoire fournit l'objet voulu avec probabilité strictement positive).

Sharp phase transition theorems for hyperbolicity of random groups, *Growth and cogrowth of generic groups*, *Concentrated spaces seen as product spaces*, *Concentration spectrale dans les graphes* relèvent du premier point de vue ; *Vitesse de convergence des opérateurs de croisement*, *Un algorithme génétique dans l'espace des arbres*, *La démographie du PRA*, du second.

Les résultats principaux de cette thèse sont les trois théorèmes sur les groupes aléatoires présentés dans *Sharp phase transition theorems for hyperbolicity of random groups*. Nous donnons ci-dessous une introduction, destinée aux non-spécialistes, à la théorie des groupes aléatoires.

Growth and cogrowth of generic groups résout une question qui se posait naturellement au vu des théorèmes du texte précédent. Elle concerne la croissance et la cocroissance des groupes aléatoires et étend un théorème de Champetier [Ch], et peut simplifier certains aspects de la démonstration du texte [Gro4] de Gromov. Certaines parties de ce texte ne sont encore qu'ébauchées.

On a small cancellation theorem of Gromov, le seul texte entièrement déterministe de ce recueil, donne une démonstration élémentaire d'un théorème annoncé par Gromov dans [Gro4], qui étend la théorie ordinaire de la petite simplification et est à la base de la construction par Gromov de groupes dont le graphe de Cayley contient une famille d'expanseurs. Ce texte ne fait donc que donner une démonstration combinatoire d'un résultat de [Gro4].

Le court *Concentrated spaces seen as product spaces* est une étude en cours sur le lien entre espaces concentrés et espaces produits (voir plus bas une introduction à cette problématique). Il est bien connu que les espaces produits sont

concentrés. Nous obtenons un début de réciproque : sur un espace concentré, on peut trouver une ébauche de structure produit. Ce résultat n'est que le commencement d'une recherche et peut très probablement être amélioré.

Concentration spectrale dans les graphes est un texte plus ancien. Il applique des méthodes connues en concentration de la mesure à un cas apparemment non traité dans la littérature mais ne présentant pas de difficulté particulière.

Les textes de la section « Autour des algorithmes génétiques » sont nettement plus anciens. *Vitesse de convergence des opérateurs de croisement* se rapproche de la biologie des populations en étudiant la vitesse du brassage des gènes introduit par la reproduction sexuée, et résout une question posée dans [RRS]. *Un algorithme génétique dans l'espace des arbres* décrit une tentative pour résoudre le problème de la reconstitution phylogénétique à l'aide d'un algorithme génétique explorant l'espace des arbres sur un ensemble de feuilles fixé, faisant intervenir un croisement d'arbres. Enfin, *La démographie du PRA* contient quelques résultats simples sur un procédé utilisé en théorie algorithmique des groupes, le « PRA », résultats inspirés par la ressemblance entre ce procédé et un algorithme génétique.

* * *

1 Groupes hyperboliques et groupes aléatoires

1.1 L'intérêt des groupes aléatoires

La théorie (si l'on peut déjà lui accorder ce nom) des groupes aléatoires s'intéresse aux propriétés d'un groupe « typique ». Il faut bien sûr commencer par donner un sens précis à cette expression.

Notons qu'outre son intérêt propre consistant à décrire les propriétés les plus fréquentes des groupes, cette théorie a déjà servi à construire un groupe aux propriétés inhabituelles répondant à une question ouverte (cf. [Gro4]). Pour un exposé général, on pourra consulter le texte du récent séminaire Bourbaki consacré à ces questions, par Étienne Ghys, [Gh].

Nous nous intéresserons ici aux groupes discrets engendrés par un nombre fini m de générateurs (et leurs inverses), disons $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Tout tel groupe peut être vu comme un quotient du groupe libre F_m à m générateurs par un ensemble de relations R. Se donner un groupe au hasard, c'est simplement se donner l'ensemble R définissant le groupe au hasard.

On est donc ramené à se donner au hasard des mots en les $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Une longueur de mots ℓ étant choisie, le plus simple consiste ensuite à choisir les mots uniformément parmi tous les mots de longueur ℓ en les générateurs. En fait, on peut toujours supposer que les relations apparaissant dans R sont des mots *réduits* (c'est-à-dire ne contenant pas de séquence $a_i a_i^{-1}$ ou $a_i^{-1} a_i$), et il est donc plus naturel de choisir nos mots aléatoires uniformément parmi l'ensemble des $(2m)(2m-1)^{\ell-1}$ mots réduits de longueur ℓ.

Ces choix étant faits, le modèle de groupe aléatoire ne dépend plus que du nombre de mots que l'on prend : plus l'ensemble de relateurs est grand, plus le groupe sera petit. Le modèle dit *à densité*, introduit par M. Gromov (cf. [Gro2]), s'est révélé extrêmement fécond et semble être la bonne manière d'évaluer la « taille » de l'ensemble R.

MODÈLE À DENSITÉ – *Choisir un nombre d entre 0 et 1. Se donner une longueur de mots ℓ très grande. Poser $N = (2m-1)^{d\ell}$. Pour l'ensemble de relations R, tirer N fois de suite (indépendamment) un mot réduit au hasard uniformément parmi les $(2m)(2m-1)^{\ell-1}$ mots réduits possibles.*

Si l'on veut pouvoir appliquer des théorèmes de probabilités, comme des lois des grands nombres, il est nécessaire d'avoir un paramètre tendant vers l'infini. C'est la raison pour laquelle on prend ℓ grand.

Comme $(2m)(2m-1)^{\ell-1}$ est presque égal à $(2m-1)^{\ell}$, on voit que ce modèle consiste à prendre un nombre N de mots égal au nombre total de mots, à la puissance d. L'exposant $d\ell$ est à interpréter comme une dimension, à savoir la dimension de l'ensemble R qu'on va tirer. En effet, la dimension d'un ensemble peut être vue comme le nombre d'équations qu'on peut s'imposer, de manière à ce que, génériquement, il existe un élément de cet ensemble vérifiant ces équations. C'est précisément ce qui se passe ici, pour la bonne notion d'« équation ». Pour des mots en certaines lettres, une équation sera de la forme « imposer la k-ième lettre du mot ». Si on se donne L équations imposant les L premières lettres d'un mot réduit, la probabilité qu'un mot réduit choisi au hasard les satisfasse est $1/(2m)(2m-1)^{L-1}$ soit environ $1/(2m-1)^{L}$. Pour un ensemble de N mots choisis au hasard, la probabilité qu'un mot au moins satisfasse les équations imposées sera donc non négligeable si N est de l'ordre de $(2m-1)^{L}$. On voit donc que pour $N = (2m-1)^{d\ell}$, on peut imposer $d\ell$ « équations » à un mot de l'ensemble.

L'intérêt du modèle à densité est justifié par le théorème suivant, dû à M. Gromov (cf. [Gro2]) :

THÉORÈME 1 : TRANSITION DE PHASE POUR LES GROUPES ALÉATOIRES – *Soit R un ensemble de relations aléatoires tirées selon le modèle à densité, et soit $G = F_m/\langle R\rangle$ le groupe aléatoire ainsi défini.*

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 $\ell \to \infty$.

Selon la densité de l'ensemble de relations, on a donc une transition de phase extrêmement précise entre des groupes infinis (dont on connaît aussi d'autres caractéristiques) et des groupes triviaux. Ce qui se passe à la densité critique $1/2$ est pour le moment totalement inconnu.

L'occurrence possible de $\mathbb{Z}/2\mathbb{Z}$ ne doit pas surprendre : si ℓ est pair, on ne met que des relations de longueur paire et donc le quotient est au moins $\mathbb{Z}/2\mathbb{Z}$.

Avant de rappeler ce qu'est un groupe hyperbolique, donnons une esquisse de preuve de la partie triviale du théorème.

Prendre $d > 1/2$ revient à prendre pour R un nombre de mots supérieur à la racine carrée du nombre total de mots possibles. Il est élémentaire et bien connu (lemme des anniversaires ou principe des tiroirs probabiliste) que si l'on tire N objets parmi moins de N^2 objets (pour N grand), avec grande probabilité on tire deux fois le même objet. Ceci signifie que dans R on rencontre deux fois le même relateur. Cela peut aussi s'interpréter en termes de dimension comme ci-dessus : la dimension de R est $d\ell$, la dimension des couples d'éléments de R est $2d\ell$, et imposer l'égalité de deux mots de longueur ℓ revient à poser ℓ équations ; « donc », si $2d\ell > \ell$, on a une chance que ces équations soient satisfaites par un couple.

Par un raisonnement analogue en n'imposant que $\ell - 1$ équations, on obtient aussi que, dans R, se rencontrent probablement deux mots $r_1 = xa_i$ et $r_2 = xa_j$ où x est un mot de longueur $\ell - 1$ et où a_i et a_j sont deux générateurs. Dans le groupe quotient $G = F_m/\langle R \rangle$, cela signifie que $xa_i = e$ et $xa_j = e$. Ceci implique $a_i = a_j$: les deux générateurs sont devenus égaux. Comme $d > 1/2$ cette situation se produit même une infinité de fois (pour ℓ grand), avec i et j tirés au hasard ; et donc, dans le groupe G, tous les couples de générateurs ainsi que leurs inverses sont égaux... il est alors facile de voir que G est soit $\{e\}$ soit $\mathbb{Z}/2\mathbb{Z}$.

1.2 Groupes hyperboliques

Il peut être utile de rappeler ce qu'est un groupe hyperbolique.

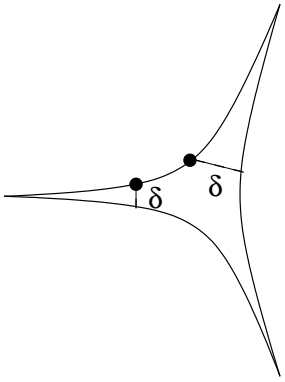
Soit G un groupe engendré par les éléments $a_1, \ldots, a_m$. Le *graphe de Cayley* de G (pour ce système générateur) est le graphe dont les sommets sont tous les éléments de G, et dont les arêtes correspondent à la multiplication à droite par l'un des générateurs $a_1, \ldots, a_m$. Par exemple, le graphe de Cayley de $\mathbb{Z}/n\mathbb{Z}$ est un cycle à n arêtes. Puisque les générateurs engendrent le groupe, ce graphe est connexe.

Le graphe de Cayley est naturellement un espace métrique : il suffit de décréter que chaque arête est de longueur 1.

Une *géodésique* entre deux points dans le graphe de Cayley est un chemin de longueur minimale (un tel chemin n'est pas nécessairement unique). Un *triangle* est la donnée de trois points du graphe de Cayley, ainsi que de trois géodésiques reliant ces points deux à deux qu'on appellera *côtés* du triangle.

DÉFINITION : 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 : GROUPES HYPERBOLIQUES – *Un groupe est dit* hyperbolique *s'il existe un nombre $\delta \geqslant 0$ tel que tous les triangles du graphe de Cayley sont δ-fins.*

C'est un théorème non trivial que l'hyperbolicité d'un groupe ne dépend pas du système générateur choisi (dont dépend le graphe de Cayley).

Cette définition n'est pas spécifique aux groupes : elle a un sens dans tout espace métrique où des géodésiques existent. La terminologie est justifiée par les deux faits suivants : le plan hyperbolique standard est hyperbolique (!!) ; et le groupe fondamental d'une variété hyperbolique compacte est hyperbolique.

Par exemple, dans un arbre tous les triangles sont 0-fins ; donc, les groupes libres, dont le graphe de Cayley (pour le système de générateurs standard) est un arbre, sont des groupes hyperboliques. Les groupes hyperboliques sont les groupes dont le graphe de Cayley, « vu de loin », ressemble à un arbre (en un sens très précis) ; ce sont donc des groupes qui, « vus de loin », ressemblent à des groupes libres.

Les groupes hyperboliques ont été introduits par M. Gromov dans [Gro1] et leur intérêt ne s'est pas démenti depuis. Pour plus de renseignements on pourra consulter l'excellent [GH].

Donnons une caractérisation très utile des groupes hyperboliques. Lorsqu'un groupe G engendré par $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$ est défini par un ensemble de relateurs R, tout mot w en les générateurs représentant l'élément neutre de $G = F_m/\langle R\rangle$ est un élément de $\langle R\rangle$, c'est-à-dire qu'il s'écrit comme un produit de conjugués d'éléments de R ou de leurs inverses :

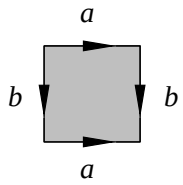
$$w = e \text{ dans } G \Leftrightarrow w = \prod u_i r_i^{\pm 1} u_i^{-1},\ r_i \in R$$

l'égalité ayant lieu dans le groupe libre (c'est-à-dire modulo les simplifications $a_i a_i^{-1}$).

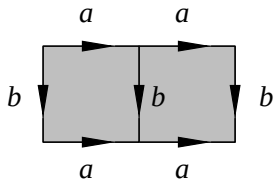
Une question naturelle qui se pose alors est : en présence d'un mot dont on sait qu'il représente l'élément neutre, combien de relateurs r_i comporte, au minimum, une telle décomposition ? Un des premiers faits de la théorie des groupes hyperboliques est le suivant.

PROPOSITION – *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.

Cet outil est à la base des théorèmes d'hyperbolicité des groupes aléatoires. Nous sommes désormais en mesure de donner l'idée de la démonstration de la partie non triviale du Théorème 1.

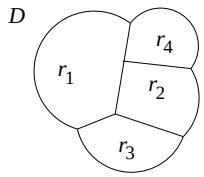
L'hyperbolicité d'un groupe dit que tout mot w représentant l'élément neutre peut être écrit comme un produit d'au plus $C.L$ relateurs où L est la longueur de w. Cela signifie qu'il existe un diagramme de van Kampen D le long du bord

duquel on lit w, et vérifiant l'inégalité isopérimétrique suivante entre son bord et son nombre de faces :

$$|\partial D| \geqslant |D| / C$$

Pour prouver l'hyperbolicité d'un groupe, il suffit donc de prouver que ses diagrammes de van Kampen satisfont une telle inégalité isopérimétrique linéaire (comparer avec l'inégalité isopérimétrique bien connue dans le plan qui lie le *carré* de la longueur du bord d'une figure à sa surface).

Revenant aux groupes aléatoires, considérons donc une présentation $G = \langle a_1, \ldots, a_m \mid R \rangle$ où R est un ensemble aléatoire de relations obtenu selon le modèle à densité. Soit D un diagramme de van Kampen pour cette présentation.



Considérons deux relateurs r_i, r_j de ce diagramme recollés sur une longueur L_{ij}. Comme ces relateurs sont choisis au hasard, la probabilité qu'un tel recollement puisse avoir lieu est $1/(2m-1)^{L_{ij}}$. Si tous les relateurs présents dans le diagramme sont distincts, donc choisis indépendamment, la probabilité d'un tel diagramme est donc $1/(2m-1)^L$ où L est la longueur interne totale du diagramme. (Traiter les relateurs apparaissant plusieurs fois est plus délicat car cela fait perdre de l'indépendance.)

Maintenant, le nombre de choix pour les $|D|$ relateurs du diagramme parmi les $|R|$ relateurs de la présentation est $|R|^{|D|} = (2m-1)^{d\ell|D|}$ par définition du modèle à densité. La probabilité d'existence d'un tel diagramme de van Kampen est donc inférieure à $(2m-1)^{d\ell|D|-L}$ où L est la longueur interne totale.

Mais la longueur du bord du diagramme est $|\partial D| = \ell\,|D| - 2L$ (sommer les longueurs des faces et enlever deux fois les longueurs des recollements). Maintenant, pour que la probabilité d'existence du diagramme ne soit pas petite (plus précisément, pas exponentiellement décroissante en ℓ), on doit avoir $L \leqslant d\ell\,|D|\,(1+\varepsilon)$ d'après notre majoration de cette probabilité. Mais alors :

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

ce qui donne l'inégalité isopérimétrique souhaitée si $d < 1/2$ (en prenant par exemple $\varepsilon = (1-2d)/4$). Autrement dit : soit un diagramme de van Kampen vérifie l'inégalité isopérimétrique, soit la probabilité que des relateurs aléatoires le forment décroît comme $(2m-1)^{-\varepsilon\ell}$.

Ceci ne concerne qu'un seul diagramme possible, alors qu'il y en a une infinité. Mais un théorème profond de géométrie hyperbolique (théorème de Cartan-Hadamard-Gromov), qui affirme que l'hyperbolicité est un phénomène « semi-local », permet de ne tester l'inégalité isopérimétrique que sur un nombre fini de diagrammes. Le Théorème 1 est donc démontré.

1.3 Transitions de phase pour les quotients aléatoires

On va désormais s'intéresser à des généralisations du Théorème 1. 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. Ceci est d'autant plus vraisemblable que, au sens ci-dessus, un groupe hyperbolique ressemble à un groupe libre.

Ce résultat constitue le théorème principal de *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) tel qu'il existe un entier $n > 1$ avec $x^n = e$. Un groupe hyperbolique est dit *élémentaire* s'il est fini ou s'il contient un sous-groupe d'indice fini isomorphe à $\mathbb{Z}$ (l'analyse des quotients aléatoires de tels groupes élémentaires est facile).

THÉORÈME 2 : TRANSITION DE PHASE POUR LES QUOTIENTS ALÉATOIRES – *Soit G_0 un groupe hyperbolique sans torsion et non élémentaire, engendré par les éléments $a_1, \dots, a_m$. Soit $0 \leqslant d \leqslant 1$ une densité et soit R un ensemble de relations aléatoires tirées selon le modèle à densité. Soit $G = G_0/\langle R \rangle$ le quotient aléatoire. Il existe une densité critique d_0 (dépendant de G_0) ayant la propriété suivante.*

Si $d < d_0$, alors le quotient aléatoire G est infini et hyperbolique avec probabilité tendant vers 1 quand $\ell \to \infty$.

Si $d > d_0$, alors le quotient aléatoire G est $\{e\}$ ou $\mathbb{Z}/2\mathbb{Z}$ avec probabilité tendant vers 1 quand $\ell \to \infty$.

Sur l'hypothèse « sans torsion » : on connaît une hypothèse moins restrictive, légèrement plus compliquée, et qui est nécessaire et suffisante pour que le théorème soit vrai. On a aussi une idée relativement claire de ce qui se passe lorsque cette hypothèse n'est pas vérifiée.

De plus, on sait caractériser explicitement la densité critique d_0 en fonction de G_0 : on a $d_0 = 1 - \eta$ où η est la *cocroissance* du groupe G_0, notion introduite par R. Grigorchuk dans [Gri]. La cocroissance d'un groupe est l'exposant de croissance du nombre de mots réduits représentant l'élément neutre du groupe.

Plus précisément, soit L une longueur paire et soit W_L l'ensemble des mots réduits en les générateurs $a_1^{\pm 1}, \dots, a_m^{\pm 1}$ qui sont égaux à e dans le groupe G_0. La cocroissance de G_0 par rapport à ce système de générateurs est

$$\eta = \lim_{\substack{L \to \infty \\ L \text{ pair}}} \frac{1}{L} \log_{2m-1} \#W_L$$

On peut montrer que la limite existe. Cette définition n'a pas de sens pour un groupe libre (W_L est vide), mais une formule liant en général la cocroissance et le rayon spectral du laplacien sur le graphe de Cayley du groupe permet de conclure que $\eta(F_m) = 1/2$, ce qui est en accord avec le fait que le Théorème 2 généralise le Théorème 1.

Dans un groupe libre, la notion de mot réduit coïncide avec celle de *mot géodésique* (de longueur minimale parmi les mots représentant le même élément). Ce n'est pas le cas dans un groupe hyperbolique quelconque. Il est donc naturel de se demander comment se comportent des quotients aléatoires par des mots géodésiques plutôt que réduits. Il se trouve que dans ce cas, la densité critique est $1/2$ pour tout groupe, mais le théorème est légèrement plus compliqué à énoncer.

Il existe aussi une troisième variante de ce théorème dans laquelle on prend des mots quelconques plutôt que réduits ou géodésiques. Ces trois variantes sont trois cas particuliers d'un théorème plus général qui contrôle les quotients aléatoires par des éléments tirés selon une mesure sur le groupe vérifiant une liste de quatre axiomes naturels mais difficiles à exprimer simplement.

1.4 Autres résultats sur les groupes aléatoires, questions en suspens

Bien d'autres propriétés des groupes aléatoires sont connues. Par exemple, au vu du théorème ci-dessus, on peut naturellement se demander quel sort subit la cocroissance lors d'un quotient aléatoire. La réponse, fournie dans *Growth and cogrowth of generic groups*, est que la cocroissance d'un quotient aléatoire est arbitrairement proche (pour ℓ assez grand) de celle du groupe de départ. Appliqué en particulier aux quotients aléatoires d'un groupe libre, ceci donne que, génériquement, la cocroissance d'un groupe aléatoire est très proche de $1/2$. Outre son intérêt propre, ce résultat simplifie le problème des quotients aléatoires successifs utilisés dans [Gro4], où le contrôle de la cocroissance des quotients successifs est obtenu par la voie détournée de la propriété T.

Mentionnons quelques thèmes étudiés en rapport avec les groupes aléatoires. La petite simplification, les éléments de torsion, la topologie du bord, la propriété T, les sous-groupes libres, la planarité du graphe de Cayley... ont été étudiés. Donnons quelques noms, outre bien sûr Gromov : Arzhantseva, Champetier, Cherix, I. Kapovich, Ol'shanskiĭ, Schupp, Schpilrain, Żuk...

Ces travaux concernent l'étude des propriétés génériques des groupes pour elles-mêmes, mais elles ont aussi des applications non probabilistes. Ainsi du dénombrement *à isomorphisme près* des groupes à un relateur, obtenu par I. Kapovich, Schupp et Schpilrain à l'aide de considérations de généricité. Ou encore, de la construction par Gromov d'un groupe dont le graphe de Cayley contient une famille d'expanseurs, ce qui a des applications en théorie des opérateurs (conjecture de Baum-Connes). Les implications de la construction ne sont pas encore bien comprises.

Dans tous ces travaux, l'hyperbolicité est omniprésente.

Une autre approche sur les groupes génériques, topologique plutôt que probabiliste, a été développée par Champetier. Là encore l'hyperbolicité est fondamentale. Cette approche a des liens importants avec la théorie du premier ordre des groupes développée entre autres par Sela. On pourra consulter le récent

séminaire Bourbaki par Frédéric Paulin sur ces sujets ([Pau]).

Donnons enfin quelques pistes de recherche suggérées par l'étude des groupes aléatoires. Commençons par deux thèmes ne relevant pas de la philosophie des propriétés génériques, mais qui sont apparus dans ce cadre.

On rencontre fréquemment, dans cette théorie, des groupes ayant une propriété un peu plus forte que l'hyperbolicité. Un groupe hyperbolique est un groupe dans lequel, étant donné un mot représentant l'élément neutre, *il existe* un diagramme de van Kampen bordé par ce mot, et vérifiant une inégalité isopérimétrique linéaire. Or, dans les groupes hyperboliques rencontrés le plus souvent dans l'étude des groupes aléatoires, *tous* les diagrammes de van Kampen (réduits) satisfont une telle inégalité. Cette propriété est bien connue pour les groupes à petite simplification, mais les groupes aléatoires dans le modèle à densité ne sont pas, en général, à petite simplification (du moins pas dans le système générateur naturel).

Cette hyperbolicité forte semble liée à des propriétés topologiques, comme le fait d'avoir une dimension du bord à l'infini égale à 1. Les liens entre cette propriété, la petite simplification, la dimension du bord et la dimension du groupe lui-même ne sont pas clairs et méritent d'être étudiés. Ce problème, bien que suggéré par les groupes aléatoires, n'a rien de probabiliste.

Un autre type de groupe surgit naturellement de ces constructions aléatoires : les limites d'une infinité de quotients aléatoires successifs. Ces groupes ne sont pas hyperboliques (car ils sont de présentation infinie). Néanmoins, ils conservent une forme d'inégalité isopérimétrique. Si D est un diagramme de van Kampen (réduit), et si on note $|\partial f|$ la longueur d'une face $f \in D$, alors ces diagrammes vérifient l'inégalité isopérimétrique linéaire

$$|\partial D| \geqslant \alpha \sum_{f \in D} |\partial f|$$

pour une certaine constante $\alpha > 0$. Ceci donne un aspect « fractal » à ces groupes. Malheureusement, le comportement de cette propriété par changement de système générateur n'est pas clair. Cette hyperbolicité faible mériterait elle aussi d'être étudiée.

Par ailleurs, un argument heuristique semble indiquer que les groupes aléatoires peuvent difficilement avoir des quotients finis (non triviaux). Préciser cet argument permettrait de résoudre la question classique de savoir s'il existe des groupes hyperboliques sans quotients finis.

Mentionnons maintenant des problèmes appartenant en propre à la problématique des groupes aléatoires.

Le modèle à densité conserve une partie de son mystère dans la mesure où on ne sait pas prouver que les groupes obtenus à différentes densités sont essentiellement différents. Cela fournirait un grand nombre de groupes non isomorphes. On aimerait trouver des invariants du groupe obtenu, dépendant de la densité. Voici deux invariants candidats : le nombre minimal de générateurs

(il est facile de montrer qu'au voisinage de la densité critique il tombe à 2) ; la cohomologie L^p pour diverses valeurs de p, qui mesure en quelque sorte le taux de branchement d'un espace, qui devrait être lié à la densité.

Les propriétés cohomologiques des groupes aléatoires sont mal connues : on sait qu'ils sont de dimension cohomologique 2, et que par ailleurs, en densité supérieure à $1/3$ ils ont la propriété T (avec évaluation des constantes de Kazhdan). Mais on ne sait pas si, en densité inférieure à $1/3$, la propriété T apparaît ou non.

Enfin, il existe des modèles généralisant le modèle à densité. Ce dernier présente l'inconvénient que toutes les relations ajoutées doivent avoir la même longueur. Cet obstacle est partiellement levé dans *Sharp phase transition theorems for hyperbolicity of random groups*, où demeure tout de même la contrainte que les longueurs des relateurs doivent rester dans un rapport borné. Un modèle plus général très naturel mais encore très mal compris est le modèle dit *à température* (que nous n'expliciterons pas), qui fournit des groupes de présentation infinie, donc non hyperboliques, avec des relateurs de toutes les longueurs. Ce modèle a la propriété intéressante de donner des groupes sans quotient fini (à part $\{e\}$), à comparer avec le même problème évoqué plus haut concernant les groupes hyperboliques.

Comme on le voit, les problèmes ouverts en théorie des groupes aléatoires, et les retombées potentielles sur les groupes non aléatoires, ne manquent pas.

2 La concentration de la mesure

La concentration de la mesure est un phénomène mêlant géométrie et probabilités, qui apporte une explication conceptuelle et des généralisations puissantes à certains des théorèmes de probabilités les plus banals, comme la loi des grands nombres ou le théorème central limite. On peut consulter l'introduction donnée par Talagrand dans [Tal]. Le point de vue géométrique est fortement développé par Gromov dans [Gro3].

Historiquement, la première observation de la concentration de la mesure est la suivante : dans la sphère S^n de grande dimension, munie de la mesure riemannienne normalisée à 1, une petite bande de largeur environ $1/\sqrt{n}$ autour d'un équateur contient presque toute la mesure, comme un calcul direct le montre aisément. Autrement dit, un petit voisinage d'une demi-sphère contient déjà presque toute la mesure.

Cette observation permet de démontrer un résultat d'apparence étrange : une fonction lipschitzienne sur la sphère S^n de grande dimension est presque constante (à $1/\sqrt{n}$ près) !

En effet, soit f une fonction 1-lipschitzienne (pour la normalisation) de S^n vers $\mathbb{R}$. Soit m la médiane de f et soient S_+ et S_- les parties de la sphère où f est respectivement supérieure et inférieure à sa médiane. Par définition de la médiane, ces deux ensembles sont de mesure supérieure à $1/2$ (strictement si f

vaut m sur une partie de mesure non nulle, auquel cas on peut tronquer un peu S_+ et S_- pour obtenir des parties de mesure $1/2$).

Dans le plan, il est bien connu que la figure minimisant la longueur de son bord, à surface donnée, est le cercle. On peut remplacer « longueur du bord » par « aire d'un petit voisinage ». On démontre de même que dans la sphère, la mesure d'un ε-voisinage d'une partie A est toujours supérieure à la mesure du ε-voisinage d'une calotte de même mesure que A. Revenant en particulier à S_+ qui est de mesure $1/2$, on voit que la mesure d'un ε-voisinage de S_+ est supérieure à la mesure d'un ε-voisinage d'une demi-sphère. Or, comme on l'a dit plus haut, pour ε de l'ordre de $1/\sqrt{n}$ cette mesure est presque 1. Autrement dit, une majorité de points sont à distance au plus $1/\sqrt{n}$ de S_+. Par définition de S_+, et comme f est 1-lipschitzienne, si un point x est à distance r de S_+, alors $f(x) \geqslant m - r$. Donc, pour une majorité (en mesure) de points x, on a $f(x) \geqslant m - 1/\sqrt{n}$. En raisonnant symétriquement avec S_-, on obtient que pour une majorité de points, f est comprise entre $m - 1/\sqrt{n}$ et $m + 1/\sqrt{n}$. (On a omis une constante devant $1/\sqrt{n}$ pour simplifier.)

En résumé, on a démontré le théorème suivant, que l'on peut faire remonter à Paul Lévy dans les années 1920 :

THÉORÈME 3 : CONCENTRATION DE LA MESURE SUR LA SPHÈRE – *Une fonction 1-lipschitzienne sur la sphère S^n, pour n grand, est presque constante à $1/\sqrt{n}$ près.*

Donnons un autre exemple plus proche des probabilités classiques et (en apparence) plus éloigné de la géométrie. Il est bien connu que, si l'on fait une série de n tirages à pile ou face avec n grand et que l'on compte la proportion de « pile » qui sont apparus, les résultats sont proches d'une gaussienne centrée en $1/2$ et d'écart-type $1/2\sqrt{n}$.

Géométrisons ce résultat. On considère le cube discret $\{0,1\}^n$ muni de la mesure de probabilité uniforme qui donne un poids $1/2^n$ à chaque point. Ce cube modélise bien sûr le résultat d'une suite de n tirages à pile ou face, et la proportion de « pile » est une fonction sur ce cube, qui varie de $1/n$ sur chaque arête du cube. Le théorème central limite affirme que cette fonction est presque constante égale à $1/2$, avec des fluctuations gaussiennes de l'ordre de $1/\sqrt{n}$.

Or ce résultat est loin de n'être valable que pour cette fonction particulière. Talagrand a ainsi démontré :

THÉORÈME 4 : CONCENTRATION DE LA MESURE SUR LE CUBE – *Soit le cube discret $\{0,1\}^n$ muni de la mesure uniforme, et de la métrique qui attribue une longueur $1/n$ à chaque arête (de manière à ce que le diamètre du cube soit 1). Soit f une fonction du cube vers $\mathbb{R}$ qui soit 1-lipschitzienne pour cette métrique. Alors f est presque constante à $1/\sqrt{n}$ près, et les fluctuations sont contrôlées par des gaussiennes. Ceci au sens où il existe un nombre m tel que pour tout $t \geqslant 0$:*

$$\Pr(|f - m| \geqslant t) \leqslant 2e^{-nt^2/2}$$

Ainsi, une fonction de n variables indépendantes dans laquelle chaque variable influe d'au plus $1/n$, est constante à $1/\sqrt{n}$ près. Ceci généralise très fortement les théorèmes habituels, où l'on ne considère que des fonctions qui sont la somme de variables indépendantes identiquement distribuées.

La démonstration de ce théorème relativement récent n'est pas très compliquée. Elle utilise le même outil isopérimétrique que nous avons présenté ci-dessus sur la sphère.

Il existe de très nombreux variantes et raffinements de ces théorèmes de concentration. Pour ne citer que le plus simple, une fonction de n variables indépendantes telle que la i-ième variable influe sur la fonction d'au plus c_i, sera constante à $\sqrt{\sum c_i^2}$ près. L'indépendance des variables correspondant au produit des espaces de probabilité sous-jacents, ce théorème signifie qu'un produit d'espaces de probabilité bornés est concentré.

La concentration de la mesure n'est pas toujours gaussienne. En particulier, obtenir une estimation de la première valeur propre non nulle du laplacien (sur une variété, sur un graphe) permet de démontrer des théorèmes de concentration où les variations ne sont plus contrôlées par une gaussienne mais par une simple exponentielle. C'est ce que nous faisons, par exemple, dans *Concentration spectrale dans les graphes* ; ce texte applique des techniques bien connues par ailleurs à un cas particulier.

Cependant, il semble que l'indépendance, c'est-à-dire la structure produit, conduise toujours à de la concentration gaussienne. Ainsi en est-il de la concentration spectrale dans les graphes : si l'on prend un produit de graphes manifestant de la concentration exponentielle, le produit montre de la concentration gaussienne au moins à petite échelle.

Sachant que la concentration gaussienne apparaît systématiquement lorsque l'on utilise des espaces produits, on peut se demander si la réciproque est vraie, à savoir : est-ce qu'un espace présentant de la concentration gaussienne est nécessairement proche d'un espace produit ? On doit cependant prendre en compte la remarque élémentaire suivante : si X, Y sont des espaces métriques mesurés, et si on a une application $f : X \to Y$ 1-lipschitzienne qui envoie la mesure de X sur la mesure de Y, alors toute forme de concentration qui existe sur X sera évidemment vérifiée sur Y.

Une contraction (en ce sens) d'un espace concentré est donc concentrée. On peut alors se demander si tout espace présentant de la concentration gaussienne est proche d'une contraction d'un espace produit.

Nous avons un début de résultat en ce sens : sur un espace présentant de la concentration gaussienne, il est possible de trouver des « coordonnées indépendantes », c'est-à-dire une famille de fonctions (non triviales) $f_1, \dots, f_k$ vérifiant une inégalité du type

$$\Pr(|f_1| \geqslant a_1, \dots, |f_k| \geqslant a_k) \leqslant A e^{-(\sum a_i^2)/C}$$

où la constante C est (à un petit facteur près) la même que celle intervenant dans l'hypothèse de concentration gaussienne sur l'espace. Nous renvoyons à *Concentrated spaces seen as product spaces* pour un énoncé précis. Ce théorème n'est que le reflet des débuts d'une recherche en cours qui, nous l'espérons, donnera bientôt des résultats plus précis.

La concentration de la mesure est un sujet récent dont toutes les implications ne sont pas encore éclaircies. Si les aspects probabiliste, statistique et analytique (liens avec la théorie des espaces de Banach, non évoqués ici) de la chose sont désormais relativement bien connus, l'aspect proprement géométrique a été quelque peu laissé de côté. Il a pourtant connu des succès indéniables, lorsque par exemple Gromov a montré que toute variété à courbure de Ricci positive, ainsi que toute variété algébrique complexe, présentait de la concentration gaussienne. Nous espérons continuer cette étude.

3 Quelques exemples d'algorithmes génétiques

3.1 Dynamique de la reproduction sexuée

La situation étudiée est la suivante. On se donne une population composée d'individus caractérisés par leur génome, un génome étant (représenté par) un élément de $\{0,1\}^n$. On suppose qu'à chaque génération, la population est remplacée par une population-fille de la manière suivante : un individu de la population-fille est obtenu en tirant au hasard deux individus dans la population-mère, et le génome de l'enfant est obtenu par mélange probabiliste de ceux des deux parents. Le mélange consiste, pour chaque position dans le génome, à décider que le codon 0 ou 1 présent à cette position sera, avec probabilité $1/2$, celui de l'un ou l'autre des parents (on étudie aussi des méthodes de croisement plus complexes).

L'objectif est d'étudier l'assertion selon laquelle la reproduction sexuée est efficace pour brasser les gènes. La réponse est positive, et le temps de brassage se comporte comme le logarithme de la taille du génome.

On peut faire fonctionner ce processus soit en population finie, soit dans le cas idéalisé d'une population infinie. Une population infinie est une mesure de probabilité sur $\{0,1\}^n$, la mesure d'un élément $x \in \{0,1\}^n$ représentant la proportion des individus de la population présentant ce génome. Une population finie à k individus est simplement un k-uplet d'éléments de $\{0,1\}^n$.

Étant donné une population infinie initiale p_0, le processus à population infinie est déterministe (sur l'espace des mesures de probabilité sur $\{0,1\}^n$). Soit p_t la mesure de probabilité sur $\{0,1\}^n$ obtenue après t générations. On peut montrer (cf. [RRS]) que le processus converge : il existe une mesure de probabilité p_∞ telle que

$$|p_t - p_\infty| \leqslant n^2/2^t$$

où on définit la distance entre deux mesures de probabilité par (distance de variation totale)

$$|p-q| = \frac{1}{2}\sum_{x\in\{0,1\}^n} |p(x)-q(x)| = \sup_{A\subset\{0,1\}^n} |p(A)-q(A)|$$

On voit sur cette expression que le temps de brassage nécessaire pour obtenir une valeur-but de $|p_t - p_\infty|$ est logarithmique en cette valeur-but, logarithmique aussi en la longueur du génome. De plus, cette estimation est essentiellement correcte.

La mesure de probabilité p_∞ peut être caractérisée simplement. Pour $1 \leqslant i \leqslant n$, soit a_i^1 la proportion dans p_0 des individus dont le i-ième bit du génome est un 1, et $a_i^0 = 1 - a_i^1$. Alors, si $x = (x_i) \in \{0,1\}^n$, la mesure p_∞ est définie par

$$p_\infty(x) = \prod_i a_i^{x_i}$$

Autrement dit, les proportions de 0 et de 1 pour chaque bit sont préservées au cours du processus, mais les bits deviennent indépendants les uns des autres. Ainsi si la population initiale est composée pour moitié de l'individu $00\ldots0$ et pour moitié de $11\ldots1$, la population finale sera la mesure uniforme sur tout $\{0,1\}^n$.

En population finie de taille k, la situation est plus complexe et les résultats donnés dans [RRS] étaient insatisfaisants. Cette fois-ci la dynamique est aléatoire sur l'ensemble des k-uplets d'éléments de $\{0,1\}^n$. Soit π_0 le k-uplet initial, supposé donné, et soit π_t le k-uplet aléatoire obtenu après t générations. On s'intéresse à la loi de π_t.

À cause du phénomène bien connu de coalescence, pour t suffisamment grand, avec très grande probabilité le k-uplet π_t sera composé d'une population-clone formée de k individus identiques : en effet, à chaque génération, avec une certaine probabilité une part de l'information génétique est perdue, et (en l'absence de mutations) la diversité génétique ne peut que diminuer. Ceci a au moins le mérite de montrer que le processus converge.

Par contre, cet individu est lui-même une variable aléatoire (connaissant π_0, on ne peut pas prévoir quelle population-clone on va obtenir). Soit q_t la loi du premier élément du k-uplet π_t. C'est sur q_t qu'on va obtenir des estimations.

Le principal résultat est une estimation de la distance entre q_t et une mesure limite p_∞ sur $\{0,1\}^n$. Le k-uplet π_0 formant la population initiale peut naturellement être vu comme une mesure p_0 sur $\{0,1\}^n$ définie par $p_0(x) = \#\{1 \leqslant i \leqslant k, \pi_0(i) = x\}$. Soit p_∞ la mesure obtenue à partir de p_0 par la même définition que ci-dessus en population infinie.

L'estimation que l'on obtient est alors

$$|q_t - p_\infty| \leqslant n^2\left(\frac{1}{k} + \frac{1}{2^t}\right)$$

En particulier, $|q_\infty - p_\infty| \leqslant n^2/k$. Ceci semble être un biais intrinsèque à la population finie. En effet, on peut montrer que pour certaines populations initiales, on a $|q_\infty - p_\infty| \geqslant n/Ck$ pour une certaine constante C.

La vitesse de convergence du processus à population finie est ainsi la même qu'en population infinie ; mais une légère différence, de l'ordre de l'inverse de la taille de la population, apparaît sur le résultat final.

On a ainsi montré que même en population finie, la reproduction sexuée est efficace pour brasser les gènes. Ces résultats s'étendent sans peine à d'autres méthodes de croisement des génomes, qui peut-être modélisent mieux le processus de *crossing-over* biologique.

3.2 L'espace des arbres phylogénétiques

Les algorithmes génétiques sont utiles pour parcourir des espaces qu'on ne connaît pas explicitement ou bien qui sont trop grands pour faire l'objet d'une énumération exhaustive. On a tenté d'exploiter cette propriété pour explorer l'espace des arbres sur un ensemble fixé de feuilles.

Ce problème se pose en particulier pour la reconstruction phylogénétique, c'est-à-dire la recherche de l'arbre évolutif entre les espèces vivantes (il peut aussi se présenter en linguistique). Le problème est le suivant : on étudie un certain nombre d'espèces, qui sont connues à travers certaines caractéristiques comme une partie de leur génome, ou bien un ensemble de traits morphologiques.

Le but est de trouver un arbre ayant ces espèces pour feuilles, et qui minimise un certain critère (comme le nombre total de mutations) censé représenter la plausibilité de l'arbre comme modèle du véritable arbre de l'évolution. Le problème majeur est que le nombre d'arbres possibles sur N feuilles données croît extrêmement vite (au moins comme $N!$ comme le montrent les arbres où chaque espèce se détache l'une après l'autre d'un tronc commun).

On s'est proposé d'écrire un programme de recherche d'arbres phylogénétiques qui fonctionne comme un algorithme génétique : on maintient en permanence une « population » d'arbres candidats, et cette population évolue par sélection, mutation et croisement.

La principale innovation de cette approche par rapport aux programmes existants consiste en l'utilisation d'un croisement entre arbres, analogue du croisement biologique et du croisement des chaînes de $\{0,1\}^n$ étudié ci-dessus. Nous le décrivons ici.

Soit F un ensemble fini (les feuilles de l'arbre, ou les espèces à étudier). Un *nœud* sera une partie non vide de F. Un *arbre* (enraciné) dont l'ensemble des feuilles est F peut être défini par l'ensemble des nœuds qu'il contient. Ainsi l'arbre sur les feuilles $\{a,b,c,d\}$ dans lequel se séparent a et b d'une part, et c et d d'autre part, a comme ensemble de nœuds $\{\{a,b,c,d\},\{a,b\},\{a\},\{b\},\{c,d\},\{c\},\{d\}\}$. Il est facile de voir qu'un ensemble E de nœuds (contenant le nœud complet ainsi que les singletons) définit bien un arbre si et seulement si pour

tous $A, B \in E$, on a soit $A \subset B$, soit $B \subset A$, soit $A \cap B = \varnothing$ (condition de compatibilité, qui exprime que les nœuds peuvent s'emboîter).

L'ensemble des arbres (enracinés, avec un ensemble de feuilles donné) est naturellement muni d'un ordre qui dit qu'un arbre est plus fin qu'un autre si son ensemble de nœuds est plus grand. Étant donné deux arbres, leur inf pour cet ordre est l'arbre dont l'ensemble de nœuds est l'intersection des ensembles de nœuds des deux arbres (consensus strict : un groupement de feuilles appartient à l'inf de deux arbres si et seulement s'il appartient aux deux arbres).

Cet inf n'est pas un bon candidat pour un croisement d'arbres dans le cadre d'un algorithme génétique. En effet il a tendance à créer des dégénérescences et perd beaucoup d'information : dès que les parents diffèrent, le caractère correspondant de l'enfant n'est pas défini ! Le croisement biologique, et le croisement sur $\{0,1\}^n$ étudié plus haut, choisissent au hasard entre les deux parents en cas de désaccord.

Un bon candidat au croisement d'arbres pourrait être le suivant. Étant donné deux arbres ayant des ensembles de nœuds E et E', considérer la réunion $E'' = E \cup E'$. En général elle ne définit pas un arbre. Ordonner au hasard les éléments de E''. Puis, parcourir la liste de nœuds ainsi obtenue ; dès qu'un nœud ne vérifie pas la condition de compatibilité avec l'un des nœuds qui le précèdent, supprimer ce nœud de la liste. Passer à l'examen du nœud suivant. Par construction, après examen et éventuelle suppression de tous les nœuds, la condition de compatibilité est respectée.

Ce croisement est une opération aléatoire puisqu'il dépend d'un choix aléatoire de l'ordre d'examen des nœuds, dont le résultat dépend en général. De plus, il vérifie la condition naturelle qu'il produit un arbre plus fin que le consensus des deux arbres parents. Enfin, il n'a pas tendance à créer trop de dégénérescences.

On a donc implémenté un algorithme génétique sur l'espace des arbres phylogénétiques utilisant ce croisement, ainsi que des opérateurs de sélection et de mutation standard. On a comparé les résultats à ceux d'un logiciel classique dans le domaine. Le programme a donné des résultats de qualité comparable, mais avec des temps de calcul supérieurs ; en tout état de cause l'intérêt de l'utilisation du croisement n'a pas été clairement démontré. Hors de l'algorithme génétique, qui semble très gourmand en temps de calcul, il pourrait être intéressant de reprendre ce croisement pour l'intégrer à un logiciel classique.

3.3 Le Product Replacement Algorithm

Le Product Replacement Algorithm est une heuristique utilisée en théorie algorithmique des groupes pour produire un élément aléatoire uniformément réparti dans un groupe fini. Le groupe est donné comme une « boîte noire », c'est-à-dire qu'on fournit trois routines effectuant respectivement la multiplication de deux éléments, l'inversion d'un élément, la comparaison d'un élément avec l'élément neutre, ainsi qu'un système générateur. (Ce peut être le cas, typi-

quement, d'un groupe de matrices qu'on ne sait pas décrire explicitement mais dont on connaît un système générateur).

Le but de l'algorithme est de fournir un élément « typique » du groupe, c'est-à-dire un élément aléatoire dont la loi soit (proche de) la loi uniforme sur le groupe. Construire de tels éléments est utile en théorie algorithmique des groupes pour tester (probabilistement) certaines propriétés algébriques.

Le PRA (Product Replacement Algorithm) est un algorithme qui semble fonctionner extrêmement bien en pratique. Ses fondements théoriques sont mal compris (voir [Pak] pour un survol). Son fonctionnement est le suivant.

On se donne un k-uplet générateur $s_1, \ldots, s_k$ du groupe. Ce k-uplet n'est pas forcément minimal et peut par exemple contenir plusieurs fois le même élément. À chaque étape, on modifie ce k-uplet de la manière suivante. On tire au hasard deux indices distincts $1 \leqslant i, j \leqslant k$. Puis on remplace, dans le k-uplet, le générateur s_j par, soit $s_j s_i$, soit $s_j s_i^{-1}$, soit $s_i s_j$, soit $s_i^{-1} s_j$ (en choisissant au hasard entre ces quatre possibilités). On ne touche pas à s_i. Il est immédiat de voir que si le premier k-uplet engendrait le groupe, le nouveau l'engendre encore.

L'idée est que les éléments d'un tel k-uplet vont s'éloigner très vite du k-uplet initial. Si l'on raisonne en termes de longueur, si au temps t les éléments du k-uplet sont à distance en moyenne r de l'origine, au temps $t+1$ la distance moyenne à l'origine sera $r(1+1/k)$ parce qu'on a multiplié un des éléments par un autre. D'où une croissance supposée exponentielle de la distance à l'origine. Ceci s'oppose fortement à l'algorithme le plus simple, la marche aléatoire sur le groupe, où à chaque étape la distance à l'origine augmente au plus de 1.

L'analogie avec un algorithme génétique est claire : on maintient une population d'éléments et on les croise. Ceci a permis d'obtenir un résultat de convergence du PRA lorsque la taille de la population est très grande. Cependant, cette évaluation n'a pas d'intérêt pratique car la méthode utilise qu'en très grande population, une partie du PRA « simule » une marche aléatoire. On ne prouvera donc pas par cette méthode (du moins sans modification importante) que le PRA fait mieux que la marche aléatoire.

On a aussi exhibé un invariant curieux (mais sans application connue) du PRA sur les groupes cycliques, ainsi qu'une sorte de borne inférieure à la qualité de l'algorithme.

* * *

Après avoir donné un aperçu des sujets abordés ici avec plus ou moins de bonheur dans la recherche, nous laissons désormais au lecteur le labeur de parcourir les textes qui suivent.

Références

[Ch] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), No. 5, 438–444.

[Gh] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003).

[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.

[Pak] I. Pak, *What do we know about the product replacement algorithm?*, in *Groups and Computation III*, eds. W. Kantor, A. Seress, de Gruyter, Berlin (2001), 301–347.

[Pau] F. Paulin, *Sur la théorie élémentaire des groupes libres*, Séminaire Bourbaki **922** (2003).

[RRS] Y. Rabani, Y. Rabinovich, A. Sinclair, *A computational view of population genetics*, Random Structures and Algorithms **12** (1998), No. 4, 313–334.

[Tal] M. Talagrand, *A new look at independence*, Ann. Prob. **24** (1996), No. 1, 1–34.

Table des matières

I
Théorie des groupes

Contenant :

Sharp phase transition theorems for hyperbolicity of random groups

Growth and cogrowth of generic groups

On a small cancellation theorem of Gromov

Sharp phase transition theorems for hyperbolicity of random groups

With 31 illustrations

Abstract

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

Introduction

What does a generic group look like?

The study of random groups emerged from an affirmation of M. Gromov that "almost every group is hyperbolic" (see [Gro1]). More precisely, fix m and N and consider the group G presented by $\langle a_1, \dots, a_m \mid r_1, \dots, r_N \rangle$ 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 $\inf \ell_i \to \infty$. The first proof of this theorem was given by A.Y. Ol'shanskiĭ in [Ols1], and independently by C. Champetier in [Ch1], thus confirming Gromov's statement.

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

This model goes as follows: Choose at random N 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 the presentation $\langle a_1, \dots, a_m \mid R \rangle$.

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 $a_1^{\pm 1}, \dots, a_m^{\pm 1}$.*

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

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

There was a small mistake in the original proof of Gromov: the proof uses van Kampen diagrams, and the case when some relator appears several times in a given van Kampen diagram was forgotten (this mistake was apparently first detected by R. Kenyon); when no relator appears twice there is much more independence in the probabilities and the proof is easier. 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 assimilate $d\ell$ to a dimension. That is, $d\ell$ represents the number of "equations" we can impose on a random word so that we still have a reasonable chance to find such a word in a set of $(2m-1)^{d\ell}$ randomly chosen words (compare to the basic intersection theory for random sets stated in section 5.2).

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

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

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

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

Our theorems precisely state that for each hyperbolic group (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 15). 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: Appendix C proves that Theorem 4 does not hold for some hyperbolic groups with harmful torsion [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 $a_1, \dots, a_m$. 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.

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*

[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 Appendix C.

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

Let θ be the gross cogrowth of the group G.

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

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

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

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

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

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

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

Discussion of the models. Of course, the three theorems given above are not proven 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 Theorem 3 still holds when quotienting by words taken in the ball rather than in the sphere, or 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 not the case $d = 0$ of our model, since in this model the ratio of lengths can be unbounded, which completely prevents the use of some geometric methods. However, this model can probably be obtained by iterating the process of taking a random quotient at $d = 0$, or by using the relative small cancellation techniques later developed by Delzant in [D] and by Gromov in [Gro4].

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

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

geodesic form, and that there exists a finite automaton which recognizes exactly the words which are normal forms of elements of the group.

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

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^*-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 the counterexamples of Appendix C, 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 at receiving me.

1 Definitions and notations

1.1 Basics

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

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

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

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

A word is said to be *reduced* if it does not contain a generator $a \in S$ 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 (see e.g. [GdlH] or [GK] for background and open problems). Cogrowth has been introduced by R. Grigorchuk in [Gri], and independently by J. Cohen in [C]. For some examples see [Ch2] or [W1]. Gross cogrowth is linked (see below) to the spectrum of the random walk on the group, which, since the seminal work by H. Kesten (see [K1] and [K2]), has been extensively studied (see for example the numerous technical results in [W2] and the references therein).

DEFINITION 5 (GROWTH, COGROWTH, GROSS COGROWTH) –

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

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

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

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

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

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

Let us state some properties of these quantities. All of them are proven 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 $(2m-1)^\ell$.

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 states that the return probability at time t is roughly equal to $(2m)^{-(1-\theta)t}$. This will be ubiquitous in our text.

ALTERNATE DEFINITION OF GROSS COGROWTH – *Let P_t be the probability that a random walk on the group G (with respect to the generating set $a_1, \dots, a_m$) 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, cogrowth is the growth (in base $2m-1$) of N. 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.)

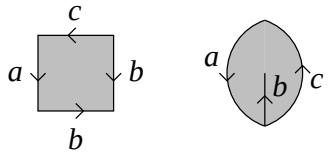
Cogrowth and gross cogrowth depend on the generating set. For example, adding trivial generators $a_i = e$ makes them arbitrarily close to 1.

1.3 Diagrams

A *filamenteous van Kampen diagram* in the group G with respect to the presentation $\langle S \mid R \rangle$ 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 6.3 (more decoration is added).

A davKd is said to be *fulfillable* w.r.t. presentation $\langle\, S \mid R \,\rangle$ 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 $\langle\, S \mid R \,\rangle$.

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.

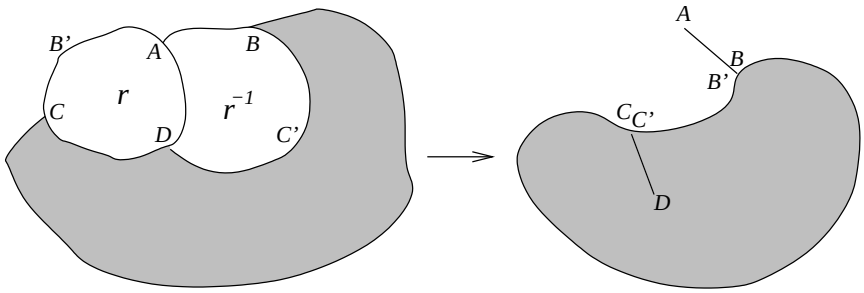
A word w is equal to the neutral element e in G if and only if some no-hole, maybe filamenteous, davKd with border w is fulfillable (see [LS]).

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[2].

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 strictly bigger 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 big enough ℓ, all couples of generators $a, b \in S$ will satisfy $a = b$ in the quotient group. As S is symmetric, in particular they will satisfy $a = a^{-1}$.

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

This proves the second part of Theorem 1.

[2]Since the proof included here was written and diffused, a similar but somewhat simpler proof has been published in [Z] for a slightly different model in which relators are of length 3 but the number of generators m tends to infinity.

2.2 Hyperbolicity for $d < 1/2$

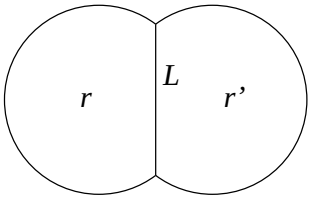
We proceed as follows: We will show that the (reduced) davKd's which are fulfillable by a random presentation necessarily satisfy some linear isoperimetric inequality. This is stronger than proving that only minimal diagrams satisfy an isoperimetric inequality: in fact, *all* reduced diagrams in a random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

Thus we will evaluate the probability that a given decorated abstract 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 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 theorem.

The intuitive 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 bigger than $(1-2d)n\ell$ which is the isoperimetric inequality we were looking for.

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

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

PROPOSITION 6 – *Let D be a reduced davKd. The probability that D can be fulfilled by relators of a random presentation is at most $(2m-1)^{(|\partial D| - \ell|D|(1-2d))/2|D|}$.*

PROOF – 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, $n \leqslant |D|$. (The original proof by Gromov was valid only when $n = |D|$, so that all relators are chosen independently, which simplifies the proof. If $n < |D|$ then we cannot simply multiply probabilities as in the basic picture.) 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.

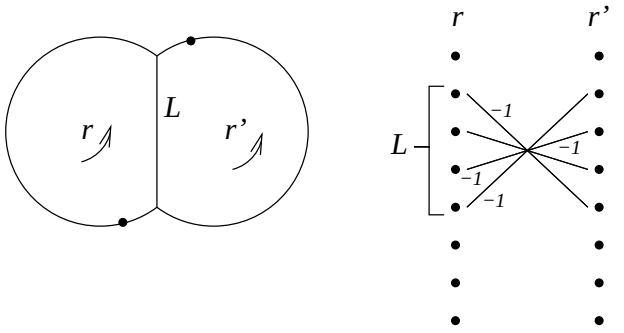
Thus, a -1 edge between the k-th vertex of the i-th part and the k'-th vertex of the j-th part means that the k-th letter of relator r_i has to be the inverse of the k'-th letter of relator r_j.

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:

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

Before concluding we need a further purely combinatorial lemma.

LEMMA 7 –

$$|\partial D| \geqslant \ell\,|D|\,(1-2d) + 2\sum d_i(m_i - m_{i+1})$$

where $m_i, 1 \leqslant i \leqslant n$ is the number of faces of D bearing relator number i.

Before proving the lemma let us end the proof of the proposition. We are free to choose the order of the construction, and we may suppose that the m_i's are non-increasing, i.e. that we began with the relator appearing the biggest number of times in D, etc., so that $m_i - m_{i+1}$ is non-negative.

If all d_i's are non-negative, then we have the isoperimetric inequality $|\partial D| \geqslant \ell\,|D|\,(1-2d)$ and the proposition is true since the probability at play is at most 1.

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

PROOF OF THE LEMMA – 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

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

(where the equality $|\partial D| = |D|\,\ell - 2A$ holds when D has no filaments).

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,

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

as was needed. □

COROLLARY 8 – *Let D be a davKD. Then, either D satisfies the isoperimetric inequality*

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

or the probability that is can be fulfilled by relators of a random presentation is at most $(2m-1)^{-\ell(1/2-d)/2}$.

Hence the interest of taking $d < 1/2$...

This was for a given davKd D. 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 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 $N(K, \ell)\,(2m-1)^{-\ell(1/2-d)/2}$.

PROPOSITION 9 – *For fixed K, the number $N(K, \ell)$ grows polynomially with ℓ. Hence, the probability $N(K, \ell)\,(2m-1)^{-\ell(1/2-d)/2}$ tends exponentially to 0 as $\ell \to \infty$.*

PROOF – 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 section 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 $N(K, \ell) \leqslant G(K)(2K)^K \ell^{4K}$. □

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.

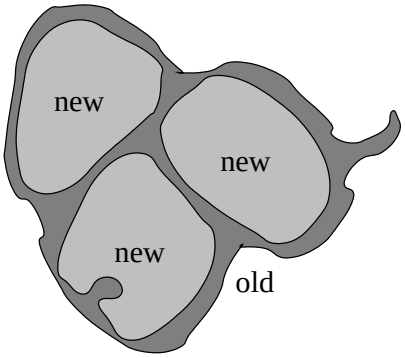
We will give a general theorem for hyperbolicity of random quotients by words taken from some probability measures on the set of all words. We will need somewhat technical axioms on the measures (for example, that they weight only long words). Here we give a heuristic justification of why these axioms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

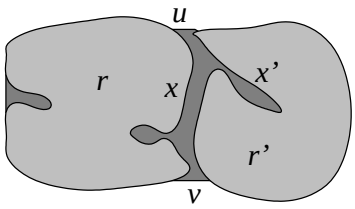
Say the new relators have length of order ℓ where ℓ is much bigger than the hyperbolicity constant of G. (This will be Axiom 1.)

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



3.1 A basic picture

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



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

If, as in the standard case, there were no glue (no old relators) and r and r' were uniformly chosen random reduced words, the probability that r and r' could be glued along subwords 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 then x and x' are equal, not as words, but as elements of G (and up to small words u and v, which we will neglect).

If, for example, the relators are uniformly chosen random words, then x and x' are independent subwords, and the probability that x and x' are (almost) equal in G is the probability that $xx'^{-1} = e$; 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 alternate definition of gross cogrowth in section 1.2).

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

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

Axiom 4 will deal with the special case when $r = r'^{-1}$, so that the words x and x' above are equal, and not at all chosen independently as we implicitly assumed above. In this case, the 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 bigger than 2ℓ. This is exactly what we want to prove: either the boundary is big, or the probability of existence of the diagram is small.

This is comparable to the former situation with random quotients of the free group: in the free group, imposing two random relators to glue along subwords of lengths L and $L' = L$ results in L "equations" on the letters. 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 $A > n\ell d/\beta$, the probability decreases exponentially with ℓ; otherwise, the boundary is longer than $n\ell(1 - d/\beta)$.

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

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

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

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

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

4.1 Asymptotic notations

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

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

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

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 10 – *Let P_ℓ be a family of properties of words, indexed by the integer ℓ. We say that*

$$\textit{for any subword } x \textit{ under } \mu_\ell,\ \Pr(P_\ell(x)) \lesssim p(\ell)$$

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

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

or

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

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 11 – *Let P_ℓ be a family of properties depending on two words, indexed by the integer ℓ. We say that*

$$\text{for any two disjoint subwords } x, y \text{ under } \mu_\ell,\ \Pr(P_\ell(x,y)) \lesssim p(\ell)$$

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

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

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

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

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 κ_1 such that μ_ℓ weights only words of length between ℓ/κ_1 and $\kappa_1\ell$.*

Note this axiom applies to words picked under μ_ℓ, and not especially subwords, so it does 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 not probably represent short elements of the group.

AXIOM 2 – *There are constants κ_2, β_2 such that for any subword x under μ_ℓ, for any $t \leqslant 1$, 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 function $n = n(\ell)$, for any two disjoint subwords x, y under μ_ℓ, 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 function $n = n(\ell)$, for any subword x under μ_ℓ, 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 $(2m)^{\gamma n}(2m)^{-\beta_4|x|}$*

If G has big 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 function $n = n(\ell)$, for any subword x under μ_ℓ, 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 12 – 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 (i.e. by what the rest of the word is up to the position of the subword), 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.

In [Gro4], p. 139–141, M. Gromov uses similar-looking properties. His $\mathbf{pr_1}$ is similar to our Axiom 2, and his $\mathbf{pr_3}$ controls the same kind of event as our Axiom 4. We no not use any analogue of his $\mathbf{pr_2}$, and analogues of our Axioms 1 and 3 are indeed present in [Gro4] but in a more "diffuse" way in the paper. Also note that in [Gro4] emphasis is put on very small densities, so that the properties considered therein are of the form "such event is realized with probability exponentially close to 1", whereas since we work in large densities we have to get a precise control of the tails of the distributions, and so our axioms take the form "the probability of a deviation of size L from such event is at most $\exp(-\beta L)$", with a tight value of β needed. So our axioms (which have been found independently of [Gro4]) are more precise quantitatively.

4.4 The Theorem

Our main tool is the following

THEOREM 13 – *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 14 – Remark 12 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 15 (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 16 – *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 61 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 17 – *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 18 – *Suppose G has strongly harmless torsion, and that Axioms 1, 3 and 4' are satisfied. Set $\beta = \min(\beta_3, \beta_{4'})$.*

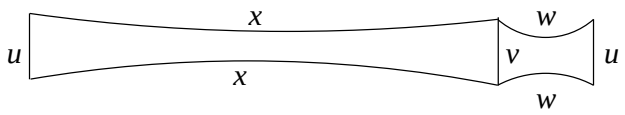
There is a constant γ such that for any subword x under μ_ℓ, the probability that there exist words u, v of length at most $n = n(\ell)$, such that $ux = xv$ in G, with u, v not equal to e, is roughly less than $(2m)^{\gamma n - \beta|x|}$.

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 61 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 16. 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 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 62 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 17 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 60, 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 17 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 64, $vxwv^{-1}$ 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 13 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 13 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 proven.

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

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

Furthermore, proving triviality of a quotient involves small scale phenomena, which are very different in our three models of random words (think of a

random quotient of $\mathbb{Z}$ by random words of ℓ letters ± 1 or by elements of size exactly ℓ).

These are the reasons why the next three sections are divided in cases, and why we did not include these properties in a general and technical theorem such as Theorem 13.

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

5.1 Satisfaction of the axioms

5.1.1 The case of plain random words

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

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

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

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

We will show that

PROPOSITION 19 – *Axioms 1, 2, 3, 4' are satisfied by plain random uniformly chosen words, with exponent $1-\theta$ (in base $2m$).*

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

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

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

PROPOSITION 20 – *The probability that B_ℓ is equal to e is roughly less than $(2m)^{-(1-\theta)\ell}$.*

PROOF – Alternate definition. □

PROPOSITION 21 –

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

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 22 – *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 23 – *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 21 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 24 – *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 alternate model of random geodesic words.

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

Let $x \in G$ such that $\|x\| = \ell$. We have $\Pr(X_\ell = x) \approx (2m)^{-g\ell}$. Let $x = x_1x_2\dots x_N$ where each x_i is of length ℓ/N. The probability that the i-th segment

of X_ℓ^N is equal to x_i is roughly $(2m)^{-g\ell/N}$. Multiplying, we get $\Pr(X_\ell^N = x) \approx (2m)^{-g\ell}$.

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

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

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

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

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

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

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

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

PROPOSITION 25 – *Let $x \in G$ and $L \leqslant \ell$. We have*

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

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

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

and

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

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 25 to X_ℓ and x^{-1} and write out the definition of the Gromov product. □

PROPOSITION 27 – *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 26, 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 28 – *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 27 starting at e, this occurs with probability $(2m)^{-g(|x|-L)/2}$. Now we have $\|xu\| \geqslant L - \|u\|$, but $\|xuy\| = \|v^{-1}\|$. By Proposition 27 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 29 – *Axiom 4′ is satisfied for random geodesic words, with exponent $g/2$.*

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

Note that applying Proposition 27 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 27 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 30 – *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 bigger 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 bigger 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 13 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 31 – *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 32 – *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 13 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 14, 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 13.

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 $\langle\, S \mid Q \,\rangle$. 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 $\langle\, S \mid Q \,\rangle$, 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 notations even heavier, *we will suppose that all the words taken from μ_ℓ are of length ℓ*. So R is made of $(2m)^{d\ell}$ words of length ℓ. This is the case in all the applications given in this text.

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.

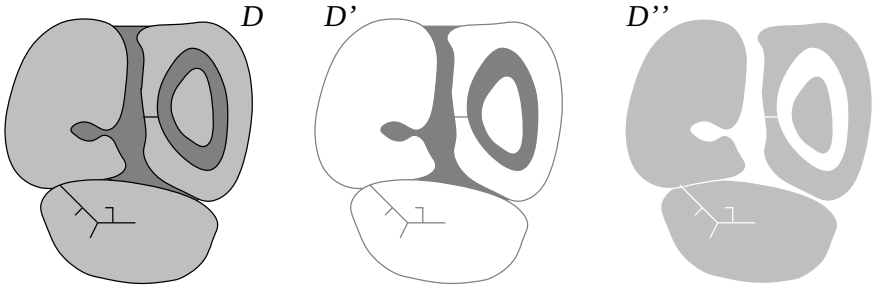
All these properties are guaranteed by Axiom 1.

6.2 Combinatorics of van Kampen diagrams of the quotient

We now proceed to the application of the program outlined in section 3. The reader may want to refer to this section while reading the sequel of this text.

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).

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 sense than previously, which we define now.

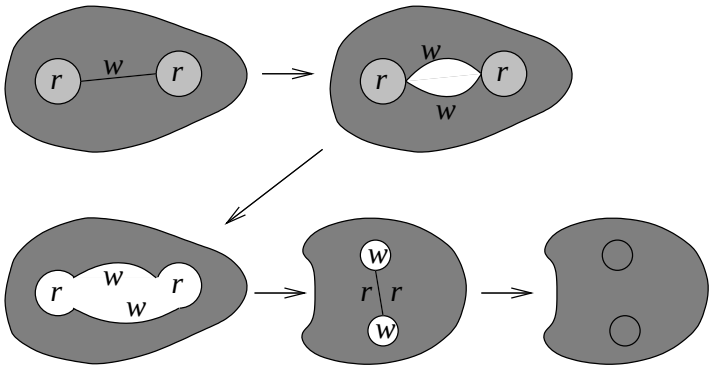
DEFINITION 33 – *A van Kampen diagram $D = D' \cup D''$ on $G/\langle R \rangle$ (with D' and D'' as above) is said to be* strongly reduced *with respect to G if there is no pair of faces of D'' bearing the same relator with opposite orientations, such that their marked starting points are joined in D' by a simple path representing the trivial element in G.*

In particular, a strongly reduced diagram is reduced.

PROPOSITION 34 – *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. □

Often in geometric group theory, problems arise when two relators are conjugate (or when a conjugate of a relator is very close to another relator), and such cases are typically excluded by reinforcing the definition of "strongly reduced". In the case of random presentations, however, below the critical density it never occurs that two relators are conjugated. So we do not have to care about these problems: these cases are automatically wiped off by our axioms.

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 theorem, cf. Appendix A), it is enough to show it for diagrams having less than

some fixed number of faces. This crucial point considerably simplifies the geometric and probabilistic treatment. More precisely, we will show the following.

PROPOSITION 35 – *There exist constants $\alpha, \alpha' > 0$ (depending on G, d and the random model 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

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

(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 36 – *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 35, 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*

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

REMARK 37 – Proposition 36 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 53. Take for property P in this proposition "to be strongly reduced". Recall the notations of the appendix: $L_c(D) = |\partial D|$ is the boundary length of D, and $A_c(D)$ is the area of D in the sense that a relator of length L has area L^2. Note that $\ell \, |D''| + |D'| \geqslant A_c(D)/\ell$.

Take a 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 35 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 53 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

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

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

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

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

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

is arbitrarily close to the original isoperimetry constant in G.

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

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

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

Now for the proof of Proposition 35.

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. Similarly, if the old relators are much more numerous that the new ones, then isoperimetry of G is enough. Namely:

LEMMA 38 – *Proposition 35 holds for diagrams satisfying* $|D'| \geqslant 4\,|D''|\,\ell/C$.

PROOF OF THE LEMMA – 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 67 for diagrams with holes in G, we have that $|\partial D'| \geqslant C\,|D'| - |D''|\,\lambda(2 + 4\alpha \log |D'|)$. So, for ℓ big enough,

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

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

So *we now suppose that* $1 \leqslant |D''| \leqslant K$ *and that* $|D'| \leqslant 4\,|D''|\,\ell/C$. In particular, the boundary length of D is at most $|D''|\,\ell + |D'|\,\lambda \leqslant \ell\,|D''|\,(1 + 4\lambda/C)$.

6.3 New decorated abstract van Kampen diagrams

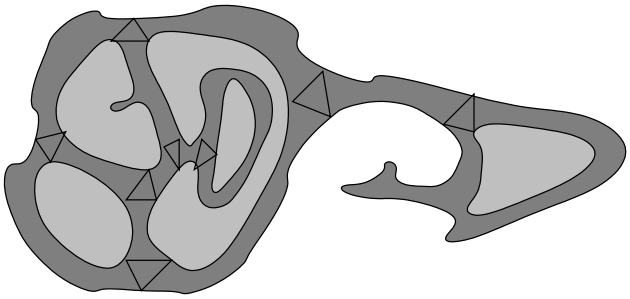
We now redefine decorated abstract van Kampen diagrams so that they better fit our needs (the definition given in the introduction fits the case of free groups only). The idea is that since D' is very narrow (at the scale of ℓ), at scale ℓ D looks like a van Kampen diagram with respect to the new relators, with some narrow "glue" (that is, old relators) between faces. This intuition will be formalized using Proposition 72 in Appendix B, which will help tell which parts of the boundary words of the new relators are facing which.

The diagram D' has at most K holes. First, after Corollary 69, we can suppose that D' is $E_1 \log \ell$-narrow for some constant E_1 depending on G and K but not on ℓ (here we used $|D'| \leqslant 4K\ell/C$ to get logarithmic dependence on ℓ).

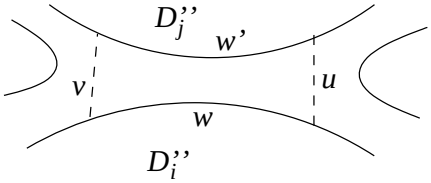
Besides, we can apply Proposition 72 to D'. This defines a $(8K, E_2 \log \ell)$-matching X (see Definition 70) between at most $8K$ subwords of the boundary words of D', for some constant E_2 depending on G and K but not on ℓ (here

again we used $|D'| \leqslant 4K\ell/C$ to get logarithmic dependence on ℓ). Set $E = \max(E_1, E_2)$.

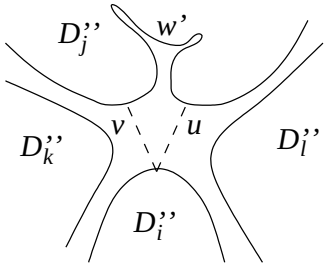
The boundary words of D' are precisely the new relators on one side, and the boundary word of D on the other side.



Each match in X is a pair of two subwords w, w' of one of the new relators (or of the boundary word), together with two short words u, v of length at most $E \log \ell$, such that $w = uw'v$ in G.



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.



Intuitively, we can reconstruct D "at scale ℓ" if we know this matching X: simply take the new relators and glue them along the matches in X. This leads to redefining what a decorated abstract van Kampen diagram is. Knowing D, the associated abstract diagram $\mathcal{D}$ will keep the combinatorial and geometric information but will forget what are the precise values of the new relators. Namely, given D we only keep the following information: How many new faces there are (that is, $|D''|$, which is at most K); Which new faces bear the same new relator or not (this can be done by attributing a number between 1 and $|D''|$ to

each face, two faces getting the same number if and only if they bear the same new relator); Which subword is matched to which one (that is, where the cuttings of the subwords were done and what the pairing is). This leads to the following (compare the definition of a davKd given in the introduction page 40, together with Proposition 72).

DEFINITION 39 – *A* decorated abstract van Kampen diagram *(davKd for short) $\mathcal{D}$ is the following data:*

- *An integer k (the number of faces), also denoted $|\mathcal{D}|$; any number between 1 and k will be called a* face *of $\mathcal{D}$.*
- *An integer $|\partial\mathcal{D}|$ between 0 and $|\mathcal{D}|\,\ell(1+4\lambda/C)$, called the* boundary length *of $\mathcal{D}$.*
- *A set of k integers between 1 and k (which faces bear the same relator).*
- *For each face, a number between 1 and ℓ (a starting point for the relator) and an orientation ± 1.*
- *A partition of the set $\{1,\dots,\ell\}\times\{1,\dots,k\}\cup\{1,\dots,|\partial\mathcal{D}|\}\times\{k+1\}$ into $8k$ subsets (some of which may be empty) of the form $\{i,i+1,\dots,i+j\}\times\{p\}$. The subsets $\{1,\dots,\ell\}\times\{p\}$ will be called* words *in $\mathcal{D}$, with $\{1,\dots,|\partial\mathcal{D}|\}\times\{k+1\}$ being the* boundary word *and the others* internal words. *The elements of the partition will be called* subwords *in $\mathcal{D}$, and the* length *of a subword $\{i,i+1,\dots,i+j\}\times\{p\}$ will be $j+1$.*
- *A partition of the set of subwords into two parts and a bijection between these parts (which subword is matched to which). A pair of two bijected subwords will be called a* match *in $\mathcal{D}$.*

A very important fact is the following one.

PROPOSITION 40 – *For a fixed K, the number of different davKd's with at most K faces is less than some polynomial in ℓ.*

PROOF – This is at most $K.K\ell(1+4\lambda/C).K^K.\ell^K.2^K.((K+1)\ell)^{8K}.(8K)^{8K}$. □

We will still add some decoration below in section 6.8. This further decoration will again be polynomial in ℓ.

We just saw that to our van Kampen diagram D we can associate a decorated abstract van Kampen diagram $\mathcal{D}$, coming from Proposition 72. This we will call the *davKd associated to D*. This davKd sums up all quasi-equalities in G imposed by the diagram on the new relators.

DEFINITION 41 – *Let $\mathcal{D}$ be a davKd. We say that a van Kampen diagram $D = D'\cup D''$ of $G/\langle R\rangle$* fulfills *$\mathcal{D}$ if $\mathcal{D}$ is the davKd associated to D and if D satisfies the assumptions of Proposition 35 and Lemma 38 that is:*

- *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;*
- $|D'| \leqslant 4|D''|\,\ell/C$.

A davKd $\mathcal{D}$ is said to be fulfillable *if some van Kampen diagram fulfills it. A davKd $\mathcal{D}$ is said to satisfy an α-isoperimetric inequality if*

$$|\partial\mathcal{D}| \geqslant \alpha\ell\,|\mathcal{D}|$$

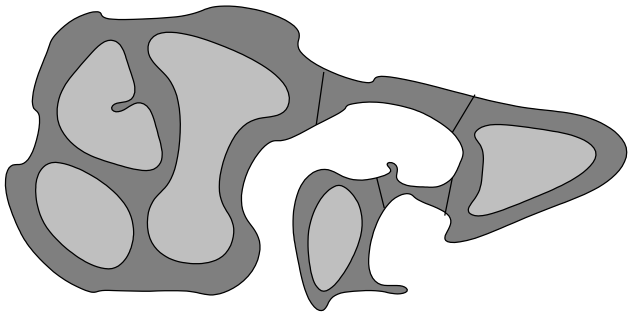
PROPOSITION 42 – *If some davKd $\mathcal{D}$ satisfies an α-isoperimetric inequality, then any van Kampen diagram $D = D' \cup D''$ fulfilling $\mathcal{D}$ satisfies the isoperimetric inequality*

$$|\partial D| \geqslant \alpha\ell\,|D''|/2 + C\alpha\,|D'|/8$$

PROOF – Indeed, since $|D'| \leqslant 4|D''|\,\ell/C$ we have $\alpha\ell\,|D''|/2 + C\alpha\,|D'|/8 \leqslant \alpha\ell\,|D''|$. □

Thus, to prove Proposition 35 we have to show that, with high probability, any fulfillable davKd satisfies some linear isoperimetric inequality (with some isoperimetric constants depending on G, the density d and the random model but not on K or ℓ).

In the matching X of D, there may be matches between subwords of the boundary, as in the following figure. Such parts of the diagram always improve isoperimetry (up to $2E\log\ell$). So *in the following we consider that all matches in X match either two subwords of the new relators or a subword of a new relator and a subword of the boundary*.



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, then by gluing "up to small words" the faces along the subwords desribed in the davKd, 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. Internal subwords in D are identified with successive vertices of Γ (with a reversal if the face in D to which the subword belongs is negatively oriented).

We now explain what to take as edges of Γ.

Let f be a match in D. First, suppose that this is a match between two internal subwords in D. Say it is a match between subwords of faces of D bearing numbers i and i'. These two subwords in D correspond to two sets of successive vertices in the i-th part and the i'-th part of Γ.

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

(This may result in double edges if $i = i'$. We will deal with this problem later, but for the moment we keep the double edges.)

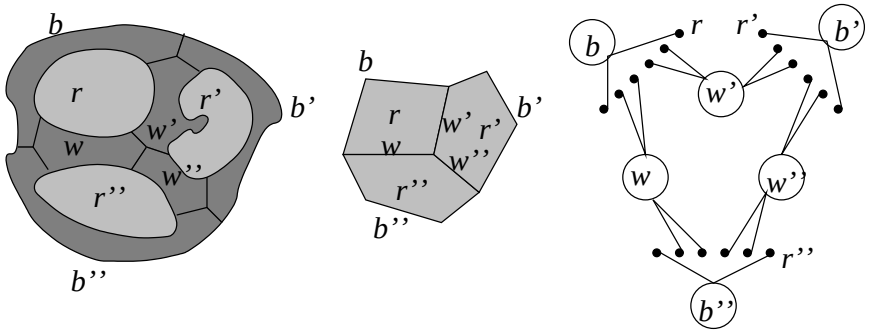
Follow this process for all matches between internal subwords of D. Each translator so obtained is connected with two (or maybe one if $i = i'$) parts of Γ.

As several faces of D may 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 face of D.

For each match in D involving a boundary subword and an internal subword adjacent to, say, face i, add a special vertex b to Γ, called a *boundary translator*. Add edges between b and the vertices of the i-th part of Γ corresponding to the internal subword of the match at play.

At the end of the process, the number of edges in Γ is equal to the cumulated length of all internal subwords of D, which is $\ell\,|D|$.

Here is an example of a simple van Kampen diagram on $G/\langle R\rangle$, its associated davKd (represented graphically by a diagram "at scale ℓ"), and the associated graph Γ.



In a davKd associated to some van Kampen diagram with at most K faces, as we only consider, since the number of matches is at most $4K$, the number of internal and boundary translators in Γ is at most $4K$ as well.

Note that each translator corresponds, via Proposition 72, to a word in the van Kampen diagram which is equal to e in G: translators are nothing else but matches of the davKd. Indeed, fulfillability of the davKd implies that for each (say internal) 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 van Kampen diagram, two nearly adjacent faces bear the same relator.

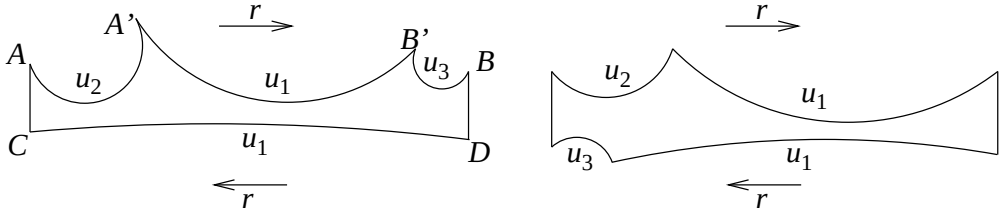
Doublets are annoying since the two sides of the translator are not chosen independently, whereas our argument requires some degree of independence. We will split the corresponding translators to control the occurrences of such a situation.

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 $u = u_2 u_1 u_3$ and $u' = u_1$, or $u = u_2 u_1$ and $u' = u_1 u_3$.



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.

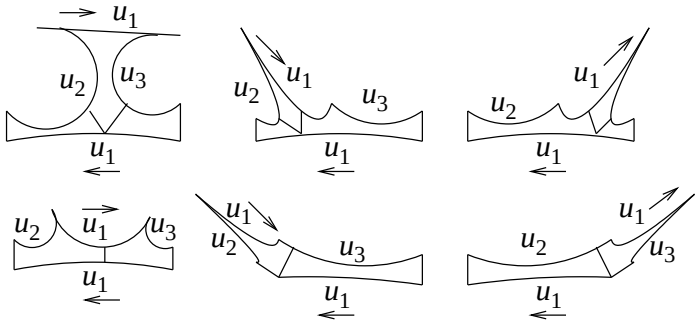
By elementary hyperbolic geometry, and given that the two lateral sides are of length at most $2E \log \ell$, any point on CD is $(2\delta + 2E \log \ell)$-close to some point on AA' or $B'B$, or 2δ-close to some point on $A'B'$.

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.

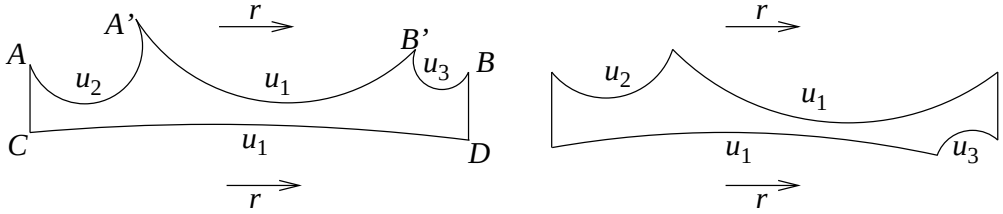
There are six cases depending on whether C_{i_0}' and C_{i_0-1}' belong to AA', $A'B'$ or $B'B$. In each of these cases we can cut the diagram in at most three parts, in such a way that no part contains two copies of some subword of u_1 (except perhaps up to small words of length at most 8δ at the extremities). 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 $u' = u_1u_3$.



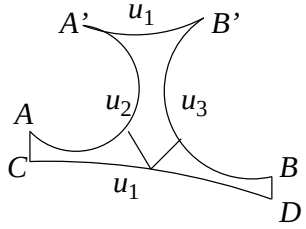
We will only treat the first case, as the second is similar.

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.

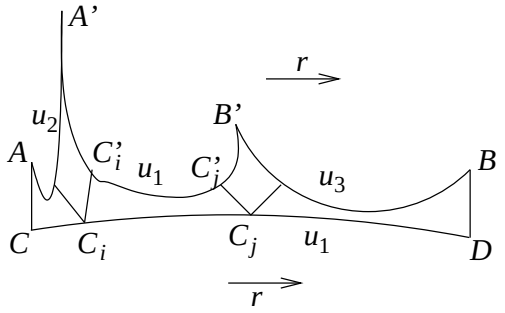


Otherwise, let $L = |u_1|$ and for $0 \leqslant i \leqslant L$, let C_i be the point of CD at distance i from C. Now assign to i a number $\varphi(i)$ between 0 and L as follows: C_i is close to some point C_i' of AB, set $\varphi(C_i) = 0$ if $C_i' \in AA'$, $\varphi(C_i) = L$ if $C_i' \in B'B$, and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

It follows from elementary hyperbolic geometry (approximation of the quadrilateral $CA'B'D$ by a tree) that $\varphi : [0; L] \to [0; L]$ is an increasing function up

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

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



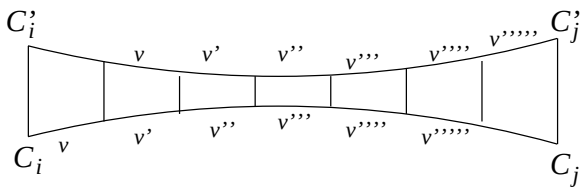
This way, we get a figure in which only the middle part $C_iC_jC_j'C_i'$ of the figure contains two copies of a given piece of u_1. Indeed (from left to right in the figure) the first part contains letters 0 to i of the lower copy of u_1 and no letter of the upper u_1; the second part contains letters 0 to $\varphi(i)$ of the upper u_1 and no letter of the lower u_1; the third part $C_iC_jC_j'C_i'$ 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 $C_iC_jC_j'C_i'$: 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 79.

PROPOSITION 43 – *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 $\varepsilon_2 |u|$.
- There are no double edges except those at commutation translators.
- There are at most $4K/\varepsilon_2$ translators.
- The total number of edges of Γ is at least $|D''| \ell(1 - \varepsilon_1 - 2\varepsilon_2)$.

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 davKd associated to the van Kampen diagram (up to some dividing done for the elimination of doublets; say we add some decoration to the davKd indicating how this was done).

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

6.7 Apparent length

The line of the main argument below is to fulfill the 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 10) there exist words u and v of length at most $2E\log\ell$ such that $xuyv = e$.

DEFINITION 44 (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 45 (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 46 – *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 10.

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 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 35.

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 44) 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 $4K/\varepsilon_2$ 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 $\varepsilon_3\ell.8K/\varepsilon_2$. Setting $\varepsilon_3 = \varepsilon_2^2/8K$ ensures that the total length of these sides is less than $\varepsilon_2\ell$.

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

Throw away all zero-sided translators from the graph Γ. This throws away a total length of at most $\varepsilon_2\ell$, and do not call sides any more the small sides of one-sided translators. Now we have two-sided translators, one-sided translators, commutation translators and boundary translators, all sides of which have length at least $\varepsilon_3^2\ell$. So if ℓ is large enough (depending on ε_3) we can apply the probability evaluations of Axioms 1-4 without trouble.

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

Add a further decoration to Γ (and to the davKd): 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.

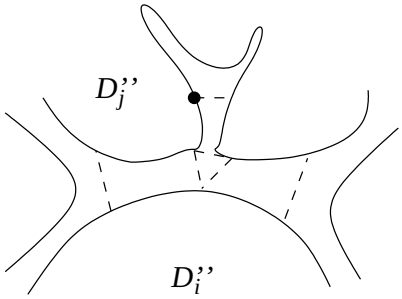
Now say that a random set of relators $r_1, \ldots, r_n$ *fulfills* the conditions of Γ up to letter i if for any internal or commutation translator t which is finished at time i, the corresponding word $w(t)$ is trivial in G; and if, for any half-finished two-sided translator t, the apparent length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

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.)

For a given relator r, there may be some translators having a side made of an initial and final piece of r, so that the side straddles the first letter of r. As we will fill in letters one by one starting with the first ones, we should treat these kind of translators in a different way. The simplest way to treat this little problem is a further cutting of the translators that straddle the beginning of

a word, using Proposition 72, as is best explained by a figure (the thick dot represents the beginning of some relator).



Up to now there are three free variables in our argument: K, the maximal number of new cells in diagrams we consider; and ε_1 and ε_2, which are linked to the way we cut translators to eliminate doublets.

PROPOSITION 47 – *For every density $d < \beta$, for every K, there exists $\varepsilon_1, \varepsilon_2 > 0$ such that, if ℓ is large enough, then, for any davKd $\mathcal{D}$, either $\mathcal{D}$ satisfies a $\frac{\kappa_2}{4}(1 - d/\beta)$-isoperimetric inequality (in the sense of Definition 41), where κ_2 is the constant in Axiom 2, or the probability that $\mathcal{D}$ is fulfillable is less than $(2m)^{-\ell(\beta-d)/4}$.*

Before proceeding to the proof of this proposition, let us show how it implies Proposition 35, via Proposition 42.

If we know that the number of distinct davKd's associated to a van Kampen diagram satisfying the assumptions of Proposition 35 is polynomial in ℓ, then summing the probability evaluation of Proposition 47 on all such davKd's we can conclude: in this case, the probability that there exists a davKd violating the isoperimetric inequality is exponentially small, and so any van Kampen diagram will satisfy an isoperimetric inequality, since any van Kampen diagram satisfying the assumptions of Proposition 35 and 38 has an associated davKd. So we will evaluate the number of davKd's with at most K faces.

But by Proposition 40, the number of possible davKd's is polynomial in ℓ at fixed K. We have to beware we added some extra decoration to the davKd in between: in the elimination of doublets (we made at most K/ε_2 more cuttings, which can be kept track of by as many numbers between 1 and $K\ell$), and when prescribing an apparent length for each internal translator (at most $4K/\varepsilon_2$ numbers between 1 and 4ℓ). So the number of possibilities remains polynomial in ℓ (all other things being fixed).

This proves that the probability that there exists a davKd violating the isoperimetric inequality decreases exponentially with ℓ, hence Proposition 35.

PROOF OF PROPOSITION 47 – Choose some integer K. It is time to fix the parameters ε_1, ε_2. Recall we set $\varepsilon_3 = \varepsilon_2^2/8K$. Also recall that the sides of translators are of length at least $\varepsilon_3^2\ell$, so that we will take ℓ large enough depending on ε_3 (that is, depending on K and on the axioms).

With foresight, let $\varepsilon = \varepsilon_1 + 3\varepsilon_2 + \gamma_4\varepsilon_2/\beta + \varepsilon_3/\kappa_2$ where $\kappa_2, \gamma_4, \beta$ are the constants appearing in the axioms. Choose ε_1 and ε_2 small enough so that $\varepsilon \leqslant (1 - d/\beta)/4$. These choices depend on K, d and G but not on ℓ neither on any diagram.

Let P_i be the probability that *some fixed choice* of n relators $r_1, \dots, r_n \in R$ under our law μ_ℓ fulfills Γ up to letter i. This does not take into account the choice of n relators among the $(2m)^{d\ell}$ relators of the presentation. The quantity P_i depends only on the davKd $\mathcal{D}$ and on the law μ_ℓ of the 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 m_a be the number of times relator a appears 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 $r_1, \dots, r_a$ 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 $f_{i_0}, \dots, f_{i_f}$ 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 $A_a - A_{a-1}$.

LEMMA 48 – *For any davKd $\mathcal{D}$ with at most K faces, for any $1 \leqslant a \leqslant n$ we have*

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

where the constant implied in $o(\ell)$ depends on K but not on the diagram $\mathcal{D}$.

Before proving this lemma, let us finish the proof of Proposition 47.

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. From $(\star)$ we get

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

Compare this to the equation linking dimension and number of edges on page 46 (and recall that here A_a is not the number of edges but the apparent length, which varies the opposite way, and that we want it to be big).

Summing the inequalities above for a from 1 to n gives

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

The number of summands is $n \leqslant K$, so that the constant in $o(\ell)$ is controlled by K again.

At the end of the process, all translators are finished, so by definition 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 (if $\mathcal{D}$ is ever fulfillable) 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 45 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 $\mathcal{D}$ this implies

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

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

So using the lower bound for A_n above we get

$$|\partial\mathcal{D}| \geqslant |\mathcal{D}|\,\ell\,(1-\varepsilon-d/\beta)\,\kappa_2 + \frac{\kappa_2}{\beta}\sum d_a(m_a - m_{a+1}) + o(\ell)$$

Recall we managed to choose $\varepsilon \leqslant (1-d/\beta)/4$. Also take ℓ large enough so that the $o(\ell)$ term is less than $\ell(1-d/\beta)/4$ (such an ℓ depends on K). The inequality above rewrites

$$|\partial\mathcal{D}| \geqslant |\mathcal{D}|\,\ell\,(1-d/\beta)\,\kappa_2/2 + \frac{\kappa_2}{\beta}\sum d_a(m_a - m_{a+1})$$

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

$$|\partial\mathcal{D}| \geqslant \frac{\kappa_2}{2\beta}\,|\mathcal{D}|\,\ell\,(\beta - d + 2\inf d_a/\ell)$$

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

$$|\partial\mathcal{D}| \geqslant \frac{\kappa_2}{4}\,\ell\,|\mathcal{D}|\,(1-d/\beta)$$

as needed.

Or, second, suppose $\inf d_a < -\ell(\beta - d)/4$. This means that the probability that the davKd is fulfillable is less than $(2m)^{-\ell(\beta-d)/4}$.

This proves Proposition 47 assuming Lemma 48. □

PROOF OF LEMMA 48 – 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:

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

(where Δ denotes the difference between before and after filling the subword). This will be the motto of 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. The difference between A_a and A_{a-1} is due to internal translators which are half-finished at time i_0 and are finished at time i_f, to internal translators which are not begun at time i_0 and are half-finished at time i_f, and to boundary translators not begun at time i_0 but finished at time i_f: that is, all internal or boundary translators joined to a letter between i_0 and i_f.

First, consider a two-sided translator t which is not begun at time i_0 and half-finished at time i_f. Let $\Delta_t A_a$ be the contribution of this translator to $A_a - A_{a-1}$, we have $\Delta_t A_a = L(t)$ by definition. Taking notations as above, we have an equality $e = u\delta_1 v\delta_2$ in G. By assumption, to fulfill the conditions imposed by Γ we must have $\mathbb{L}(u) = L(t)$. The word u is a subword of the part a of Γ at play. But Proposition 46 (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 notations as above, with p the first letter of u and q the last one, we have

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

or, taking the log and decomposing u into letters:

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

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

This is the rigorous form of our motto above.

Second, consider an internal translator t which is half-finished at time i_0 and finished at time i_f. Let $\Delta_t A_a$ be the contribution of this translator to $A_a - A_{a-1}$, we have $\Delta_t A = -L(t)$. Taking notations as above, we have an equality $e = u\delta_1 v\delta_2$ in G. By assumption, we have $\mathbb{L}(u) = L(t)$. But the very definition of apparent length (Definition 44) 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 $(2m)^{\gamma_4\varepsilon_2|u|-\beta|u|}$ where γ_4 is some constant. Take $\varepsilon_4 = \gamma_4\varepsilon_2/\beta$, 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!), 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:

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

Recall m_a is 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

$$A_a - A_{a-1} \geqslant m_a \left(\ell(1 - \varepsilon_4 - \varepsilon_5 - \varepsilon_1 - 3\varepsilon_2) + \frac{\log_{2m} \mathrm{P}_{i_f} - \log_{2m} \mathrm{P}_{i_0 - 1}}{\beta} \right) + o(\ell)$$

(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 set $\varepsilon = \varepsilon_1 + 3\varepsilon_2 + \gamma_4 \varepsilon_2/\beta + \varepsilon_3/\kappa_2 = \varepsilon_1 + 3\varepsilon_2 + \varepsilon_4 + \varepsilon_5$, which is exactly what we get here. So Lemma 48 is proven. □

All pending proofs are finished; 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

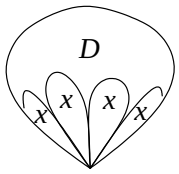
Let $d < \beta$. We will show that the probability that the random quotient is finite decreases exponentially as $\ell \to \infty$.

We know from hyperbolicity of the quotient (Proposition 36) that the probability that there exists a van Kampen diagram of the quotient whose part made of old relators is reduced and which is strongly reduced with respect to G, violating some isoperimetric inequality, is exponentially close to 0.

Imagine that $G/\langle R\rangle$ is finite. Then any element of the quotient is a torsion element. Let x be an element of the quotient, this means that there exists a van Kampen diagram D bordered by x^n for some n.

Now take for x a random word picked under μ_ℓ. We will show that such a random word is very probably not a torsion element in the quotient. Instead of applying the previous section's results to the random quotient of G by R, consider the random quotient of G by $R \cup \{x\}$. Since x is taken at random, $R \cup \{x\}$ is a random set of words, whose density is only slightly bigger than d; this density is $d' = \frac{1}{\ell}\log_{2m}\left((2m)^{d\ell}+1\right)$ which, if ℓ is large enough, is smaller than β if d is.

Now, if $G/\langle R\rangle$ is finite then x is of torsion. Set $N = |R| = (2m)^{d\ell}$. Consider the following family of diagrams. Let D be any abstract 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 34 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 35 and 36) that, in the random quotient $G/\langle R\cup\{x\}\rangle$ at density d', the existence of such a diagram is of probability exponentially close to 0 as ℓ tends to infinity. This ends the proof.

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

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

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

Say (from Proposition 36) that any strongly reduced diagram D of the group $G/\langle R' \rangle$ satisfies an isoperimetric inequality $|\partial D| \geqslant \alpha \ell \, |D''|$ for some positive α, notations as above.

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

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

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

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

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

PROPOSITION 49 – *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 50 – *A random quotient of a hyperbolic group by plain random words for $d < \beta$ is not quasi-isometric to $\mathbb{Z}$.*

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

If $\mathbb{Z}$ is a quotient of the group, then its abelianized is at least $\mathbb{Z}$, which contradicts the previous proposition. If D_∞ is a quotient, note that the abelianized of

D_∞ is $D_2 = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, which is still incompatible with the previous proposition. So we are done if $d > 0$.

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

This ends the proof of Theorem 13.

A Appendix: The local-global principle, or Cartan-Hadamard-Gromov theorem

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

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 51 (P. PAPASOGLU, CF. [PAP], AFTER M. GROMOV) – *Let X be a simplicial complex of dimension 2, simply connected. Suppose that for some integer $K > 0$, any circle S drawn in X whose area lies between $K^2/2$ and $240K^2$ satisfies*

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

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

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

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

termed by Gromov *hyperbolic Cartan-Hadamard theorem* or *local-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 52 (P. PAPASOGLU, CF. [PAP], AFTER M. GROMOV) – *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 53 – *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

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

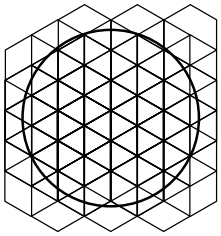
Then any disk D drawn in X, having P, with $A(D) \geqslant K^2$, satisfies

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

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 who come from a disk drawn in X, and those which there exists faces of X that are only partially contained in.

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

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

We will need the following isoperimetric lemmas:

LEMMA 54 – *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 55 – *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 56 – *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 55, we have $L(B') \leqslant 32L(B)$.

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

As a first case, suppose that the cumulated area of all such D which are included in B' is at least one half of the area of B'. Define B'' by amputing B' from all faces D of X which are not totally included in B'. By assumption, we have $A(B') \geqslant A(B'') \geqslant A(B')/2$. And it follows from Lemma 54 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 56, 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. For general references on hyperbolic groups and spaces we refer to [BH], [CDP] or [GH].

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 57 – *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 $2\delta + 1$.*

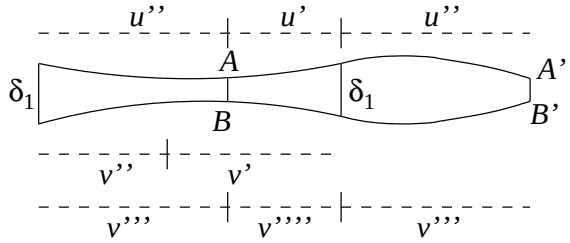
This immediately extends to:

PROPOSITION 58 – *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 57. 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 $v''v'$.



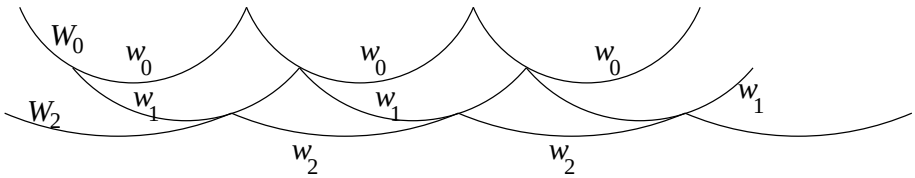
Let A be the endpoint of u''. The point A is $(2\delta + |\delta_1|)$-close to some point B on $v''v'$. Let δ_2 be a path connecting A to B. The point B divides $v''v'$ into two words v''' and v'''', and we have $u = u'u'' = \delta_2 v'''' v''' \delta_2^{-1}$ which ends the proof of the proposition.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

We will need the following

PROPOSITION 59 – *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 $(|w| - |z|)(\delta + 1/2) + 4\delta$.*

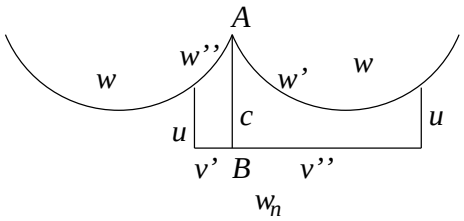
PROOF – Set $w_0 = w$ and construct a sequence w_n of geodesic words by induction. If w_n is cyclically geodesic, stop. If not, then write $w_n = w'_n w''_n$ such that $w''_n w'_n$ is not geodesic. Then set w_{n+1} to a geodesic word equal to $w''_n w'_n$. As length decreases at least by 1 at each step, 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. □

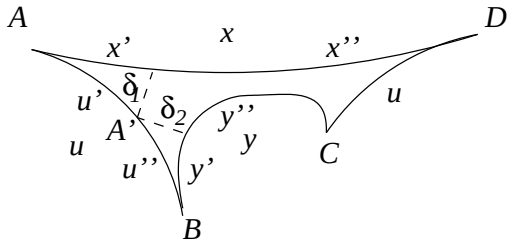
Now, in the spirit of Proposition 57, let $C_c = \max_{x,y} \min\{|u|, x = uyu^{-1}\}$ where the range of the maximum is the set of all couples of conjugate words of length at most $8\delta + 1$. As this set is finite we have $C_c < \infty$. Let $C'_c = C_c + 4\delta^2 + 12\delta + 2$.

PROPOSITION 60 – *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.

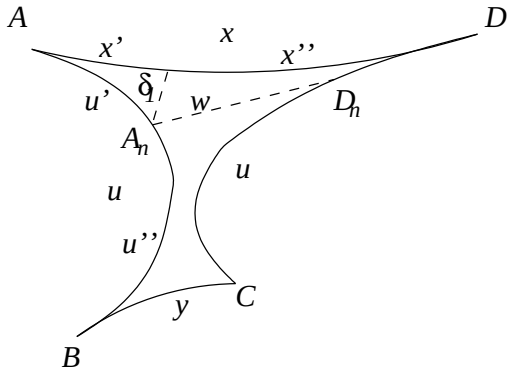


Let A' be the first point on AB which is 2δ-close to BC. Considering the point just before A', we know that A' is $(2\delta + 1)$-close to AD.

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 58. If not, use Proposition 59 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 58. □

COROLLARY 61 – *Let x be any word and y be a conjugate of x of minimal length. Then some cyclic conjugates of x and y are conjugate by a word of length at most $\delta \log_2 |x| + C'_c + 1$.*

PROOF – This is because in a hyperbolic space, a geodesic joining the ends of any curve of length ℓ stays at distance at most $1 + \delta \log_2 \ell$ 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 $1 + \delta \log_2 |x|$. □

B.2 Cyclic subgroups

We will also need the following.

PROPOSITION 62 – *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 63 – *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 60 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 $u^k = u'x^{-1}v^k xu'^{-1}$, 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 64 – *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 $R(\Delta)$.*

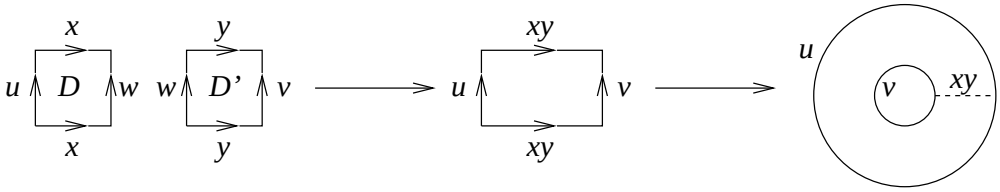
B.3 One-hole diagrams

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.

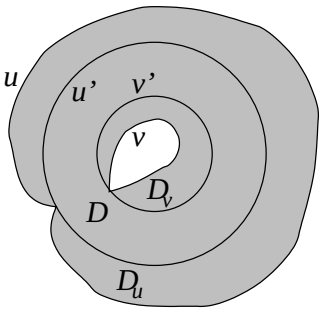
PROPOSITION 65 – *There exists a constant $C' > 0$ such that for any two conjugate words u and v, there exists a one-hole diagram D bordered by u and v, such that $C'|D| \leqslant |u| + |v|$.*

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 60, 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 65. (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 66 – *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 65 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. □

PROPOSITION 67 – *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 68 – *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 ball 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 66, 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 LEMMA 68 – *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 69 – *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 66. 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. □

B.5 Coarsenings of diagrams

If D is a very narrow diagram with holes, then we have an intuitive feeling of which parts of its boundary "face" which. This intuition can be made clear using the approximation of hyperbolic spaces by trees. We now give for this intuition a mathematical setting fitted to our needs.

DEFINITION 70 – *Let $w_1,\dots,w_n$ be n geodesic words in G. A (k,ε)-*matching *of these words is a set of words $w'_1,\dots,w'_k$, some of which may be empty, together with a partition I_1, I_2 of $\{1,\dots,k\}$ and a bijection ψ between I_1 and I_2, such that:*

- *The words $w'_1,\dots,w'_k$ form a partition of the words $w_1,\dots,w_n$.*
- *For all $i \in I_1$, there exist words δ_1 and δ_2 of length at most ε such that $w'_i = \delta_1 w'_{\psi(i)}\delta_2$ in G (we will say that w'_i and $w'_{\psi(i)}$ ε-*match*).*

This means that we cut the words w_i into at most k subwords such that each subword "faces" another one up to ε. Typically ε is of order δ. We have to leave open the possibility that some w_i' are empty since, for example, if one of the w_i's is very short, it could have to be matched with the empty word.

The following proposition is basically equivalent to the approximation of finite hyperbolic spaces by trees.

PROPOSITION 71 – *Let $w_1, \dots, w_n$, for $n \geqslant 2$, be n geodesic words in G such that $w_1 \dots w_n = e$. There exists a $(4n, 4n\delta)$-matching of these words.*

For $n = 3$ this closely resembles the definition of thin triangles.

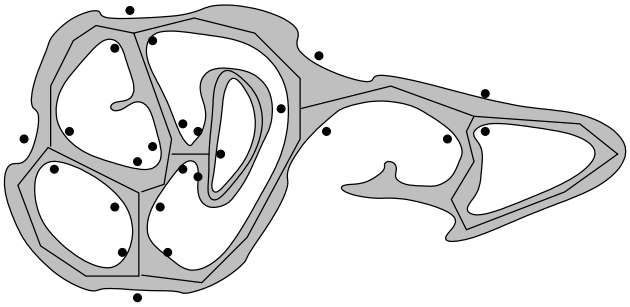
PROOF – Work by induction on n. The result is clear for $n = 2$. Suppose that $n+1$ words $w_1, \dots, w_n, w_{n+1}$ forming a closed piecewise geodesic path in G are given. Let x be a geodesic word equal to $w_n w_{n+1}$. The three geodesic words x, w_n, w_{n+1} form a δ-thin triangle. Let $x = x_1 x_2$ where the endpoint of x_1 lies at distance at most 2δ from both sides w_n and w_{n+1} of the triangle. Now apply the induction assumption to $w_1, \dots, w_{n-1}, x$. This gives a matching involving a partition of the word x into subwords x_i', $i \in I$. At most one of the words x_i' straddles the endpoint of x_1. If some x_i' is included in x_1 or x_2 and is $4n\delta$-matched to w' where w' is a subword of the w_i's, then using thinness of the triangle x, w_n, w_{n+1} we can $(4n+2)\delta$-match w' with a subword of w_n or w_{n+1}. If x_i' straddles the endpoint of x_1, and x_i' is $4n\delta$-matched to w', then we can write $x_i' = x_i'' x_i'''$ where the endpoint of x_i'' is that of x_1, and also write $w' = w'' w'''$ such that w'' $(4n+2)\delta$-matches with x_i'' and w''' matches with x_i'''; using thinness of the triangle x, w_n, w_{n+1} we can $(4n+4)\delta$-match w'' and w''' with subwords of w_n and w_{n+1} respectively. Last, the two parts of w_n and w_{n+1} which are not 2δ-close to x can be matched together. □

Doing the induction more cleverly, one can even obtain a $(4n, 4\delta \log_2 n)$-matching.

We are to apply this construction to diagrams with n holes. In order to symmetrize the role of holes and of the boundary in the following proposition, we view a n-hole diagram as a $(n+1)$-hole diagram embedded in the sphere.

PROPOSITION 72 – *There exists a constant B (depending on G) such that, for any minimal $(n+1)$-hole diagram D embedded in the sphere, there exists a $(8n, Bn \log |D|)$-matching of the boundary words of D. This matching is called the* coarsening *of D.*

The coarsening of D is best visualized as a planar graph as in the following picture (black dots mark the points where we partition the boundary words). The planar graph can be precisely defined but we do not need it.



PROOF – First, using Corollary 61, and the fact already used above that any geodesic joining the endpoints of a curve of length ℓ stays $(\delta \log \ell + 1)$-close to that curve, we can suppose that the boundary words of D are cyclically geodesic: the length ℓ of any boundary word of D is at most $\lambda |D|$ and so a (k, ε)-matching for the cyclically reduced words will give a $(k, \varepsilon+2+C'_c+2\delta \log(\lambda |D|))$-matching of the original words.

Use the corollary of Lemma 68 to cut D into a simply connected diagram D'. The boundary word of D' is made of $n' \leqslant 2n$ pieces $w_1, \ldots, w_{n'}$ partitioning the boundary words of D, with little words $x_1, \ldots, x_{n'}$ of cumulated length at most $B_1 n \log |D|$ in between (for some constant B_1 depending on G). Define y_i to be a geodesic word equal to $w_i x_i$.

Now apply the previous proposition to get a $(4n', 4n'\delta)$-matching of the words $y_1, y_2, \ldots, y_{n'}$. Since the x_i's are of cumulated length at most $B_1 n \log |D|$, this matching defines a $(4n', 4n'\delta + 2\delta + B_1 n \log |D|)$-matching of the w_i's. □

C Appendix: Cases of harmful torsion

Here we show that the assumption of harmless torsion cannot be removed. Examples of hyperbolic groups with harmful torsion include such groups as $(F_m \times \mathbb{Z}/2\mathbb{Z}) \star F_m$ with $m \geqslant 2$, since the $\mathbb{Z}/2\mathbb{Z}$ factor has a centralizer which is a free group of rank m. More precisely:

PROPOSITION 73 – *Theorem 4 does not hold for the hyperbolic group* $(F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$.

PROOF – Consider the two groups $G_1 = (F_m \times \mathbb{Z}/2\mathbb{Z}) \star F_m$ and $G_2 = (F_m \star F_m) \times \mathbb{Z}/2\mathbb{Z}$. In each of these, denote by u a generator for $\mathbb{Z}/2\mathbb{Z}$ and respectively by $a_1, \ldots, a_m$ and $b_1, \ldots, b_m$ a standard set of generators for the first and second factor F_m. Let A_1 and A_2 be the subgroups of G_1 and G_2, respectively, generated by the a_i's, and define B_1 and B_2 similarly.

It is immediate to see that these groups are hyperbolic.

For any group G generated by k elements, let $\lambda(G)$ denote the spectral radius of the random walk on G (w.r.t. the k generators), and let $\theta = 1 + \log_{2k} \lambda$ be the gross cogrowth of G.

The spectral radius for the free group F_k is $\lambda(F_k) = \sqrt{2k-1}/k$. By Lemma 4.1 of [K1], the spectral radius for the group $F_k \times \mathbb{Z}/2\mathbb{Z}$ is equal to $(1+k\lambda(F_k))/(k+1) = (1+\sqrt{2k-1})/(k+1)$.

In particular, the spectral radius of G_2 is $(1+\sqrt{4m-1})/(2m+1)$.

Take $m = 4$. We have $\theta(G_2) = 1 + \log_{4m+2} \lambda(G_2) \approx .788$. In particular, the critical density d^2_{crit} for random quotients of G_2 by plain random words is about $1 - .788 = .212$.

Since G_2 is a quotient of G_1 we have of course $\lambda(G_1) \leqslant \lambda(G_2)$. But Theorem 1 of [K1] states that quotienting a group by (the normal closure of) a non-amenable subgroup strictly increases the spectral radius. The kernel of the quotient map $G_1 \to G_2$ contains the two elements $ub_1u^{-1}{b_1}^{-1}$ and $ub_2u^{-1}{b_2}^{-1}$ which generate a free non-cyclic subgroup in $\mathbb{Z}/2\mathbb{Z} \star B_1$. Hence the kernel is non-amenable.

Thus, we have $\lambda(G_1) < \lambda(G_2)$, so that if Theorem 4 holds for G_1, the critical density d^1_{crit} for random quotients of G_1 satisfies

$$d^1_{\text{crit}} > d^2_{\text{crit}} \approx .212$$

But we are going to prove that random quotients of G_1 are very probably trivial as soon as $d > d^2_{\text{crit}}$.

Let R be a set of randomly chosen words in $u, a_1, \ldots, a_m, b_1, \ldots, b_m$, of length ℓ, at density d. (Note that for the model of random quotients by plain random words, the law of the relators depends only on the generators and not on the initial group.) We now study the random quotient $G_1/\langle R \rangle$ and consider the elements of R as elements of G_1.

Let us compute the probability that one of the relators in R belongs to $C = A_1 \times \mathbb{Z}/2\mathbb{Z} \subset G_1$. The number of words of length ℓ belonging to C is at least $(2m+2)^\ell$, so that the corresponding density is at least $\log_{4m+2}(2m+2) \approx .797$. So there exists a density $d_C \leqslant 1 - .797 = .203$ such that if $d > d_C$, there will very probably be some element of R lying in C. Note that this is below the critical density d^1_{crit} for random quotients of G_1 predicted by Theorem 4 (if it holds). Also, d_C is not 0 since $G_1/\langle C\rangle$ is not amenable.

By the same argument, for $d > d_C$ it is very probable that R contains a relator r of the form $r = xc$ where x is one of the generators of G_1 and c is a word of length $\ell - 1$ with $c \in C$. As the random words are sampled uniformly, when ℓ is big enough this will occur for *each* of the relators x of G_1.

Let H be the random quotient $G/\langle R\rangle$. By definition of C, u commutes with c in G_1, so in H we have

$$uxu^{-1}x^{-1} = uxcu^{-1}c^{-1}x^{-1} = uru^{-1}r^{-1} = e$$

since $r = e$ in H by definition.

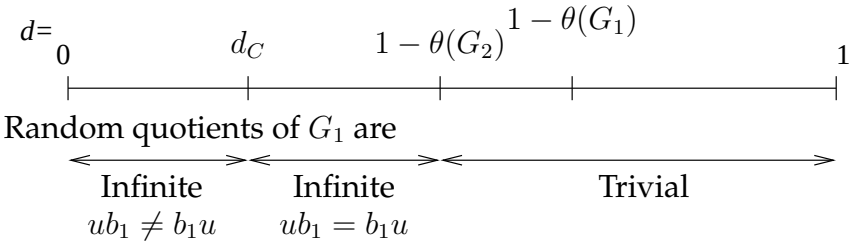
Thus, in H, the generator u commutes with all the generators of G_1. Let $S \subset G_1$ be the set of the commutators of u with these generators, we have

$$H = G_1/\langle R\rangle = G_1/\langle R \cup S\rangle = \left(G_1/\langle S\rangle\right)/\langle R\rangle = G_2/\langle R\rangle$$

But $G_2/\langle R\rangle$ is a random quotient of G_2 (this is because for random quotients by plain words, the law of R is independent on the initial group). In particular, if $d > d^2_{\text{crit}} \approx .212$ this group will very probably be trivial, whereas if Theorem 4 were valid for G_1, the critical value would be $d^1_{\text{crit}} > d^2_{\text{crit}}$.

This ends the proof. □

So random quotients of G_1 behave in a different manner than that of Theorem 4. For densities between 0 and $d_C < .203$ they behave "normally" (in particular, Axiom 4 is satisfied). But for densities between d_C and $.212$, Axiom 4 is not satisfied, and the random quotients behave like random quotients of G_2, and they vanish as soon as $d > .212$, whereas the expected critical density in Theorem 4 would be higher. (The gap between $.203$ and $.212$ can be made larger by taking bigger m.)



The two phases are really different: indeed, a difference can be seen in the ball of radius 2 in the Cayley graph since, in the random quotient, the relation

$ub_1 = b_1 u$ (notation as above) holds in the second phase but not in the first one (since in the first phase, the "ordinary" theory of random quotients holds and in particular, the radius of injectivity grows with ℓ).

More than three phases can probably be arranged, using groups such as

$$(((F_m \times \mathbb{Z}/2\mathbb{Z}) \star F_p) \times \mathbb{Z}/2\mathbb{Z}) \star F_q$$

with different critical densities equal to the densities of the centralizers of the different torsion elements.

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).

[Bow] 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.

[C] J.M. Cohen, *Cogrowth and Amenability of Discrete Groups*, J. Funct. Anal. **48** (1982), 301–309.

[CDP] M. Coornaert, T. Delzant, A. Papadopoulos, *Géométrie et théorie des groupes : les groupes hyperboliques de Gromov*, Lecture Notes in Mathematics *1441*, Springer-Verlag, Berlin (1990).

[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.

[D] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), No. 3, 661–682.

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

[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.

[GK] R.I. Grigorchuk, P.F. Kurchanov, *Some questions of group theory related to geometry*, in *Algebra VII, Combinatorial group theory, applications to geometry*, Encyclopaedia of mathematical sciences **58**, Springer, Berlin (1993), 167–240.

[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, *Growth and cogrowth of generic groups*, this volume.

[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., *Notes on word hyperbolic groups*, in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 3–63.

[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.

Contents

Growth and cogrowth of generic groups

Abstract

We prove that growth and cogrowth of a random quotient of a group stay close to those of this group. In particular, we prove that for all ε, having cogrowth at most $1/2 + \varepsilon$ and growth at least $1 - \varepsilon$ (in base $2m - 1$ with m the number of generators) is a generic property of groups (in the density model of random groups).

In [Ch], Champetier shows that groups defined by a presentation satisfying the small cancellation condition (more precisely, a somewhat weaker assumption), with long enough relators, has a cogrowth arbitrarily close to $1/2$ (see also [Gri]).

We get the same conclusion for *generic* groups in a precise probabilistic meaning: that of the density model of random groups introduced in [Gro1] and defined below. We also show that the growth exponent of a generic group is arbitrarily close to 1 (in base $2m - 1$ where m is the number of generators).

More generally we show similar results for random quotients of torsion-free hyperbolic groups. A generic group is only a random quotient of a free group. For non-free groups there are several variants of the density model of a random quotient: one may quotient by plain words, reduced words, or geodesic words (see the precisions below). We show that for any torsion-free hyperbolic group, a random quotient of it has the same cogrowth and gross cogrowth (when quotiented by plain or reduced words), or the same growth (when quotiented by geodesic words).

As one of our tools we use results about locality of growth and cogrowth in hyperbolic groups, which can be of independent interest.

The primary motivation for these results is the study of generic properties of groups. Nevertheless, besides, these results happen to be especially interesting in contexts where it is necessary to iterate the operation of taking a random quotient, as is the case in the construciton by Gromov of groups whose Cayley graph contains a family of expanders (see [Gro2]). In this last construction, the group is obtained as a limit of an infinite series of quotients, and it is absolutely crucial (and not at all easy) to get a *uniform* control on the cogrowths of the successive quotients. More generally, since the critical density for a random

quotient is controlled by the cogrowth of the initial group (see [Oll]), it is natural to ask what happens to this critical density through a random quotient. The results presented here solve this question.

In this text we constantly refer to the methods of [Ch] and of [Oll], the knowledge of which is thus necessary to understand the proofs presented here.

We refer to [Oll] and to the references therein for our notations about growth, cogrowth and gross cogrowth of groups (see also [GdlH] for a discussion of these notions), and for a general discussion of the density model of random groups.

This text is preliminary and the proofs therein are very sketchy.

We recall the density model for random groups. We consider a random quotient of a torsion-free hyperbolic group $G = \langle a_1, \dots, a_m \mid Q \rangle$ by a set R of randomly uniformly chosen words of length ℓ in the $a_i^{\pm 1}$'s. We have three variants: for a fixed $0 \leqslant d \leqslant 1$ we take either $(2m-1)^{d\ell}$ reduced words, or $(2m)^{d\ell}$ plain words, or $|S_\ell|^{d\ell}$ elements of norm ℓ (actually $\ell \pm 1$ to avoid small scale phenomena) where S_ℓ is the number of elements of norm ℓ in G. We designate these three variants as random quotients by reduced, plain, or geodesic words respectively.

Let η and θ be the cogrowth and gross cogrowth of G w.r.t. this presentation, i.e. the exponent (in base $2m-1$ or $2m$ respectively) of growth of the number of (reduced or plain) words representing the trivial element of G. Let g be the growth exponent (in base $2m$) of G.

It is known (see [Oll]) that if $d < 1 - \eta$, the random quotient $G' = G/\langle R \rangle$ by reduced words is infinite and hyperbolic with probability tending to 1 as $\ell \to \infty$, and similarly for $d < 1 - \theta$ for a random quotient by plain random words, and similarly for $d < 1/2$ for a random quotient by elements of a given norm.

Cogrowth and gross cogrowth cannot stay exactly unchanged under a random quotient since, for example, a free group on m generators is uniquely characterized by its cogrowth or gross cogrowth.

We show here that

THEOREM 1 – *Suppose $d < 1 - \eta$. For any $\varepsilon > 0$, the probability that the cogrowth of a random quotient of G by reduced words is at most $\eta + \varepsilon$ tends to 1 as $\ell \to \infty$.*

Suppose $d < 1 - \theta$. For any $\varepsilon > 0$, the probability that the gross cogrowth of a random quotient of G by plain words is at most $\theta + \varepsilon$ tends to 1 as $\ell \to \infty$.

Suppose $d < 1/2$. For any $\varepsilon > 0$, the probability that the growth of a random quotient of G by geodesic words is at least $g - \varepsilon$ tends to 1 as $\ell \to \infty$.

This theorem implies that cogrowth does not change much in the plain word model and that gross cogrowth does not change much in the reduced word model. Indeed, cogrowth and gross cogrowth are linked by the formula $(2m)^\theta = (2m-1)^\eta + (2m-1)^{1-\eta}$ (cf. [Gri] or [Ch]).

This cannot be interpreted by simply saying that at scales less than ℓ, a random quotient looks like the original group and that the first relations occur at scale ℓ. Indeed, growth and cogrowth are asymptotic invariants, and thus our evaluations take into account the (non-trivial) geometry of the quotient at scale ℓ and further.

The theorem applied to a free group yields the following.

COROLLARY 2 – *For any $\varepsilon > 0$, the property of having cogrowth at most $1/2+\varepsilon$ is generic for the density model of random groups. In particular, the spectral gap of the discrete Laplacian on the Cayley graph is arbitrarily close to that of a free group.*

COROLLARY 3 – *For any $\varepsilon > 0$, the property of having growth at least $1-\varepsilon$ (in base $2m-1$) is generic for the density model of random groups.*

When $d < 1/12$ the group is probably a small cancellation group and so in this case the corollaries follow from a theorem of Grigorchuk and Champetier ([Gri], [Ch]) in the case of cogrowth or from a theorem of Shukhov ([Shu]) in the case of growth. For greater d, the second corollary provides new examples of groups with nearly-maximal growth, as is asked for in [GdlH].

1 Locality of cogrowth in hyperbolic groups

Here $G = \langle a_1, \dots, a_m \mid R \rangle$ 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 λ be the maximal length of a relation in R.

Let D be a van Kampen diagram w.r.t. this presentation and define the *area* of D to be

$$\mathcal{A}(D) = \sum_{f \text{ face of } D} |\partial f|$$

Suppose that G satisfies the following isoperimetric inequality: for some $C > 0$, any minimal van Kampen diagram D over this presentation satisfies

$$|\partial D| \geqslant C\mathcal{A}(D)$$

Of course $C \leqslant 1$. Set $\alpha = 1/\log(1/(1-C))$.

It can be shown that the hyperbolicity constant of such a group is at most $\mathrm{Cst}.\lambda/\alpha$ where Cst is a universal constant.

LEMMA 4 – *Let D be a minimal van Kampen diagram. Then D is $\alpha\lambda \log \mathcal{A}(D)$-narrow.*

Note that this is less than $\alpha\lambda(\log|\partial D| + \log 1/C)$.

PROOF – Easy: remove cells adjacent to the boundary and induce backwards. □

Working the same way, we can prove that any van Kampen diagram is the union of an $\alpha\lambda\log(\mathcal{A}(D)/\lambda)$-narrow annular diagram glued along the boundary of a diagram of area at most λ.

LEMMA 5 – *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 $2\alpha\lambda\log\mathcal{A}(D)$ such that each of D' and D'' contains at least one quarter of the boundary of D.*

PROOF – 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 $\alpha\lambda\log\mathcal{A}(D)$-narrow, there exists a path of length at most $2\alpha\lambda\log\mathcal{A}(D)$ joining either a point of AB to a point of CD or a point of AD to a point of BC. □

Using the variant above with a $\alpha\lambda\log(\mathcal{A}(D)/\lambda)$-narrow annular diagram, we get a decomposition into three pieces: one of area at most λ, and two pieces having at least one quarter of the boundary each, and cut along a path of length at most $\lambda + 2\alpha\lambda\log(\mathcal{A}(D)/\lambda)$.

COROLLARY 6 – *Up to parity problems*

$$|W_\ell| \leqslant \sum_{\ell/4\leqslant\ell'\leqslant 3\ell/4} \left|W_{\ell'+2\alpha\lambda(\log\ell+\log 1/C)}\right| \left|W_{\ell-\ell'+2\alpha\lambda(\log\ell+\log 1/C)}\right|$$

and

$$\begin{aligned}|W_\ell| &\leqslant |W_\lambda| \sum_{\ell/4\leqslant\ell'\leqslant 3\ell/4} \left|W_{\ell'+2\alpha\lambda(\log(\ell/\lambda)+\log 1/C)+\lambda}\right| \left|W_{\ell-\ell'+2\alpha\lambda(\log(\ell/\lambda)+\log 1/C)+\lambda}\right| \\ &\leqslant \frac{\ell}{\lambda}|W_\lambda| \max_{\ell/4\leqslant\ell'\leqslant 3\ell/4} \left|W_{\ell'+2\alpha\lambda(\log(\ell/\lambda)+\log 1/C)+3\lambda}\right| \left|W_{\ell-\ell'+2\alpha\lambda(\log(\ell/\lambda)+\log 1/C)+3\lambda}\right|\end{aligned}$$

(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.)

PROPOSITION 7 – *Suppose that, for some $A > 1$, for any $A\lambda/4 \leqslant \ell \leqslant A\lambda$ one has*

$$|W_\ell| \leqslant (2m-1)^{\eta\ell}$$

Then for any $\ell \geqslant A\lambda/4$,

$$|W_\ell| \leqslant (2m-1)^{(\eta+o(1)_{A\to\infty})\ell}$$

where the constant implied in $o(1)$ depends only on C.

PROOF – First, choose ℓ between $A\lambda$ and $4A\lambda/3$. By Corollary 6, we have

$$|W_\ell| \leqslant (2m-1)^{\eta(\ell+4\alpha\lambda(\log(\ell/\lambda)+\log 1/C)+7\lambda+\log_{2m}(\ell/\lambda))}$$

Let B be such that for any $L \geqslant B\lambda$, one has

$$(2m-1)^{\eta(-L+4\alpha\lambda(\log(L/\lambda)+\log 1/C)+7\lambda+\log_{2m}(L/\lambda))} \leqslant 1$$

Since everything is homogeneous in L/λ, such a B depends only on C. Suppose that $A > B$. Then

$$|W_\ell| \leqslant (2m-1)^{\eta(\ell+B\lambda)} \leqslant (2m-1)^{\eta\ell(1+B/A)}$$

We have just shown that if $|W_\ell| \leqslant (2m-1)^{\eta\ell}$ for $\ell \leqslant A\lambda$, then $|W_\ell| \leqslant (2m-1)^{\eta\ell(1+B/A)}$ for $\ell \leqslant (4A/3)\lambda$. Thus, iterating the process shows that for $\ell \leqslant (4/3)^k A\lambda$ we have

$$|W_\ell| \leqslant (2m-1)^{\eta\ell \prod_{0\leqslant i<k}\left(1+\frac{B}{A}\left(\frac{3}{4}\right)^i\right)}$$

and we are done as the product $\prod_i \left(1+\frac{B}{A}\left(\frac{3}{4}\right)^i\right)$ converges. □

2 Application to random groups: the free case

Here we first treat the case when the initial group G is the free group F_m on m generators.

2.1 Fulfilling of diagrams

We show elsewhere (see [Oll]) that for any K fixed in advance, for any reduced van Kampen diagram D of the quotient with at most K faces, filled by n different relators from R (say relator i is used m_i times in D and that the m_i's are chosen non-increasing), one can associate numbers d_i such that the probability that this diagram is fulfilled by the random relators is less than $(2m-1)^{\inf d_i}$ and the diagram satisfies the isoperimetric inequality

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

As by definition we have $m_n \geqslant 1$ and $m_{n+1} = 0$, and as for any $\varepsilon > 0$ we can suppose that $\inf d_i \geqslant -\varepsilon\ell$ (taking ℓ large enough for the probability $(2m-1)^{-\varepsilon\ell}$ to be small), we get that (with $\varepsilon = (1-2d)/4$)

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

Now, $(2m-1)^{d_n}$ is not only an upper bound of the probability that there exists n relators in R fulfilling the diagram. More precisely, $(2m-1)^{d_n - nd\ell}$ is the probability that a given set of n relators can fulfill the diagram. By definition of the model there are $(2m-1)^{nd\ell}$ such n-tuples of relators. So the expected number S of n-tuples of relators fulfilling the diagram is $(2m-1)^{d_n}$. (The probabilities that two n-tuples fulfill the diagram are independent only when the n-tuples are disjoint, but expectancy is linear anyway.)

By Markov's inequality, the probability that $S \geqslant (2m-1)^{\varepsilon'\ell}(2m-1)^{d_n}$ is less than $(2m-1)^{-\varepsilon'\ell}$.

Thus, with probability exponentially close to 1 as $\ell \to \infty$, we can suppose that a given (hence any, since the number of decorated diagrams with less than K faces grows subexponentially) diagram can be filled in at most $(2m-1)^{\varepsilon'\ell}(2m-1)^{d_n}$ different ways by relators of R.

But for any diagram we have that

$$d_n \leqslant \frac{1}{2}\left(|\partial D| - (1-2d)\ell\,|D|/2\right)$$

or as $|D| \geqslant 1$

$$d_n + \varepsilon'\ell \leqslant \frac{1}{2}\left(|\partial D| - (1-2d-\varepsilon')\ell\,|D|/2\right)$$

so if we take $\varepsilon' < 1-2d$, this is at most $|\partial D|/2$.

The conclusion is that for each K, if ℓ is large enough, for each van Kampen diagram with at most K faces, the number of ways to fulfill it with relators of R is less than

$$(2m-1)^{\eta'|\partial D|}$$

with $\eta' \leqslant 1/2$. So what?

2.2 Evaluation of the cogrowth

Consider a reduced word w of length ℓ in the generators $a_i^{\pm 1}$, representing e. This word is the boundary word of some van Kampen diagram D. It may be the case that D has filaments i.e. some edges of D do not lie on the boundary of a face.

We can suppose (up to exponentially small probability in ℓ) that any diagram satisfies the inequality

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

where C depends only on the density d (basically $C = 1-2d$ divided by the constants appearing in the Cartan-Hadamard-Gromov theorem) and not on ℓ.

By Proposition 7 it is enough to suppose that $|w| \leqslant A\ell$ for some (large) A which the quality of the final result will depend on. As $|w| = |\partial D| \geqslant C\ell\,|D|$ we have $|D| \leqslant A/C$, which is bounded independently of ℓ. So take $K = A/C$ and take ℓ large enough so that any diagram with at most K faces can be fulfilled in at most $(2m-1)^{\eta'|\partial D|}$ ways by relators of R, with $\eta' \leqslant 1/2$.

The word w is determined by choosing a relator in R for each face of D, and a reduced word to put on each filament.

Decompose D into filaments and connected subdiagrams D_i such that D_i has no filaments. Now w is determined by the following data: a set of relators fulfilling each D_i, a set of reduced words to put on the filaments and the combinatorial information for how to choose the boundary lengths of the D_i's and connect the D_i's using the filaments. This situation is precisely the one analyzed in [Ch]. It is shown there 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 is less than

$$\frac{|w|}{L}\,|w|\,(eL)^{2|w|/L}(2eL)^{|w|/L}(3eL)^{2|w|/L}$$

whereas the number of choices for the graphs D_i's is obviously less than $N(K)^{|w|/L}$ where $N(K)$ is the (finite!) number of planar graphs with at most K faces.

Observe that for L large enough this behaves like $(2m-1)^{|w|O(\log L/L)}$. But here for any D_i we have $|\partial D_i| \geqslant C\ell\,|D_i| \geqslant C\ell$ hence we can take $L = C\ell$.

Now the number of ways to fill the D_i's is at most $\prod(2m-1)^{\eta'|\partial D_i|} = (2m-1)^{\eta'\sum|\partial D_i|}$. The total length of the filaments is $(|w| - \sum|\partial D_i|)/2$ (each edge of a filament counts twice in the boundary), thus the number of ways to fill in the filaments is trivially at most $(2m-1)^{\frac{1}{2}(|w|-\sum|\partial D_i|)}$.

So the total number of possibilities is

$$(2m-1)^{|w|O(\log \ell/\ell)+\frac{1}{2}(|w|-\sum|\partial D_i|)+\eta'\sum|\partial D_i|}$$

and since $\eta' \leqslant 1/2$, for any ε this is at most $(2m-1)^{(1/2+\varepsilon)|w|}$.

3 The non-free case

Here we work in a random quotient of an arbitrary hyperbolic group G, either by random plain words or random reduced words. Let us work in the case of a random quotient by plain random words and study the gross cogrowth θ (the case of cogrowth for a quotient by reduced words is similar). Set $\beta = 1-\theta$.

Keep the notations of [Oll]: for a vK diagram D, let D'' be the subdiagram made of cells bearing new relators. Suppose that $|D''| \leqslant K$ where K is a constant to be chosen later. By considering the coarsening as defined in [Oll], this subdiagram can be seen as a van Kampen diagram at scale ℓ, with each edge representing a $E\log\ell$-narrow strip of cells from the presentation of G (for some constant E depending on K and G).

Let D'' be a van Kampen diagram at scale ℓ, involving n different relators r_i with multiplicities m_i (and choose the order of enumeration such that the m_i's are non-increasing). This diagram is decorated by imposing the apparent lengths of all edges (the number of possible decorations is subexponential in ℓ). Let d_i be the log-probability that there exists an i-tuple of relators of R fulfilling the conditions imposed by D'' up to the i-th relator. Let d'_i be the log-probability (beware $d_i \leqslant 0$) that a fixed choice of i relators fulfills these conditions. Of course $d_i \leqslant d'_i + id\ell$. By definition, we can suppose that $d_i \geqslant -\varepsilon\ell$ (otherwise the probability that the diagram can be fulfilled is exponentially small) for all i. Hence $d'_i + id\ell \geqslant -\varepsilon\ell$.

One can show as above that with overwhelming probability, the number of n-tuples of relators fulfilling the diagram is at most $(2m)^{d'_n+nd\ell}$.

Let A be the sum of the apparent lengths of the boundary edges of D''. Summing equation $(\star)$ (section 6.8) of [Oll] we get

$$A \geqslant |D''|\,\ell\,(1 - o(\ell)/\ell) + \frac{1}{\beta}\sum m_i(d'_i - d'_{i-1})$$

hence, neglecting the $o(\ell)$ term,

$$\begin{aligned}
A &\geqslant |D''|\,\ell + \frac{1}{\beta}\sum d'_i(m_i - m_{i+1}) \\
&= |D''|\,\ell + \frac{1}{\beta}\sum (d'_i + id\ell + \varepsilon\ell)(m_i - m_{i+1}) - \frac{1}{\beta}\sum (id\ell + \varepsilon\ell)(m_i - m_{i+1}) \\
&= |D''|\,\ell + \frac{1}{\beta}\sum (d'_i + id\ell + \varepsilon\ell)(m_i - m_{i+1}) - \frac{d\ell}{\beta}\sum m_i - \frac{\varepsilon\ell}{\beta}m_1 \\
&\geqslant |D''|\,\ell + \frac{d'_n + nd\ell + \varepsilon\ell}{\beta}m_n - \frac{d\ell + \varepsilon\ell}{\beta}\sum m_i \\
&\quad \text{as } d'_i + id\ell + \varepsilon\ell \geqslant 0 \text{ and } m_i - m_{i+1} \geqslant 0 \\
&\geqslant |D''|\,\ell\left(1 - \frac{d+\varepsilon}{\beta}\right) + \frac{1}{\beta}(d'_n + nd\ell)
\end{aligned}$$

as by definition $\sum m_i = |D''|$ and $m_n \geqslant 1$.

So $d'_n + nd\ell \leqslant \beta A - |D''|\,\ell(\beta - d - \varepsilon)$, hence the number of choices for the relators fulfilling the diagram is at most

$$(2m)^{\beta A - |D''|\ell(\beta - d - \varepsilon)}$$

Suppose here that D'' is connected (up to the small strips of thickness $E\log\ell$), that is, D'' has only one boundary component.

Choosing the relators is not enough to determine the boundary word y of the diagram D, since for a given boundary word x of D'' there are many words y which are (quasi-)equal to x in G (this corresponds to the $E\log\ell$-narrow strip of old relators surrounding D'' in D). But the apparent length of x is A (with an $o(\ell)$ approximation controlled by K, since we sum up all the apparent lengths of the boundary edges of D''). The probability that a random word y of length $|\partial D|$ is (quasi-)equal to x in G is thus by definition at most $(2m)^{-\beta(|\partial D| + A)}$, so the number of such words y, for a given choice of x, is at most

$$(2m)^{(1-\beta)|\partial D| - \beta A}$$

and the total number of possibilities (including the choice of relators) is at most

$$(2m)^{(1-\beta)|\partial D| - \beta A + \beta A - |D''|\ell(\beta - d - \varepsilon)} \leqslant (2m)^{(1-\beta)|\partial D|}$$

with $1 - \beta = \theta$.

To reach the conclusion, reason as above in the free case: take a diagram D (maybe with filaments) and decompose it into filaments and parts without

filaments. For this to work two things are important: first, that we are allowed to use the local-global principle for cogrowth, and second, that each component of D without filaments has a large enough length (compare the L above).

We know from [Oll] that we can suppose that any diagram D satisfies

$$|\partial D| \geqslant \kappa\ell\,|D''| + \kappa'\,|D \setminus D''|$$

where κ is a constant depending only on G and d and κ' is (arbitrarily close to) the original isoperimetric constant of G. So by Corollary 7 it is enough to compute the cogrowth on words of length less than $A\ell$ for some constant A depending on G and d but not on ℓ. By the isoperimetric inequality above, such a word is the boundary word of a diagram D with $|D''| \leqslant A\ell/\kappa$, so we can choose the constant K above to be A/κ, independently on ℓ. This is what guarantees that we can apply the local-global principle.

Then, take the diagram D and decompose it into parts D_i linked by filaments (filaments are $2E\log\ell$-narrow parts) such that D_i'' is connected. As each D_i has boundary length at most $|\partial D_i| \geqslant \kappa\ell\,|D_i''| \geqslant \kappa\ell$, the combinatorics of the parts and filaments linking them is controlled as above (with $L = \kappa\ell$ in the evaluation) and behaves like $(2m)^{|\partial D| O(\log\ell/\ell)}$.

To determine the boundary word of D it is enough to choose a boundary word for each ∂D_i and a boundary word for each filament. The number of boundary words for the filaments is controlled by the cogrowth of the original group (up to a subexponential term resulting from gluings of size $E\log\ell$ of the filaments on the D_i's) and behaves like $(2m)^{\theta(|\partial D| - \sum|\partial D_i|)}$. We showed above that the number of choices for each component D_i is at most $(2m)^{\theta|\partial D_i|}$ hence the evaluation.

4 The case of growth

4.1 Locality of growth in hyperbolic groups

Let S_ℓ denote the set of elements of norm ℓ in the hyperbolic group G and let B_ℓ denote the elements of norm at most ℓ.

Consider also, for homogeneity reasons, the annulus $S_{\ell,a} = B_\ell \setminus B_{\ell-a}$.

Everything here is up to parity problems on the lengths.

Let $(,)$ denote the Gromov product in G (with origin at e).

PROPOSITION 8 – *Suppose that for some g, for some $\ell_0 \geqslant 4\delta + 16/g$ and $\ell_1 \geqslant A\ell_0$ we have*

$$|B_{\ell_0}| \leqslant (2m)^{1.2g\ell_0}$$

and

$$|B_{\ell_1}| \geqslant (2m)^{g\ell_1}$$

Then the growth of G is at least $g(1 - 300/A)$.

Note that the occurrence of $1/g$ in the scale upon which the proposition is true is natural: indeed, an assumption such as $|B_\ell| \geqslant (2m)^{g\ell}$ 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.

PROOF –

PROPOSITION 9 – *Let $g \in B_\ell$. The number of elements g' in S_ℓ or B_ℓ such that $(g, g') \geqslant a$ is at most $|B_{\ell-a+\delta}|$.*

PROOF – Easy: such elements are at distance at most $\ell - a$ (up to δ) of the point at distance a on a geodesic joining e to g. □

COROLLARY 10 – *Let $g \in S_{\ell,a}$. The number of elements g' in $S_{\ell,a/2}$ such that $\|gg'\| \geqslant 2\ell - 2a$ is at least $\left|S_{\ell,a/2}\right| - \left|B_{\ell-a/4+\delta}\right| \geqslant |B_\ell| - 2\left|B_{\ell-a/4+\delta}\right|$.*

PROOF – If $\|g\| \geqslant \ell - a$, $\|g'\| \geqslant \ell - a/2$ and $(g, g') \geqslant a/4$, then $\|gg'\| \geqslant 2\ell - a - a/2 - a/2$. □

PROPOSITION 11 – *Let $x \in S_{2\ell,2a}$. The number of couples (g, g') in $S_{\ell,a} \times S_{\ell,a/2}$ such that $x = gg'$ is at most $|B_{4a+2\delta}|$.*

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 $4a + 2\delta$-close to h (and then g' is determined). □

COROLLARY 12 –

$$|S_{2\ell,2a}| \geqslant \frac{1}{|B_{4a+2\delta}|} |S_{\ell,a}| \left(\left|S_{\ell,a/2}\right| - \left|B_{\ell-a/4+\delta}\right|\right)$$

and so

$$|B_{2\ell}| \geqslant \frac{1}{|B_{4a+2\delta}|} \left(|B_\ell| - 2\left|B_{\ell-a/4+\delta}\right|\right)^2$$

LEMMA 13 – *Suppose that for some g, for some ℓ_0 and $\ell_1 \geqslant 200\ell_0$ we have $|B_{\ell_0}| \leqslant (2m)^{1.2g\ell_0}$ and $|B_{\ell_1}| \geqslant (2m)^{g\ell_1}$. Let $a \leqslant \ell_0$. There exists $0.52\ell_1 \leqslant \ell \leqslant \ell_1$ such that*

$$|B_\ell| \geqslant (2m)^{g\ell}$$

and

$$|B_\ell| \geqslant (2m)^{ga/2} |B_{\ell-a}|$$

PROOF OF THE LEMMA – First, note that by subadditivity, the inequality $|B_{\ell_0}| \leqslant (2m)^{1.2g\ell_0}$ implies that for any ℓ, writing $\ell + r = k\ell_0$ ($k \in \mathbb{N}, 0 \leqslant r < \ell_0$) we have $|B_{\ell_0}| \leqslant (2m)^{1.2kg\ell_0}$. Especially for $\ell \geqslant 100\ell_0$ we have $k\ell_0/\ell \leqslant 101/100$ and so $|B_\ell| \geqslant (2m)^{1.3g\ell}$. In particular, if $\ell_1 \geqslant 200\ell_0$ then $|B_{0.52\ell_1}| \leqslant (2m)^{0.68g\ell_1}$.

Suppose that for all $0.52\ell \leqslant \ell \leqslant \ell_1$ with $\ell = \ell_1 - ka$ ($k \in \mathbb{N}$) we have $|B_\ell| < (2m)^{ga/2} |B_{\ell-a}|$. Write $\ell_1 - 0.52\ell_1 + r = qa$ with $0 \leqslant r < a$. Then we get

$$\begin{aligned} |B_{\ell_1}| &< (2m)^{ga/2} |B_{\ell_1-a}| < (2m)^{ga} |B_{\ell_1-2a}| < \cdots \\ &< (2m)^{gqa/2} |B_{0.52\ell_1-r}| \leqslant (2m)^{g(\ell_1-0.52\ell_1)/2+ga/2} |B_{0.52\ell_1}| \\ &\leqslant (2m)^{g(0.48\ell_1)/2+g\ell_1/400+0.68g\ell_1} < (2m)^{g\ell_1} \end{aligned}$$

hence a contradiction.

So take the largest $\ell \leqslant \ell_1$ satisfying $|B_\ell| \geqslant (2m)^{ga/2} |B_{\ell-a}|$. Since for $\ell \leqslant \ell' \leqslant \ell_1$ we have $|B_{\ell'}| \leqslant (2m)^{ga/2} |B_{\ell'-a}|$ we get $|B_{\ell_1}| \leqslant (2m)^{g(\ell_1-\ell)/2} |B_\ell|$ hence the result using $|B_{\ell_1}| \geqslant (2m)^{g\ell_1}$. □

LEMMA 14 – *Suppose that for some g, for some $\ell_0 \geqslant 4\delta + 16/g$ and $\ell_1 \geqslant A\ell_0$ we have $|B_{\ell_0}| \leqslant (2m)^{1.2g\ell_0}$ and $|B_{\ell_1}| \geqslant (2m)^{g\ell_1}$. Then there exists $\ell_2 \geqslant 1.04\ell_1$ such that*

$$|B_{\ell_2}| \geqslant (2m)^{g\ell_2(1-11/A)}$$

PROOF OF THE LEMMA – Consider the ℓ provided by Lemma 13, and take $a = \ell_0$. As $|B_\ell| \geqslant (2m)^{ga/2} |B_{\ell-a}|$ we have, by Corollary 12

$$|B_{2\ell}| \geqslant \frac{1}{|B_{4\ell_0+2\delta}|} |B_\ell|^2 \left(1 - 2(2m)^{-g(\ell_0/4-\delta)/2}\right)^2$$

which, if $\ell_0 \geqslant 4\delta + \frac{16}{g}$, is at least

$$\frac{1}{4\,|B_{4\ell_0+2\delta}|} |B_\ell|^2$$

We have $|B_{4\ell_0+2\delta}| \leqslant |B_{5\ell_0}| \leqslant |B_{\ell_0}|^5$ by subadditivity. So by assumption

$$|B_{2\ell}| \geqslant \frac{1}{4\,|B_{\ell_0}|^5} |B_\ell|^2 \geqslant (2m)^{2g\ell-10g\ell_0-2} \geqslant (2m)^{2g\ell(1-11/A)}$$

□

Now the proposition is clear: start from ℓ_1 and construct by induction a sequence ℓ_i with $\ell_{i+1} \geqslant 1.04\ell_i$ as in the lemma; thus

$$|B_{\ell_i}| \geqslant (2m)^{g\ell_i \prod_{k=1}^{i-1}(1-11/(A.1.04^k))}$$

and note that the infinite product converges to a value that is arbitrarily close to 1 when A is taken large enough. (The estimate $1 - 300/A$ is easy.) □

4.2 Growth of random quotients

A possible way to show that the growth of a quotient of a group G is the same as the growth of G is as follows: There are roughly $(2m)^{gL}$ elements of norm L in G. Some of these elements are identified in G/R. Let N be the number of equalities of the form $x = y$ which hold in G/R but did not hold in G, where x and y are two distinct elements of norm L in G. Each such equality decreases the number of words of length L by at most 1, hence the number of elements of norm L in G/R is at least $(2m)^{gL} - N$. So if we can show that $N \leqslant (2m)^{g'L}$ with $g' < g$, we are done.

We showed above (Proposition 8) that growth in a hyperbolic group is local (at scale δ). So in a random quotient where $\delta \approx \ell$ it is enough to work with words of length at most $A\ell$ for some constant A depending on the g we want to obtain. Take $L \leqslant A\ell$.

In this paragraph density is expressed in base $2m$ instead of base $(2m)^g$, so that the critical density is $g/2$ rather than $1/2$.

Let D be a (reduced) van Kampen diagram in the quotient with boundary word xy^{-1} where x and y are geodesic words of length L in the group G. This boundary word is geodesic except perhaps at the two junctions between x and y, but these can be cut and treated in G (this means that x and y are already close in G), so we don't consider this problem.

Define D'' as above. As we consider relations which did not already hold in G we can suppose that D'' is non-empty.

The geodesic model of random quotient satisfies the same properties as the random quotient by plain words, with $\beta = g/2$ instead of $\beta = 1 - \theta$. In this model apparent length is just the usual length. As the boundary word of D is geodesic we have, up to $o(\ell)$, that $|\partial D| = |\partial D''|$. So in this model we have (up to $o(\ell)$)

$$|\partial D| = |\partial D''| = 2L$$

and as above

$$d'_n + nd\ell \leqslant \beta 2L - |D''|\,\ell(\beta - d - \varepsilon)$$

and as we trivially have $|\partial D''| \leqslant |D''|\,\ell$ we get

$$d'_n + nd\ell \leqslant \beta 2L - |\partial D''|\,(\beta - d - \varepsilon)$$

As $|\partial D| = |\partial D''| = A$, we get, choosing $\varepsilon = (\beta - d)/2$

$$d'_n + nd\ell \leqslant |\partial D|\,(\beta/2 + d/2)$$

So the number of possible ways to fill a diagram of boundary length $2L$ expressing an equality between two words of length L in the quotient, is at most $(2m)^{d'_n + nd\ell} \leqslant (2m)^{L(\beta+d)}$ and we are done as $\beta = g/2$ and as $d < \beta$.

(Note that the above only applies when D'' is non-empty: otherwise, there are of course roughly $(2m)^{gL}$ diagrams expressing the equality $x = x$ in G, but none of them appear as a reduced diagram having at least one new relator.)

We have to sum this probabilistic evaluation on all diagrams. By locality it is enough to consider $L \leqslant A\ell$, and so $|D''| \leqslant 2A$, hence a bounded summation.

References

[Ch] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), No. 5, 438–444.

[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.

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

[Gro2] M. Gromov, *Random Walk in Random Groups*, Geom. Funct. Anal. **13** (2003), No. 1, 73–146.

[Oll] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, this volume.

[Shu] A.G. Shukhov, *On the dependence of the growth exponent on the length of the defining relation*, Math. Notes **65** (1999), No. 3-4, 510–515.

Contents

On a small cancellation theorem of Gromov

Abstract

We give a combinatorial proof of a theorem of Gromov, which extends the theory of small cancellation to group presentations arising from labelled graphs.

This short paper contains a combinatorial proof of a small cancellation theorem stated by M. Gromov in [G2], which is a generalization of ordinary small cancellation. It deals with presentations of groups obtained by reading the words on the cycles of some graph each edge of which is labelled by a generator of the group (see statement below). Our aim is to complete the six-line-long proof given in [G2] (which invokes geometric arguments).

The theorem extends the usual conclusions of small cancellation theory (see [GH] or [LS]) to much more general situations. For example, ordinary small cancellation theory cannot deal with such 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.

1 Statement and discussion

Let S be a finite set which is the disjoint union of two sets S' and S'', with a bijection from S' to S'' called *being inverse*. 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 complex* is a finite unoriented cellular 2-complex in which each oriented edge bears a letter, such that opposite edges bear inverse letters. (Each unoriented edge is considered as a couple of two oriented edges.) Thus each face defines a word (up to inversion and cyclic permutation) read on its boundary. 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 boundary label — this amounts to considering maps between the corresponding oriented complexes).

A *labelled graph* is a 1-dimensional labelled complex.

A labelled complex 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 complex is a word which has two different immersions in the labelled complex. (An immersion is a locally injective map of labelled complexes. Two immersions are considered different if they are different as maps of 2-complexes.) This is analogue to the traditionnal 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.

We are now in a position to state the theorem.

THEOREM 1 (M. GROMOV) – *Let Γ be a reduced 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 (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 equal to e in G, some cyclic permutation of w contains a subword of a word read on a circle immersed in Γ, of length at least $(1 - 3\Lambda/g) > \frac{1}{2}$ 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. But it is clear from the proof below that the assumption in the theorem can be replaced by the following slightly weaker one: for each piece, its length is less than $1/6$ the minimal length of the cycles of the graph on which the piece appears. With this assumption, the theorem reduces to ordinary small cancellation when the graph is a disjoint union of circles.

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 [G2] in the form we give but using a much more abstract and more powerful formalism of "rotation families of groups". 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 (instead of $\lambda < 1/6$), using geometric rather than combinatorial tools, can be found in [G1].

In [G2], this theorem is applied to a random labelling (or rather a variant of this theorem given below, in which reducedness is replaced with quasi-geodesicity). It is not difficult, using for example the techniques described in [O], to check that a random labelling satisfies the small cancellation and quasi-geodesicity assumptions.

2 Proof

We now give some more definitions which are useful for the proof.

A *tile* is a planar labelled complex with only one face and no other edge than the boundary edges of this face (but not necessarily simply connected). We do not fix the embedding in the plane. 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 has a non-trivial automorphism. 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 torsion because with this convention asphericity of a presentation implies torsion-freeness and asphericity of the Cayley complex. Note that our definition of asphericity is thus slightly stronger than the one in [LS].

The *length* of a tile is the length of its boundary.

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 (which is constructed by cutting 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.

A *tile of a labelled complex* is the tile associated to any of its 2-faces.

A *piece* with respect to a set of tiles is a word which has immersions in the boundary two different tiles, or two distinct immersions in the boundary of one tile.

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).

A puzzle is said to be *minimal* if it has the minimal number of tiles for a given 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 boundary 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 boundary word of these faces (in other words, there is no trivial gluing). This corresponds to reduced van Kampen diagrams (see [LS]). (Incidentally, a reduced puzzle is van Kampen-reduced, though the converse is not necessarily true.)

PROOF OF THE THEOREM – Let Γ be a reduced labelled graph. Γ defines a presentation R by taking all the words read along its cycles. 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.

Note that the assumption on pieces implies that no two distinct cycles of Γ bear the same word. We will implicitly use this fact below for uniqueness of lifts to Γ.

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. If $\mathcal{C}$ is taken standard, Γ_2 has no homotopy in degree 2.

By our definitions above, a tile of Γ_2 is a topological disk whose boundary is labelled by some word of R.

Let D be a simply connected puzzle with respect to the tiles of Γ_2. We are going to show that there exists a constant $C > 0$ such that for any D, if D is van Kampen-reduced, then D 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 [S]).

We can safely assume that all edges of D lie on 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. In particular, we suppose that any edge is adjacent to some face).

Let e be an internal edge of D, between 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 boundary word as f_1 and f_2 respectively (maybe up to inversion). These faces are unique, since no two distinct faces of Γ_2 can bear the same word as it would contradict the $1/6$ condition on pieces.

The edge e belongs to 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 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 the $1/6$ assumption.

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 graph D into parts having all 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 2 – *Let $\Delta = \operatorname{diam}(\Gamma)$. Suppose that $\mathcal{C}$ was chosen to be the set of closed paths embedded 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 2 – 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 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 3 – *For any choice of R, there exists a constant α 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 3 – Indeed, the existence of an isoperimetric constant does not depend on the presentation. □

These last affirmations only express in terms of diagrams the fact that the fundamental group of Γ, which is 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 [GH], or [LS]), stated in the form we need. We include a short proof here for completeness.

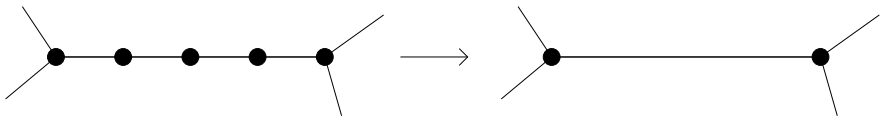
LEMMA 4 – *Let R be a set of simply connected reduced tiles. Suppose that any piece with respect to two tiles $t, t' \in R$ is a word of length at most λ times the smallest boundary length of t and t', for some constant $\lambda < 1/6$.*

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.*

The value $1/6$ is optimal, as a hexagonal tiling shows.

PROOF – Let D be a van Kampen-reduced puzzle w.r.t. R. We can suppose that its boundary word is reduced (isoperimetry for the reduction implies isoperimetry for D a fortiori). Its skeleton is a planar graph. Define a metric graph D' as follows: start with D, but for each pair of adjacent faces, replace all consecutive edges between these two faces by a single edge. Define the length of this new edge as the number of edges it replaces. So D' is a metric planar graph having the same faces as D but in which every vertex is of degree at least 3. The small cancellation assumption states that the length of any edge in D' is at most λ times the smallest boundary length of the two adjacent faces.



Let V, E and F denote the number of vertices, edges and faces of D', respectively. Let E_i and E_e denote the number of internal and external edges, and F_i, F_e the number of internal faces and faces adjacent to the boundary. Let F_e^1 and F_e^2 denote the number of external faces with exactly one, or at least two, external edges, respectively.

As every vertex has degree at least 3, we have $E \geqslant 3V/2$. By definition we have $E_e \geqslant F_e^1 + 2F_e^2$. Since $\lambda < 1/6$, any internal face has at least 6 edges.

We now prove that there are at least two external faces with exactly one external edge and at most three internal edges (this will prove the first assertion of the lemma). Let F_e' and F_e'' be the number of external faces with exactly one external edge and, respectively, at most three or at least four internal edges. Since any internal face has at least six internal edges, and since any external face with two external edges has at least two internal edges, we get $E_i \geqslant \frac{1}{2}\left(F_e' + 4F_e'' + 2F_e^2 + 6F_i\right)$.

The Euler formula writes

$$\begin{aligned} 1 &= V - E + F = V - \frac{2}{3}E - \frac{1}{3}E + F \\ &\leqslant 0 - \frac{1}{3}E_i - \frac{1}{3}E_e + F_e + F_i \\ &\leqslant -\frac{1}{6}\left(F_e' + 4F_e'' + 2F_e^2 + 6F_i\right) - \frac{1}{3}\left(F_e' + F_e'' + 2F_e^2\right) + F_e' + F_e'' + F_e^2 + F_i \\ &= \frac{1}{2}F_e' \end{aligned}$$

Hence there exist at least two faces with exactly one external edge and at most three internal ones.

By the small cancellation assumption, the internal edges of such a face have cumulated length at most 3λ times the length of this face. This proves the first assertion of the lemma.

The second assertion is easy. Suppose that the boundary word of some puzzle is a proper subword w of the boundary word x of some tile t. We know that w contains a subword which contains a proportion at least $(1 - 3\lambda) > 1/2$ of the boundary word of some tile t'. If $|w| \leqslant |x|/2$ we have $t' \neq t$, so t and t' share a common subword of length more than one half the boundary length of t', which contradicts the small cancellation assumption. If $|w| > |x|/2$, then define a new puzzle by gluing the inverse tile t^{-1} to the original puzzle (and reducing the resulting puzzle): the new puzzle now contains the reduction of the word wx^{-1}, which is at least one half of the boundary word of t^{-1}; use the same argument.

Isoperimetry follows immediately. Let F be a external face of D' with at most three internal edges. Let ℓ be the length of F. Let D'' be the (maybe non connected) graph obtained from D' by removing F. We have $|\partial D'| = |\partial D''| +$

$\ell - 2\,|F \cap D''| \geqslant |\partial D''| + (1-6\lambda)\ell$ hence the conclusion by backwards induction on the number of faces.

The fact that the boundary length of D is at least the boundary length of the largest face is trivial if D has only one face, and otherwise follows from the fact that there are two external faces having at least a proportion $1 - 3\lambda > \frac{1}{2}$ of their length on the boundary. Remove such a face (but not the largest one): this decreases the boundary length. Iterate this process up to a diagram containing only the largest face. □

COROLLARY 5 – *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.

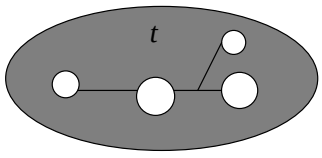
PROOF OF THE COROLLARY –

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 interior of t contains no other non-simply connected tile.

The interior of t is embedded and is homeomorphic to a disc with some finite number k of holes. Let $D'_1, \ldots, D'_n$ be 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.

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 is the union of closed sets $D''_1, \ldots, D''_q$ 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. Since tiles are reduced, 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 $p_1, \ldots, p_r$ on the boundary of D''_i such that p_j belongs to some D''_k for $k \neq j$. Now define T_i to be made of segments $p_0 p_j \subset D''_i$ 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 $p_1, \ldots, p_r$. 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 the previous lemma: 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 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.

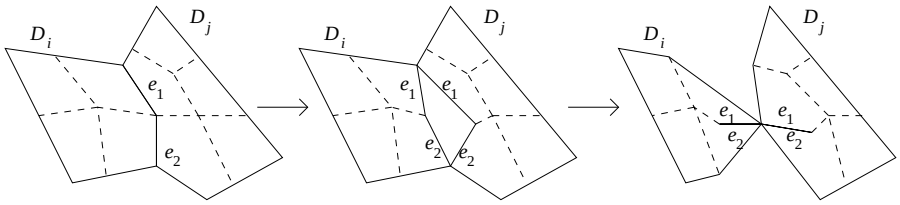
First, define a disjoint union of puzzles D' by taking the disjoint union of all tiles of D and gluing them along the internal edges of D originating from Γ_2. All internal edges of D' originate from Γ_2.

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 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 [CH]. 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 diagram (denoted D again) has the same number of faces as before, and no more boundary edges. Thus, proving isoperimetry for the modified diagram 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.) Note that these tiles are not necessarily simply connected.

These tiles satisfy the condition of Corollary 5. 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 5 they are simply connected.

Then by Lemma 4, the boundary of D is at least $1 - 6\lambda$ times the sum of the boundary lengths of the D_i's. Since D is minimal, each D_i is as well, and as D_i is simply connected, by Corollary 3 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 R is standard (so that Γ_2 is aspherical) and that there exists a van Kampen-reduced spherical diagram D. Define the D_i's as above. As we showed

above that the boundary length of D (which is 0) is at least the boundary length of any D_i, this means that the boundary length of all the D_i's is 0. Hence each D_i is a spherical diagram itself. But by definition D_i lifts to Γ_2. We conclude with the following lemma.

LEMMA 6 – *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 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. □

The last assertions of the theorem follow easily from the last assertions of Lemma 4. The smallest relation in the group presented by $\langle\, S \mid R \,\rangle$ is the boundary length of the smallest puzzle, which by Lemma 4 is at least the smallest boundary length of the D_i's, which is at least g. Similarly, any reduced word representing the trivial element in the group is the boundary of a reduced diagram, 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 in the graph labelling a word x is equal to a shorter word y in the quotient. This means that there exists a puzzle with boundary word xy^{-1}, made up of tiles with cycles of Γ as boundary words. 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 $|xy^{-1}| < |w|$ and in particular, the puzzle bordering xy^{-1} cannot contain a tile with boundary word w. Glue a tile with boundary word $(xz)^{-1}$ to the puzzle xy^{-1} along the subwords x and x^{-1}. This results in a (reduced) puzzle bordering $z^{-1}y^{-1}$, containing a tile $z^{-1}x^{-1}$. This is impossible as by assumption $|y| < |x|$.

The computation of the Euler characteristic immediately follows, using that the cohomological dimension is at most 2.

This proves the theorem. □

3 Further remarks

REMARK 7 – 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 8 – 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 9 (M. GROMOV) – *Let Γ be a 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 is standard) 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.

(In the notation of [GH], 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 10 – The same kind of theorem hold 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.

REMARK 11 – Using the techniques in [D] or [O], 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.

Thanks to Thomas Delzant for having brought the problem to my attention and to Étienne Ghys and Pierre Pansu for helpful discussions and comments on the manuscript.

References

[CH] J.D. Collins, J. Huebschmann, *Spherical diagrams and identities among relations*, Math. Ann. **261** (1982), No. 2, 155–183.

[D] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), No. 3, 661–682.

[G1] 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.

[G2] M. Gromov, *Random Walk in Random Groups*, Geom. Funct. Anal. **13** (2003), No. 1, 73–146.

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

[LS] R.C. Lyndon P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[O] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, this volume.

[S] 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.

Contents

II
Espaces métriques mesurés concentrés

Contenant :

Concentrated spaces seen as product spaces

Concentration spectrale dans les graphes

Concentrated spaces seen as product spaces

Abstract

It is well-known that a product of probability spaces has Gaussian concentration. We show that, in a weak sense, the converse is true: on any space with Gaussian concentration it is possible to define a structure resembling a product space.

Independence (exact or approximate) has long been known to be a key feature to prove Gaussian concentration inequalities. Namely, product spaces are concentrated (see [Tal] for an introductory paper), and this fact has many known refinements (cf. the reference book [Led]). Here we are interested in the converse: is a concentrated space anyhow close to a product space?

Asking for all concentrated spaces to be product spaces (or close to product spaces in a measured-metric sense) would probably be too much. Indeed, concentration is stable by a contraction (i.e. a 1-Lipschitz measure-preserving map). We could reasonably ask whether a concentrated space is close to a contraction of some product space.

A fact pointing in this direction is the following: smooth convex bodies, as manifolds with positive curvature, are concentrated (by a theorem of Gromov, see [Gro]). By Dvoretzky's theorem (which itself is linked to concentration), they admit big nearly-spherical sections. But the sphere itself is very close to a product space ($\mathbb{R}^n$ with Gaussian measure).

Here we construct a kind of dual product structure on a concentrated space. Namely, if some space X has Gaussian concentration with "observable diameter" σ, we prove that it is possible to find some functions $f_1, \dots, f_n$ satisfying an inequality of the form:

$$\mathbb{P}(f_1 > a_1, \dots, f_k > a_k) \leqslant A \prod e^{-a_i^2/C\sigma^2}$$

where the right-hand side is simply a product of the natural estimates for each of the $\mathbb{P}(f_i > a_i)$. (Because concentration is stable by contractions, as noted above, it is probably not possible to hope for more than an inequality and for this estimate to give a correct order of magnitude on all spaces.)

Of course, this is trivial if the f_i are constant functions, so it is necessary to add some more constraint saying that the f_i's do vary. For the moment being we only have a constraint on the range of these functions; we hope at least to be able to give a lower bound on their variance.

This text represents only the beginning of our investigations.

1 Notations and statement

Let (X, d, μ) be a mm-space of mass 1. Suppose X has Gaussian concentration: there exists constants A and σ such that for any 1-Lipschitz function $f : X \to \mathbb{R}$ with $\mathbb{E}f = 0$, for any $\lambda \geqslant 0$ we have

$$\mathbb{E}e^{\lambda f} \leqslant Ae^{\lambda^2\sigma^2/2}$$

where σ is to be interpreted as the observable diameter of X. (This is more comfortable and essentially equivalent to the equally traditional assumption that $\mathbb{P}(|f - \mathbb{E}f| > t) \leqslant Ae^{-t^2/2\sigma^2}$.)

Let also R be the mean radius of X, that is $\mathbb{E}\,\mathrm{dist}(\cdot,\cdot)$.

Let N be the "statistical dimension" $N = (R/\sigma)^2$.

THEOREM 1 – *Let m be a positive integer. There exists a family of functions $f_i : X \to \mathbb{R}$, for $1 \leqslant i \leqslant m$, such that*

- *Each f_i is 2-Lipschitz and satisfies $|\mathbb{E}f_i| \leqslant \sigma\sqrt{2\log 8mA}$*
- *The values of f_i range in an interval of size at least $R\left(1 - \sqrt{8(\log 8mA)/N}\right)$*
- *For each $1 \leqslant k \leqslant m$, for each k-tuple $1 \leqslant i_1 < \ldots < i_k \leqslant m$, for each k-tuple t_j with $0 \leqslant t_j \leqslant 2R$ we have*

$$\mathbb{P}\left(\left|f_{i_j}\right| \geqslant t_j\ \forall j\right) \leqslant 2\left(8e\,m\,A\log_2 N\right)^k e^{-\left(\sum t_j^2\right)/4\sigma^2}$$

Asymptotically when $N \to \infty$ (with fixed m and A), better constants can be achieved and the theorem reads

THEOREM 1' – *Let m be a positive integer. There exists a family of functions $f_i : X \to \mathbb{R}$, for $1 \leqslant i \leqslant m$, such that*

- *Each f_i is 2-Lipschitz and satisfies $|\mathbb{E}f_i| = O(\sigma)$*
- *The values of f_i range in an interval of size at least $R(1 - O(1/\sqrt{N}))$*
- *For each $1 \leqslant k \leqslant m$, for each k-tuple $1 \leqslant i_1 < \ldots < i_k \leqslant m$, for each k-tuple t_j with $\sigma\sqrt{\log\log N} \ll t_j \leqslant 2R$ we have*

$$\mathbb{P}\left(\left|f_{i_j}\right| \geqslant t_j\ \forall j\right) \leqslant e^{-\left(\sum t_j^2\right)/(2+o(1))\sigma^2}$$

Of course, without the condition on the range of the functions (which is not so weak in a concentrated space), the assertion would be trivial since $f_i = 0$ would do. One would like to get better properties e.g. the variance of f_i compared to σ (at least when N is large) and some understanding of how these functions separate the points of X.

2 Proof

For $x \in X$, let $R_x = \mathbb{E}\,\mathrm{dist}(x, .)$ be the mean distance to x. This is a 1-Lipschitz function, of mean R, hence the probability that a randomly chosen x satisfies $|R_x - R| > t$ is at most $2Ae^{-t^2/2\sigma^2}$. The probability that a randomly chosen m-tuple of points $(x_i)_{1 \leqslant i \leqslant m}$ satisfies $\sup_i |R_{x_i} - R| \leqslant t$ is thus at least $1 - 2mAe^{-t^2/2\sigma^2}$. For $t = \sigma\sqrt{2\log 8mA}$ this is at least $3/4$.

Let $k \leqslant m$. Fix some $z \in X$. Let a_i, $1 \leqslant i \leqslant k$ be positive coefficients. Let f be the function on X^k defined by

$$f : (x_1, \ldots, x_k) \mapsto \sum_i a_i(\mathrm{dist}(z, x_i) - R_z)$$

The function $e^{\lambda f}$ is a product of independent functions and so

$$\mathbb{E}_{(x_i)} e^{\lambda f} = \prod \mathbb{E} e^{\lambda a_i(\mathrm{dist}(z,.) - R_z)} \leqslant A^k e^{\lambda^2(\sum a_i^2)\sigma^2/2}$$

by assumption, since each component of f has zero mean.

Hence for any $t \geqslant 0$

$$\mathbb{P}_{(x_i)}(f \geqslant t) \leqslant A^k e^{-t^2/\left(2\sigma^2 \sum a_i^2\right)}$$

For $1 \leqslant i \leqslant k$ let $t_k \geqslant 0$. We have (the probabilities are w.r.t. (x_i))

$$\begin{aligned} \mathbb{P}_{(x_i)}\left(\forall i, \mathrm{dist}(z, x_i) \geqslant R_z + t_i\right) &\leqslant \mathbb{P}(f \geqslant \sum a_i t_i) \\ &\leqslant A^k e^{-(\sum a_i t_i)^2/\left(2\sigma^2 \sum a_i^2\right)} \\ &\leqslant A^k e^{-\left(\sum t_i^2\right)/2\sigma^2} \end{aligned}$$

taking $a_i = t_i$.

Integrating with respect to z we get that for any (t_i) we have

$$\mathbb{P}_{z,(x_i)}\left(\forall i, \mathrm{dist}(z, x_i) \geqslant R_z + t_i\right) \leqslant A^k e^{-\left(\sum t_i^2\right)/2\sigma^2}$$

or (symmetrizing in the definition of the f above)

$$\mathbb{P}_{z,(x_i)}\left(\forall i, |\mathrm{dist}(z, x_i) - R_z| \geqslant t_i\right) \leqslant (2A)^k e^{-\left(\sum t_i^2\right)/2\sigma^2}$$

Consider this as a function of the (x_i)'s: set

$$P_k((x_i)) = \mathbb{P}_z\left(\forall i, |\mathrm{dist}(z, x_i) - R_z| \geqslant t_i\right)$$

for a given (x_i). Thus for any (t_i) we have

$$\mathbb{E}_{(x_i)} P_k \leqslant (2A)^k e^{-\left(\sum t_i^2\right)/2\sigma^2}$$

and the Markov inequality states that

$$\mathbb{P}_{(x_i)}\left(P_k \geqslant B(2A)^k e^{-\left(\sum t_i^2\right)/2\sigma^2}\right) \leqslant 1/B$$

Thus as soon as $B > 1$, there exists a k-tuple (x_i) such that $P_k \leqslant B(2A)^k e^{-\left(\sum t_i^2\right)/2\sigma^2}$ that is

$$\mathbb{P}_z\left(\forall i, |\mathrm{dist}(z, x_i) - R_z| \geqslant t_i\right) \leqslant B(2A)^k e^{-\left(\sum t_i^2\right)/2\sigma^2}$$

as was needed.

We will set $f_i = \{z \mapsto \mathrm{dist}(z, x_i) - R_z\}$.

But this was for fixed (t_i). The k-tuple (x_i) may not work for all k-tuples (t_i) as we need.

Consider the family T of k-tuples (t_i) where each t_i is of the form $t_i = R/2^j$ with $0 \leqslant j \leqslant \log_2(R/\sigma) = \log_2 N$. Any k-tuple (t_i) with $\sigma \leqslant t_i \leqslant 2R$ can be approached by this family up to a factor 2. This means that if we have a k-tuple $(x_i) \in X^k$ satisfying the property above for all $(t_i) \in T$, then for any $t_i \leqslant 2R$ we have

$$\mathbb{P}_z\left(\forall i, |\mathrm{dist}(z, x_i) - R| \geqslant t_i\right) \leqslant B(2A)^k \prod_i e^{-t_i^2/4\sigma^2 + 1}$$

where the $+1$ accounts for integer parts in $\log_2 N$ and for the case $t_i \leqslant \sigma$.

There are $(\log_2 N + 1)^k \leqslant (2\log_2 N)^k$ such k-tuples in T. So, taking $B = B'(\log_2 N)^k$, the probability (in (x_i)) that there exists one k-tuple (t_i) such that

$$P_k((x_i)) > B'(2\log_2 N)^k (2A)^k e^k\, e^{-\left(\sum t_i^2\right)/4\sigma^2}$$

is at most $1/B'$.

This was for a given $k \leqslant m$. Now, if we pick a random m-tuple $(x_1, \ldots, x_m)$, there are at most m^k choices for a k-tuple $(x_{i_j})_{1 \leqslant j \leqslant k}$. Taking $B' = 2(2m)^k$, for a given k the probability (in $(x_1, \ldots, x_m)$) that there exists a choice of $(x_{i_j})_{1 \leqslant j \leqslant k}$ and a k-tuple (t_j) such that

$$P_k((x_{i_j})) > 2(2m)^k (2\log_2 N)^k (2A)^k e^k\, e^{-\left(\sum t_j^2\right)/4\sigma^2}$$

is at most $1/2^{k+1}$.

Summing over $1 \leqslant k \leqslant m$, the probability that, for a random m-tuple $(x_1, \ldots, x_m)$, there exists a $k \leqslant m$, a k-tuple $(x_{i_j})_{1 \leqslant j \leqslant k}$ and a k-tuple (t_j) such that

$$P_k((x_{i_j})) > 2\left(8e\, m\, A \log_2 N\right)^k\, e^{-\left(\sum t_j^2\right)/4\sigma^2}$$

is at most $1/2$. Hence there exists a m-tuple satisfying all the opposite inequalities.

Now set $f_i = \{z \mapsto \mathrm{dist}(z, x_i) - R_z\}$. This is a 2-Lipschitz function, of mean $R_{x_i} - R$. So if we choose x_i such that $|R_{x_i} - R| \leqslant \sigma\sqrt{2\log 8mA}$ (which holds for at least $3/4$ of the m-tuples as was shown above), we have $|\mathbb{E}f_i| \leqslant \sigma\sqrt{2\log 8mA}$ and $f_i(x_i) = -R_{x_i} \leqslant -R + \sigma\sqrt{2\log 8mA}$, so the range of f_i is of size at least $R - 2\sigma\sqrt{2\log 8mA} = R\left(1 - \sqrt{8(\log 4mA)/N}\right)$.

This ends the proof.

REMARK – In the above, we only used concentration of functions of the form $\mathrm{dist}(x, .)$. So this also works when σ is defined with respect to these functions only.

References

[Gro] M. Gromov, *Isoperimetric inequalities in Riemannian manifolds*, in *Asymtotic theory of finite dimensional normed spaces*, V. Milman and G. Schechtman, Lecture Notes in Mathematics **1200**, Springer-Verlag, Berlin (1986).

[Led] M. Ledoux, *The concentration of measure phenomenon*, Mathematical surveys and monographs **89**, Amer. Math. Soc., Providence (2001).

[Tal] M. Talagrand, *A new look at independence*, Ann. Prob. **24** (1996), No. 1, 1–34.

Contents

Concentration spectrale dans les graphes

Résumé

On donne une évaluation spectrale de la concentration de la mesure sur des produits de graphes mesurés. On montre qu'à petite échelle la concentration est gaussienne, et exponentielle à grande échelle.

On se place sur un graphe G, dont on note $\sim$ la relation d'adjacence. On note v le nombre maximal d'arêtes adjacentes à un même sommet, et on suppose $v < \infty$.

On se donne une mesure de probabilité μ sur G. On note $\mathbb{E}$ l'espérance sous μ.

Soit f une fonction de $L^1(G,\mu) \cap L^2(G,\mu)$. On pose

$$E(f) = \sum_{x \in G} \mu(x) \sum_{y \sim x} (f(x) - f(y))^2$$

On suppose que pour un certain $C > 0$ on a l'inégalité de Poincaré :

$$E(f) \geqslant C \left(\mathbb{E}(f^2) - (\mathbb{E}f)^2 \right)$$

(vérifiée par exemple si μ est uniforme avec pour C la première valeur propre non nulle du laplacien sur le graphe).

On note également $\|f\|_{Lip} = \sup_{y \sim x} |f(y) - f(x)|$.

Si G n'est pas réduit à un point (auquel cas l'inégalité est vérifiée pour tout C...), on a, en considérant la fonction dirac en un point de mesure inférieure à $1/2$, que $C \leqslant 2v$.

On définit la quantité

$$D = \frac{1}{2 \log \left(1 + \sqrt{\frac{C}{2v}} \right)}$$

à interpréter comme le diamètre observable de (G,μ). Comme $C \leqslant 2v$ (si G n'est pas réduit à un point), on a $D \leqslant \sqrt{2v/C} / \log 2$.

En particulier, la variance d'une fonction 1-lipschitzienne est inférieure à D^2.

1 Inégalité de concentration

On redémontre ici pour les graphes un théorème de Gromov et Milman pour des espaces métriques probabilisés continus.

THÉORÈME 1 – *Soit f une fonction 1-lipschitzienne. Alors f est dans $L^1(G,\mu)$ et sous μ :*

$$\mathbb{P}\left(|f-\mathbb{E}f|\geqslant t\right)\leqslant 12\,e^{-t/D}$$

et pour tout $0\leqslant\lambda\leqslant 1/D$:

$$\mathbb{E}e^{\lambda f}\leqslant 6e^{\lambda\mathbb{E}f}$$

Notons que pour tout λ, par convexité, $\mathbb{E}e^{\lambda f}\geqslant e^{\lambda\mathbb{E}f}$
En utilisant $D\leqslant\sqrt{2v/C}/\log 2$, l'inégalité se récrit

$$\mathbb{P}\left(|f-\mathbb{E}f|\geqslant t\right)\leqslant 12.\,2^{-t\sqrt{\frac{2C}{v}}}$$

DÉMONSTRATION –
Observons d'abord que la première inégalité découle de la seconde par l'inégalité de Tchebychev exponentielle.
On se place d'abord dans le cas où f est bornée.
Pour tous réels x et y, on a $|e^y-e^x|=|e^x\left(e^{y-x}-1\right)|\leqslant e^x\left(e^{|y-x|}-1\right)$.
Pour toute fonction g sur le graphe, on a

$$\begin{aligned}E(e^g) &= \sum_{x\in G}\mu(x)\sum_{y\sim x}\left(e^{g(y)}-e^{g(x)}\right)^2\\ &\leqslant \sum_{x\in G}\mu(x)\sum_{y\sim x}\left(e^{\|g\|_{Lip}}-1\right)^2 e^{2g(x)}\\ &\leqslant \left(e^{\|g\|_{Lip}}-1\right)^2 v\,\mathbb{E}e^{2g}\end{aligned}$$

Par conséquent, si f est 1-lipschitzienne :

$$\mathbb{E}e^{2\lambda f}-\left(\mathbb{E}e^{\lambda f}\right)^2\leqslant 1/C\;E\left(e^{\lambda f}\right)\leqslant v/C\;\left(e^{\lambda}-1\right)^2\mathbb{E}e^{2\lambda f}$$

ou encore

$$\mathbb{E}e^{\lambda f}\leqslant\frac{1}{1-\left(e^{\lambda/2}-1\right)^2\,v/C}\left(\mathbb{E}e^{\frac{\lambda}{2}f}\right)^2$$

En itérant :

$$\mathbb{E}e^{\lambda f}\leqslant\left(\mathbb{E}e^{f\lambda/2^k}\right)^{2^k}\prod_{1\leqslant i\leqslant k}\left(1-\left(e^{\lambda/2^i}-1\right)^2\,v/C\right)^{-2^{i-1}}$$

Or $\left(\mathbb{E}e^{f\lambda/2^k}\right)^{2^k}\to e^{\lambda\mathbb{E}f}$ quand $k\to\infty$ (prendre le logarithme, développer à l'ordre 1), car f est bornée.

Quant au produit infini, pour $\lambda \leqslant 2\log\left(1+\sqrt{\frac{C}{2v}}\right)$, il converge, et sa valeur est inférieure à sa valeur pour ce choix de λ, soit

$$\prod_{k\geqslant 1}\left(1-\frac{v}{C}\left(\left(1+\sqrt{\frac{C}{2v}}\right)^{2^{1-k}}-1\right)^2\right)^{-2^{k-1}}$$

On a $(1+x)^a \leqslant 1+ax$ pour $x\geqslant 0, a\leqslant 1$; ce produit est donc inférieur à

$$\prod_{k\geqslant 1}\left(1-\frac{v}{C}\left(\frac{1}{2^{k-1}}\sqrt{\frac{C}{2v}}\right)^2\right)^{-2^{k-1}} = \prod_{k\geqslant 1}\left(1-\frac{1}{2^{2k-1}}\right)^{-2^{k-1}}$$

En utilisant $(1-x)^a \geqslant 1-ax$ pour $x\geqslant 0, a\geqslant 1$, on déduit que ce produit est encore inférieur à

$$\prod_{k\geqslant 1}\left(1-\frac{1}{2^k}\right)^{-1}\leqslant 6$$

Par conséquent, pour $\lambda \leqslant 2\log\left(1+\sqrt{\frac{C}{2v}}\right)$, on a

$$\mathbb{E}e^{\lambda f}\leqslant 6e^{\lambda \mathbb{E}f}$$

ce qui achève la démonstration quand f est bornée.

Pour f non bornée dans L^1, soit f_n, pour n entier, la fonction f où on a remplacé par n toutes les valeurs supérieures à n. La suite de fonctions $e^{\lambda f_n}$ croît vers $e^{\lambda f}$, et pour tout n on a $\mathbb{E}e^{\lambda f_n}\leqslant 6e^{\lambda \mathbb{E}f_n}\leqslant 6e^{\lambda\mathbb{E}f}$. Par théorème de convergence monotone, on a donc $\mathbb{E}e^{\lambda f}\leqslant 6e^{\lambda\mathbb{E}f}$.

Enfin, montrons que la fonction distance à un point est dans L^1. Soit f dans L^1. On a d'après ce qui précède $\mathbb{P}\left(|f-\mathbb{E}f|\geqslant t\right)\leqslant Ae^{-Bt}$ pour certains $A, B>0$. Par conséquent, si Mf est une médiane de f, on a $\mathbb{P}\left(|f-Mf|\geqslant t\right)\leqslant Ae^{BMf-Bt}$. Soit f la fonction distance à un point, et soit f_n la fonction f tronquée à la valeur n. On a $Mf<\infty$, et même $Mf_n = Mf$ à partir d'un certain rang. Les fonctions indicatrices de $\{|f_n-Mf|\geqslant t\}$ croissent vers la fonction indicatrice de $\{|f-Mf|\geqslant t\}$; leur intégrale est inférieure à Ae^{BMf-Bt}. On a donc $\mathbb{P}\left(|f-Mf|\geqslant t\right)\leqslant Ae^{BMf-Bt}$. Ensuite, on a

$$|\mathbb{E}f-Mf|\leqslant \mathbb{E}\,|f-Mf| = \int_0^\infty \mathbb{P}\left(|f-Mf|\geqslant t\right)\leqslant \int Ae^{BMf-Bt}<\infty$$

Ainsi, la fonction distance à un point et, partant, toute fonction lipschitzienne, est dans L^1. □

COROLLAIRE 2 – *Soit x un point de G. Alors l'espérance de la fonction distance à x vérifie*

$$\mathbb{E}\,\mathrm{dist}(\cdot,x)\leqslant D\,\left(-\log\mu(x)+\log 12\right)$$

Noter en particulier la dépendance logarithmique en la petitesse de la masse du point considéré.

DÉMONSTRATION – Appliquer l'estimation en x avec $t = \mathbb{E}\,\mathrm{dist}(\cdot, x)$. □

2 Lemmes

On suppose qu'est donné un graphe G ainsi qu'un nombre D tel que pour toute fonction 1-lipschitzienne f, pour tout $\lambda \leqslant 1/D$, on a $\mathbb{E}e^{\lambda f} \leqslant 6e^{\lambda \mathbb{E} f}$ (cas précédent avec $D = 1/2\log(1 + \sqrt{C/2v})$).

LEMME 3 – *Soit f une fonction 1-lipschitzienne sur G telle que $\mathbb{E}f = 0$. Alors pour tout $\lambda \leqslant 1/D$, on a $\mathbb{E}e^{\lambda|f|} \leqslant 13$.*

DÉMONSTRATION – On pose $f_+ = \max(f, 0)$ et $f_- = \max(-f, 0)$ de sorte que $f = f_+ - f_-, |f| = f_+ + f_-$. En décomposant G selon le signe de f, on obtient que $\mathbb{E}e^{\lambda f} = \mathbb{E}e^{\lambda f_+} + \mathbb{E}e^{-\lambda f_-} - 1$, et comme $\mathbb{E}e^{-\lambda f_-} > 0$, on déduit $\mathbb{E}e^{\lambda f_+} \leqslant \mathbb{E}e^{\lambda f} + 1$.

Répétant l'opération pour $|f|$, on obtient $\mathbb{E}e^{\lambda|f|} = \mathbb{E}e^{\lambda f_+} + \mathbb{E}e^{\lambda f_-} - 1$. Quitte à changer f en $-f$, on peut supposer $\mathbb{E}e^{\lambda f_+} \geqslant \mathbb{E}e^{\lambda f_-}$, et donc $\mathbb{E}e^{\lambda f_-} \leqslant \mathbb{E}e^{\lambda f} + 1$.

On obtient ainsi $\mathbb{E}e^{\lambda|f|} \leqslant 2\mathbb{E}e^{\lambda f} + 1$. Par hypothèse, on a $\mathbb{E}e^{\lambda f} \leqslant 6e^{\lambda \mathbb{E} f} = 6$ d'où le lemme. □

LEMME 4 – *Soit f une fonction 1-lipschitzienne sur G telle que $\mathbb{E}f = 0$. Alors f est dans $L^p(G)$ pour tout $1 \leqslant p < \infty$ et, pour tout entier n, on a*

$$\mathbb{E}f^n \leqslant 13n!D^n$$

DÉMONSTRATION – Posons $\lambda = 1/D$, on a

$$\sum \mathbb{E}\frac{|f|^n}{n!D^n} = \mathbb{E}e^{|f|/D} \leqslant 13$$

d'où l'inégalité du lemme en majorant chaque terme par la somme. Comme f est dans $L^n(G)$ pour tout n entier, elle est dans $L^p(G)$ pour tout $1 \leqslant p < \infty$. □

REMARQUE 5 – Sans l'hypothèse $\mathbb{E}f = 0$, on obtient $\mathbb{E}f^n \leqslant n!D^n(1 + 12e^{n\mathbb{E}f/D})$.

LEMME 6 – *Si f est une fonction 1-lipschitzienne sur G, pour $0 \leqslant \lambda < 1/D$, on a*

$$\mathbb{E}e^{\lambda f} \leqslant e^{\lambda \mathbb{E} f} e^{\lambda^2 D^2/2 + 13(\lambda D)^3/(1-\lambda D)}$$

DÉMONSTRATION – On peut supposer $\mathbb{E}f = 0$. Alors, on a

$$\begin{aligned}\mathbb{E}e^{\lambda f} &= 1 + \frac{\lambda^2}{2}\mathbb{E}f^2 + \sum_{n\geqslant 3}\frac{\lambda^n}{n!}\mathbb{E}f^n \\ &\leqslant 1 + \frac{\lambda^2 D_i^2}{2} + 13\sum_{n\geqslant 3}\lambda^n D^n \\ &= 1 + \frac{\lambda^2 D_i^2}{2} + 13\frac{(\lambda D)^3}{1-\lambda D}\end{aligned}$$

d'après le lemme 4. □

3 Graphes produits

On considère N graphes $G_1, \ldots, G_N$ munis des mesures de probabilité μ_i, de valence inférieure à v_i, et vérifiant l'inégalité de Poincaré avec des constantes C_i. On définit les diamètres D_i comme précédemment.

On construit un graphe G dont les sommets sont les n-uplets de sommets des G_i, et tel qu'il existe une arête entre $(x_1, \ldots, x_N)$ et $(y_1, \ldots, y_N)$ si et seulement s'il existe un i tel que $x_j = y_j$ pour $j \neq i$ et $x_i \sim y_i$ dans G_i. (C'est aussi le 1-squelette du produit des graphes vus comme 1-complexes.) On munit G de la mesure $\mu = \otimes\mu_i$.

Sur le graphe produit, on devrait avoir, à grande échelle, la même concentration exponentielle que sur chacun des graphes. À petite échelle, on devrait observer une concentration gaussienne. Prouvons le premier point.

THÉORÈME 7 – *Soit f une fonction 1-lipschitzienne sur G. Alors f est dans $L^1(G,\mu)$ et sous μ :*

$$\mathbb{P}\left(|f - \mathbb{E}f| > t\right) \leqslant 2.6^N\, e^{-t/\max D_i}$$

et pour tout $0 \leqslant \lambda \leqslant 1/\max D_i$:

$$\mathbb{E}e^{\lambda f} \leqslant 6^N e^{\lambda \mathbb{E}f}$$

Si on suppose que sur G_i l'estimation exponentielle en e^{-t/D_i} est correcte, on voit que la dépendance en t est correcte dans l'évaluation pour le produit : il suffit de prendre f ne dépendant que de la coordonnée i, réalisant l'estimation sur G_i. Par contre, pour la dépendance en N, on attendrait plutôt du $\sqrt{N}$ au moins pour les petites valeurs de t.

DÉMONSTRATION – Traitons le cas $N = 2$. On note $\mathbb{E}_i$ l'espérance selon la i-ième composante. On a

$$\begin{aligned}\mathbb{E}e^{\lambda f} &= \sum_{x\in G_1}\mu_1(x)\sum_{y\in G_2}\mu_2(y)e^{\lambda f(x,y)}\\ &\leqslant \sum_{x\in G_1}\mu_1(x)\,6e^{\lambda\mathbb{E}_2 f(x,\cdot)}\\ &= 6\mathbb{E}_1 e^{\lambda\mathbb{E}_2 f(x,\cdot)}\end{aligned}$$

si $\lambda \leqslant 1/D_2$. Or $\mathbb{E}_2 f(x,\cdot)$ est une fonction 1-lipschitzienne sur G_1, dont l'espérance est $\mathbb{E}f$. Par conséquent, si $\lambda \leqslant 1/D_1$, on a $6\mathbb{E}_1 e^{\lambda\mathbb{E}_2 f(x,\cdot)} \leqslant 36e^{\lambda\mathbb{E}f}$. Etc. □

LEMME 8 – *Soit f une fonction 1-lipschitzienne sur G. Alors pour $\lambda \leqslant 1/2\max D_i$:*

$$\mathbb{E}e^{\lambda f} \leqslant \exp\left(\lambda\mathbb{E}f + \frac{\lambda^2}{2}\sum D_i^2 + 13\lambda^3\sum\frac{D_i^3}{1-\lambda D_i}\right)$$

DÉMONSTRATION – Traitons le cas $N = 2$. On a

$$\begin{aligned}\mathbb{E}e^{\lambda f} &= \sum_{x\in G_1}\mu_1(x)\sum_{y\in G_2}\mu_2(y)e^{\lambda f(x,y)}\\ &= \sum_{x\in G_1}\mu_1(x)\mathbb{E}_2 e^{\lambda f(x,\cdot)}\\ &\leqslant \sum_{x\in G_1}\mu_1(x)e^{\lambda\mathbb{E}_2 f(x,\cdot)}e^{\lambda^2 D_2^2/2+13(\lambda D_2)^3/(1-\lambda D_2)}\end{aligned}$$

d'après le lemme.

Or $\mathbb{E}_2 f$ est une fonction 1-lipschitzienne sur G_1, dont l'espérance est $\mathbb{E}f$. Par conséquent, si $\lambda \leqslant 1/D_1$, on a $\mathbb{E}_1 e^{\lambda\mathbb{E}_2 f} \leqslant e^{\lambda\mathbb{E}f}e^{\lambda^2 D_1^2+13(\lambda D_1)^3/(1-\lambda D_1)}$ ce qui achève la démonstration. □

THÉORÈME 9 – *Soit f une fonction 1-lipschitzienne sur G. Soit $D = \max D_i$. Pour tout $0 \leqslant t \leqslant \frac{1}{64}\sum D_i^2/D$, on a*

$$\mathbb{P}(|f-\mathbb{E}f| \geqslant t) \leqslant 2\exp -\frac{t^2}{4\sum D_i^2}$$

Pour t plus grand que cette valeur, on a

$$\mathbb{P}(|f-\mathbb{E}f| \geqslant t) \leqslant 2\exp -\frac{t}{256D}$$

Par ailleurs pour tout $t \geqslant 0$:

$$\mathbb{P}(|f-\mathbb{E}f| \geqslant t) \leqslant 2\exp\left(-\frac{t}{D}+\sum_i\min\left(\log 6, 13\frac{(D_i/D)^2}{1-D_i/D}\right)\right)$$

(Les deux premières expressions sont égales à la frontière. La deuxième évaluation sert juste de zone de transition entre les domaines où la première et la dernière sont intéressantes.)

Illustrons le théorème par deux exemples opposés.

Si tous les D_i sont égaux, la première évaluation devient intéressante pour $t \approx D\sqrt{N}$ comme on s'y attendrait, et on a une allure gaussienne jusqu'à $t \approx ND/64$ où la probabilité décroît jusqu'à $\exp -N/16384$. La dernière évaluation devient non triviale pour $t \approx ND \log 6$, et ensuite, l'allure est en exponentielle $\exp(-t/D + N \log 6)$.

Inversement, si par exemple G_i est de diamètre observable $1/i$ (le premier facteur est beaucoup plus grand que les autres), on peut même se permettre de faire un produit infini. L'évaluation gaussienne est inintéressante et, pour tout t, la probabilité est en $\exp(-t + \log 6 + 13)$ (pour tout N !).

Enfin, remarquons que d'après la démonstration, si $t \ll \sum D_i^2/D$, le coefficient $1/4$ devant $t^2/\sum D_i^2$ dans l'évaluation gaussienne peut être remplacé par $1/2 + O(tD/\sum D_i^2)$ comme on s'y attend.

DÉMONSTRATION –

On a $\mathbb{P}(f - \mathbb{E}f \geqslant t) \leqslant \mathbb{E}e^{\lambda f - \lambda t}$ par l'inégalité de Tchebychev exponentielle.

Posons $\lambda = t/\sum D_i^2$. Si $t \leqslant \sum D_i^2/64D$, on a bien $\lambda \leqslant 1/64D$ et on peut appliquer le résultat précédent ; on obtient

$$\begin{aligned}
\mathbb{P}(f - \mathbb{E}f \geqslant t) &\leqslant \exp\left(-\lambda t + \frac{\lambda^2}{2}\sum D_i^2 + 13\lambda^3 \sum D_i^3/(1-\lambda D_i)\right) \\
&\leqslant \exp\left(-\frac{t^2}{2\sum D_i^2} + \frac{13}{(1-1/64)}\frac{t^3 \sum D_i^3}{(\sum D_i^2)^3}\right) \\
&\leqslant \exp\left(-\frac{t^2}{2\sum D_i^2} + \frac{13}{1-1/64}\frac{t^2}{\sum D_i^2}\frac{t\sum D_i^3}{(\sum D_i^2)^2}\right) \\
&\leqslant \exp\left(-\frac{t^2}{2\sum D_i^2} + \frac{13}{63}\frac{t^2}{\sum D_i^2}\frac{\sum D_i^3}{D\sum D_i^2}\right)
\end{aligned}$$

La première affirmation du théorème se déduit alors en observant que $\sum D_i^3 \leqslant D\sum D_i^2$.

Pour le second cas, on fait le même calcul avec $\lambda = 1/64D$. Le troisième cas est une combinaison des techniques du théorème 7 et du lemme 8. □

Un théorème similaire apparaît dans [ABS], pour des espaces métriques de la forme X^n où X est un espace *borné*. (Caractère qui intervient aussi de manière cruciale dans tous les théorèmes classiques à la Talagrand.) C'est la prise en compte du diamètre infini qui oblige à faire une hypothèse supplémentaire, comme celle de λ_1 minoré ($C > 0$). Cette hypothèse supplémentaire n'implique qu'une concentration exponentielle et non gaussienne, le caractère gaussien provenant du fait de considérer plusieurs variables indépendantes.

Note sur les constantes : on devrait plutôt énoncer le théorème en utilisant, à la place des D_i^2, les "constantes de dispersion" ("spread constant") introduites dans [ABS] comme le sup de la variance d'une fonction 1-lipschitzienne. La constante de dispersion est bien sûr majorée par D_i^2 si l'on part d'une inégalité de Poincaré.

Référence

[ABS] N. Alon, R. Boppana, J. Spencer, *An asymptotic isoperimetric inequality*, Geom. Funct. Anal. **8** (1998), 411–436.

Table des matières

III
Autour des algorithmes génétiques

Contenant :

Vitesse de convergence des opérateurs de croisement

Un algorithme génétique dans l'espace des arbres

La démographie du PRA

Vitesse de convergence des opérateurs de croisement

Résumé

Nous étudions la convergence de l'opérateur de croisement sur $\{0,1\}^n$, issu des algorithmes génétiques. En particulier, nous améliorons un résultat de Rabani, Rabinovich et Sinclair (cf. [RRS]) dans le cas d'une population finie ; nous donnons aussi une borne inférieure sur la divergence entre les processus à population finie et infinie, qui prouve que notre borne supérieure est proche de la valeur réelle.

Introduction, notations

On étudie ici d'un point de vue théorique la vitesse de convergence d'un opérateur de « croisement » de deux « génomes ». Le cadre est celui de la génétique des populations, ou bien des algorithmes génétiques : une population est composée d'individus déterminés par un génome, qui est une suite de symboles (que l'on prendra par commodité dans $\{0,1\}$).

L'opérateur de croisement consiste à supposer qu'une population est remplacée par une population-fille de la manière suivante : un individu de la population-fille est obtenu en tirant au hasard et indépendamment deux individus de la population précédente, et en mélangeant leurs génomes selon une règle prescrite. Cette opération est répétée indépendamment pour obtenir tous les individus de la population-fille.

Intuitivement, l'effet du croisement est de mélanger les gènes présents dans la population, et on lit généralement dans les manuels de biologie que l'intérêt de la reproduction sexuée est de maintenir une diversité supérieure à ce qu'elle serait par simple mutation. On peut donc vouloir évaluer la vitesse à laquelle cet effet se produit.

On se fixe donc une longueur de génome n, et on définit, de manière probabiliste, le croisement de deux éléments de $\{0,1\}^n$ de la manière suivante. On choisit une loi Π sur l'ensemble des parties de $\{1 \dots n\}$. On tire au hasard selon Π une partie $S \subset \{1 \dots n\}$; le croisement de $x, y \in \{0,1\}^n$ est alors un élément aléatoire $z \in \{0,1\}^n$ dont le i-ième bit z_i est égal à x_i si $i \in S$, et à y_i sinon.

Selon les distributions Π choisies, on peut obtenir différents types de croisement. Le plus simple est le croisement uniforme : Π est la distribution uniforme sur les parties de $\{0,1\}^n$, et cela revient à fixer chaque bit z_i égal à x_i ou y_i aléatoirement et indépendamment des autres avec probabilité $1/2$. On se tiendra essentiellement à ce choix de Π.

On prend un processus à population finie de taille k : à chaque instant, on a une population constituée de k individus de $\{0,1\}^n$, et la population à l'instant suivant est obtenue en tirant k fois de suite un couple d'éléments distincts dans cette population, en engendrant un enfant de ce couple et en plaçant cet enfant dans la population à l'instant suivant. (On peut aussi étudier le cas où chaque couple engendre deux enfants complémentaires.)

Soit π_t le k-uplet aléatoire de $\{0,1\}^n$ ainsi obtenu au temps t en partant d'un k-uplet initial π_0.

On cherche à comparer ce processus au processus dit « à population infinie » où on fait évoluer des lois de probabilité sur $\{0,1\}^n$: la loi d'un élément de la distribution p_{t+1} est obtenue en tirant selon la loi p_t deux individus et en les croisant. Étant donné p_0, on obtient ainsi une suite déterministe de lois de probabilité p_t sur $\{0,1\}^n$.

Le processus à population infinie p_t est assez bien connu (cf. [RRS]). Il converge vers une loi p_∞ dépendant de p_0 de la manière suivante : p_∞ est la loi où les bits d'un individu sont choisis indépendamment les uns des autres, chaque bit étant égal à 0 ou à 1 avec la même probabilité que dans p_0. Autrement dit, le processus conserve les proportions de 0 et de 1 à chaque position mais rend les positions indépendantes.

Ces auteurs donnent des bornes supérieures et inférieures essentiellement correctes concernant le processus à population infinie. Ces résultats sont rappelés dans la section 1.

En revanche, le processus à population finie est moins bien appréhendé. On peut le concevoir comme une approximation du cas de la population infinie ; mais il semble a priori que pour connaître un individu d'une certaine génération, il faut connaître ses deux parents, ses quatre grands-parents, etc., et que si la population est petite, des aïeux vont apparaître plusieurs fois dans l'arbre généalogique, ce qui introduit des corrélations non souhaitées.

Cette difficulté survient pour tous les systèmes dynamiques dits quadratiques, où l'on se donne un « croisement » probabiliste entre individus d'un certain espace donné, et où on fait évoluer une loi de probabilité sur cet espace en prenant pour loi d'un individu au temps $t+1$ la loi du croisement de deux individus tirés selon la loi au temps t. La difficulté de simuler un système dynamique quadratique a été formalisée (cf. [ARV]) : un tel système pourrait en effet résoudre en temps polynomial tout problème de classe PSpace (classe de problèmes solubles avec une mémoire polynomiale).

La comparaison entre les deux processus va comme suit : on se donne une population infinie p_0, et on tire k individus selon p_0, ce qui nous donne un k-uplet aléatoire π_0. On fait ensuite évoluer ce k-uplet, et on obtient ainsi un k-uplet π_t.

En fait, π_t vu comme une loi de probabilité sur $\{0,1\}^n$ n'est pas une bonne approximation de p_t. Ceci est du au phénomène de coalescence : au bout d'un certain temps, avec grande probabilité, π_t est constitué de k fois le même individu. En appliquant l'algorithme à population finie, on obtiendra donc très probablement une population-clone. (Cela démontre au passage que la loi de π_t converge.)

Par contre, l'individu aléatoire composant cette population ne sera pas toujours le même, et sa loi sera proche de p_t, ce que l'on souhaite. Ainsi, la loi d'un élément (par exemple le premier) de π_t, pris seul, est proche de la loi p_t.

Notons donc q_t la loi du premier élément du k-uplet aléatoire π_t.

Pour comparer des lois, on utilisera la distance de variation totale

$$|p - p'| = \frac{1}{2} \sum_{x \in \{0,1\}^n} |p(x) - p'(x)| = \sup_{X \subset \{0,1\}^n} |p(X) - p'(X)| \leqslant 1$$

1 La convergence du processus à population infinie

On rappelle ici les résultats de Y. Rabani, Y. Rabinovich et A. Sinclair (cf. [RRS]).

Soit p_0 une loi de probabilité sur $\{0,1\}^n$. Soit a_{i0} la probabilité que le i-ième bit d'un individu tiré selon p_0 soit égal à 0, et $a_{i1} = 1 - a_{i0}$.

On note alors p_∞ la loi qui attribue à un individu $x = x_1 x_2 \dots x_n$ la probabilité $p_\infty(x_1 x_2 \dots x_n) = \prod a_{ix_i}$. p_∞ est ainsi la loi où chaque bit vaut 0 ou 1 avec la même probabilité que dans p_0, mais où les bits sont choisis indépendamment les uns des autres.

On fait de plus une hypothèse de non-dégénérescence sur le croisement Π employé : il faut que quelles que soient les positions $1 \leqslant i, j \leqslant n$, ces positions aient une probabilité non nulle d'être séparées par le croisement (sinon, on peut traiter les deux positions i, j comme un seul bloc de deux gènes), i.e. il existe une partie $S \subset \{1 \dots n\}$ telle que $\Pi(S) > 0$ et à laquelle i appartient mais pas j (ou l'inverse).

Cette hypothèse naturelle est satisfaite par les croisements usuels. Les auteurs s'intéressent plus particulièrement à trois croisements :

- le croisement uniforme : Π est la loi uniforme sur les parties de $\{1 \dots n\}$, chaque bit de l'enfant est égal au bit correspondant d'un de ses parents indépendamment des autres ;
- le croisement à un point : on choisit une position $1 \leqslant i \leqslant n$ uniformément, et les bits de position inférieure à i sont identiques à ceux du parent 1, les autres à ceux du parent 2 ;

– le croisement de Poisson : on commence à la position 0, en copiant les bits du parent 1 ; on copie un bloc de bits consécutifs du parent 1, bloc dont la longueur suit une loi de Poisson (i.e. on arrête à chaque étape avec une probabilité fixée) ; puis on recopie un bloc de l'autre parent, de longueur suivant une loi de Poisson ; puis on recommence avec des bits du parent 1, etc., jusqu'à épuisement des bits.

Y. Rabani, Y. Rabinovich et A. Sinclair démontrent alors le

THÉORÈME 1 – *Le processus à population infinie p_t tend vers p_∞.* (en tant que loi de probabilité sur $\{0,1\}^n$).

De plus, une bonne évaluation de la convergence peut être obtenue. Elle dépend de l'opérateur de croisement utilisé. Suivant les notations de [RRS], soit $r_{ij}(\Pi)$ la probabilité qu'une partie $S \subset \{1 \dots n\}$ tirée selon Π ne sépare pas les positions i et j ; soit $r_\Pi = \max\limits_{i,j} r_{ij}(\Pi)$. L'hypothèse de non-dégénérescence équivaut à $r_\Pi < 1$.

Alors

THÉORÈME 2 – $|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$

Par exemple, pour le croisement uniforme, on a $r_\Pi = 1/2$ et donc $|p_t - p_\infty| \leqslant n^2/2^t$.

Si on définit le temps de relaxation τ du processus comme le temps nécessaire pour que, quelle que soit p_0, la distance $|p_t - p_\infty|$ soit inférieure à (par exemple) $1/4$, on voit que $\tau \lesssim 2 \log_{1/r_\Pi} n$ lorsque n est grand, ce qui est un bon résultat.

La dépendance en t donnée par ce théorème est correcte : en effet il existe une population p_0 telle que pour tout t, $|p_t - p_\infty| \geqslant r_\Pi^t/2$.

L'analyse menée par les auteurs de [RRS] pour obtenir des bornes inférieures sur le temps de relaxation τ dépend du détail du processus. Par exemple, pour le croisement uniforme, on a $\tau \geqslant \log_2 n - O(1)$; le résultat est donc correct à un facteur 2 près. Pour le croisement de Poisson, le résultat est correct à un facteur $O(\log\log\log n)$ près.

Nous donnons ici la preuve d'un résultat similaire mais légèrement différent concernant une borne inférieure de la convergence dans le cas du croisement uniforme, qui montre qu'on ne peut pas remplacer n^2 par une quantité plus petite que n dans l'évaluation de $|p_t - p_\infty|$ pour le croisement uniforme.

PROPOSITION 3 – *Dans le cas du croisement uniforme, pour n et t assez grands, pour une certaine population initiale p_0, on a*

$$|p_t - p_\infty| \geqslant \frac{n}{32\ 2^t}$$

« n et t assez grands » : par inspection de la démonstration, le résultat est valable dès que $n \geqslant 8$, $t \geqslant 3\log_2 n + 4$ (le temps à partir duquel la proposition est valable dépend forcément de n, sinon $n/(32\ 2^t)$ pourrait être supérieur à 1).

DÉMONSTRATION – Regardons autrement comment un individu de la génération t est engendré. Commençons par fixer les 2^t ancêtres de cet individu, tirés selon p_0. Ensuite, observons que, sous le croisement uniforme, chacun des n bits de cet individu provient de l'un de ces ancêtres, que l'on appellera ancêtre du bit en question. Dans le cas du croisement uniforme, les ancêtres de chaque bit sont choisis indépendamment et uniformément parmi les 2^t ancêtres de l'individu considéré (ceci est spécifique au croisement uniforme). Autrement dit, la distribution des ancêtres des n bits de l'individu considéré est un tirage indépendant, avec remise, de n individus parmi les 2^t ancêtres de l'individu considéré.

On va utiliser le fait que, parfois, deux bits proviennent du même ancêtre pour en tirer une déviation par rapport à la loi p_∞. Pour ce faire, on va prendre pour p_0 la loi de probabilité sur $\{0,1\}^n$ composée pour moitié de l'individu $111\ldots1$ et pour moitié de $000\ldots0$. On va regarder la loi du nombre de 1 d'un individu sous p_∞ et sous p_t et trouver une différence.

Sous p_∞, la loi du nombre de 1 est une binomiale de paramètres n et $1/2$.

Sous p_t, chaque bit d'un individu provient d'un de ses ancêtres au temps 0. Si les n ancêtres des n bits sont tous distincts, alors les bits en question sont tirés uniformément et indépendamment selon p_0, auquel cas on retrouve la même binomiale.

Si au contraire deux bits d'un individu ont le même ancêtre à la génération 0, vu notre population p_0, ces deux bits seront égaux. Ceci crée des corrélations qui résultent en une différence quantifiable dans la loi du nombre de 1 d'un individu.

On va d'abord évaluer la déviation qu'on obtient lorsqu'exactement deux bits ont le même ancêtre. Ensuite, on montrera qu'exactement deux bits ont le même ancêtre avec une probabilité assez grande, et que les cas où plus d'une corrélation se produit ont un poids négligeable quand t est grand. Le premier point fait l'objet du lemme suivant.

LEMME 4 – *Soit $n \geqslant 8$. Soit μ_1 la mesure de probabilité uniforme sur $\{0,1\}^n$. Soit μ_2 la mesure sur $\{0,1\}^n$ valant $1/2^{n-1}$ aux points $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ tels que $x_1 = x_2$ et 0 ailleurs. La distance entre μ_1 et μ_2 vérifie $|\mu_1 - \mu_2| \geqslant \dfrac{1}{2n}$.*

Plus précisément, l'écart entre les probabilités sous μ_1 et μ_2 de l'événement « le nombre de 1 de l'élément tiré est compris entre $n/2 - \sqrt{n/8}$ et $n/2 + \sqrt{n/8}$ » est supérieur à $1/2n$.

DÉMONSTRATION DU LEMME –

Sous μ_1, sans bits corrélés, la loi du nombre de 1 est une binomiale $\mathrm{C}_n^r/2^n$.

Sous μ_2, il y a un couple de bits corrélés, et la loi du nombre de 1 sera plutôt $\dfrac{1}{2}\mathrm{C}_{n-2}^{r-2}/2^{n-2} + \dfrac{1}{2}\mathrm{C}_{n-2}^{r}/2^{n-2}$ (pour les cas où le couple corrélé est composé de deux 1 et de deux 0, respectivement).

Cette probabilité est inférieure à la précédente dans une zone autour de $n/2$, et supérieure ailleurs. La différence entre les deux lois est $(\mathrm{C}_n^r - 2\mathrm{C}_{n-2}^{r-2} -$

$2\mathrm{C}_{n-2}^{r})/2^n$, soit, tous calculs faits, $\mathrm{C}_{n-2}^{r-1}/2^n \left(\frac{n-(n-2r)^2}{2(n-r)r}\right)$. La quantité en facteur est positive pour $n/2-\sqrt{n}/2 \leqslant r \leqslant n/2+\sqrt{n}/2$, elle vaut $2/n$ en $r=n/2$; elle est supérieure à $1/n$ pour $|r-n/2| \leqslant \sqrt{n/8}$.

Dans le cas où exactement deux bits ont le même ancêtre, la différence entre les probabilités, sous p_∞ et sous p_t, de l'événement « le nombre de 1 est compris entre $n/2-\sqrt{n/8}$ et $n/2+\sqrt{n/8}$ » est donc supérieure à

$$\frac{1}{n}\sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^n}\mathrm{C}_{n-2}^{r-1}$$

Sachant qu'une binomiale de paramètre $1/2$ est presque une gaussienne : $\sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^{n-2}}\mathrm{C}_{n-2}^{r-1} \sim \frac{2}{\sqrt{\pi}}\int_{-1/\sqrt{2}}^{1/\sqrt{2}} e^{-x^2/2}\mathrm{d}x \geqslant 1/2$ (valable en pratique dès que $n \geqslant 8$), on obtient que cette quantité est supérieure à $1/2n$, ce qui démontre le lemme. □

Observons que, si les n bits d'un individu de la génération t ont des ancêtres distincts au temps 0, la loi du nombre de 1 parmi ces n bits est la même que sous le μ_1 du lemme. Si exactement deux bits ont un ancêtre commun au temps 0, la loi du nombre de 1 est la même que sous le μ_2 du lemme.

Maintenant, on va tirer de ceci une évaluation sur la distance entre p_t et p_∞. Notons A_0 l'événement « tous les bits ont des ancêtres distincts », A_1 l'événement « exactement un couple de bits a un ancêtre commun », A_2 le reste (strictement plus d'une coïncidence). Notons également B l'événement « le nombre de 1 est compris entre $n/2-\sqrt{n/8}$ et $n/2+\sqrt{n/8}$ ».

Alors, d'après le lemme :

$$\begin{aligned}
|p_\infty - p_t| &\geqslant |p_\infty(B) - p_t(B)| \\
&\geqslant |(p_\infty(B) - p_t(B|A_0))\, p_t(A_0) + (p_\infty(B) - p_t(B|A_1))\, p_t(A_1)| \\
&\quad - |p_\infty(B) - p_t(B|A_2)|\, p_t(A_2) \\
&\geqslant 0 + \frac{1}{2n}p_t(A_1) - p_t(A_2)
\end{aligned}$$

(Sachant A_0, la loi du nombre de 1 sous p_t est la même que sous p_∞.)

Il s'agit donc désormais d'évaluer les probabilités qu'exactement deux bits, ou plus de deux bits, aient un ancêtre commun. On sait que les ancêtres de ces bits sont tirés uniformément et indépendamment parmi 2^t. Posons $k=2^t$. Nous posons le lemme suivant qui nous resservira.

LEMME 5 – *Si n individus (distinguables) sont placés uniformément parmi k places, la probabilité qu'exactement deux éléments occupent la même place est supérieure à $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$, ou encore, si $k \geqslant n^2$, à $\frac{n^2}{8k}$.*

DÉMONSTRATION DU LEMME – Par simple combinatoire, cette probabilité est égale à $\frac{1}{k^n}\frac{n(n-1)}{2}\,k(k-1)\dots(k-n+2)$, soit $\frac{n(n-1)}{2k}\,1\,(1-1/k)\dots(1-(n-2)/k)$, qui est supérieure à $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$. □

Ainsi, la probabilité qu'exactement deux bits d'un individu aient le même ancêtre au temps 0 est supérieure à $n^2/8k$, si on prend k assez grand. Le cas où on aurait plus de corrélations, à savoir au moins deux couples de bits ayant même ancêtre, ou bien au moins trois bits ayant le même ancêtre, a une probabilité majorée par n^4/k^2, ce qui est d'ordre supérieur en k. En effet on a en tout k^n possibilités de choisir n ancêtres parmi k ; le nombre de manières de les choisir tels que deux couples aient un ancêtre commun est inférieur à $(n(n-1)/2)^2\,k(k-1)k^{n-4}$; le nombre de cas où trois bits ont le même ancêtre est $(n(n-1)(n-2)/6)\,k^{n-2}$. Si l'on prend $k \geqslant 16n^3$ on s'assure ainsi que la probabilité de A_2 est inférieure à $n/32k$, auquel cas la quantité $1/(2n)\,p_t(A_1) - p_t(A_2)$, est supérieure à $n/32k$ comme annoncé. □

2 Comparaison entre les processus à population finie et infinie

On rappelle que q_t est la loi du premier élément du k-uplet aléatoire π_t obtenu au temps t par le processus à population finie. Dans toute la suite, on utilisera l'opérateur de croisement uniforme.

Dans [RRS], Y. Rabani, Y. Rabinovich et A. Sinclair prouvent que

$$|q_t - p_t| \leqslant \frac{4n^2t}{k}$$

Nous montrons ici, en utilisant des techniques similaires, que

THÉORÈME 6 –

$$|q_t - p_\infty| \leqslant n^2\left(\frac{1}{k}+\frac{1}{2^t}\right)$$

En particulier, $|q_\infty - p_\infty| \leqslant n^2/k$.

Si ce résultat est optimal est inconnu. On montre ci-dessous que $\lim |q_t - p_\infty| \geqslant n/Ck$ dans certains cas (C est une constante).

DÉMONSTRATION – On peut voir le processus π_t de la manière suivante : pour engendrer π_t, on laisse d'abord π_0 non spécifié, on choisit une structure généalogique menant de la génération 0 à la génération t, à laquelle on assigne la probabilité qui lui revient, et, tout à fait indépendamment, on remplit π_0 en tirant k individus suivant la loi p_0. On regarde ensuite comment les bits de la génération 0 se transmettent aux générations suivantes.

Plus précisément, on appelle structure généalogique la donnée, pour chaque $t \geqslant 1$ et pour chaque numéro d'individu i de la génération t, de deux numéros d'individus distincts i_1 et i_2 de la génération précédente, et d'un masque décrivant ceux des bits de i qui vont provenir de i_1 ou de i_2. La probabilité assignée à une telle structure est le produit des probabilités sous le croisement Π de tous les masques qui y apparaissent, divisée par $(k(k-1))^{kt}$ qui correspond aux choix des parents de tous les individus.

Une fois une généalogie tirée, on remplit aléatoirement, et indépendamment de cette généalogie, les bits de la génération 0 en suivant la loi p_0. Dans ces conditions, étant donné un bit d'un individu de la génération t, on est en mesure de dire de quel bit de quel individu de la génération 0 il provient, en remontant la généalogie.

On voit alors que si on a tiré une généalogie telle que les n bits du premier individu de la génération t proviennent tous d'individus différents de la génération 0, ces n bits proviennent de n individus tirés indépendamment selon p_0. Les valeurs 0 ou 1 de ces bits sont donc indépendantes, et le i-ième bit vaut 1 avec probabilité a_{i1} (notation ci-dessus). Autrement dit, si on a tiré une généalogie où les n bits du premier individu de π_t proviennent d'individus distincts, la loi de cet individu est exactement p_∞.

Le jeu de la définition de la distance $|q_t - p_\infty|$ montre qu'elle est donc inférieure à la probabilité que la généalogie tirée ne soit pas de ce type.

Évaluons-la. Considérons deux bits du premier individu de la génération t. Lorsqu'à une génération t', ces deux bits appartiennent à un même individu, la probabilité qu'ils proviennent du même parent de cet individu à la génération $t'-1$ est $1/2$, d'après nos lois de croisement. Lorsqu'ils appartiennent à deux individus différents de la population t', leurs parents respectifs sont tirés indépendamment dans $\pi_{t'-1}$, et la probabilité qu'ils proviennent du même individu de $\pi_{t'-1}$ est $1/k$.

On a donc une chaîne de Markov avec les probabilités de transition suivantes entre les deux états S (à une certaine date, ces deux bits appartiennent à deux individus séparés) et R (ces deux bits sont réunis dans un même individu) : $S \to S$ avec probabilité $1 - 1/k$, $S \to R$ avec probabilité $1/k$, $R \to S$ avec probabilité $1/2$, $R \to R$ avec probabilité $1/2$.

Tous calculs faits, et compte tenu du fait qu'à la génération t les deux bits sont réunis, on calcule que la probabilité de tirer une généalogie où ces deux bits sont réunis à la génération 0 est

$$\frac{2}{k+2} + \left(\frac{1}{2} - \frac{1}{k}\right)^t \left(1 - \frac{2}{k+2}\right) \leqslant \frac{2}{k} + \frac{1}{2^t}$$

Ceci pour un couple de bits du premier individu de la génération t. Il y a $n(n-1)/2$ tels couples. La probabilité pour qu'on ait tiré une généalogie où il existe deux bits de cet individu qui proviennent du même individu de la génération 0 est donc inférieure à $n(n-1)/2\,(2/k + 1/2^t)$, d'où le théorème. □

La même démonstration s'adapte sans changement à un opérateur de croisement Π différent, en tenant compte du paramètre r_Π défini à la section 1. Le résultat devient :

THÉORÈME 7 –

$$|q_t - p_\infty| \leqslant n^2 \left(\frac{1}{1 + k(1 - r_\Pi)} + r_\Pi^t \right)$$

En particulier, $|q_\infty - p_\infty| \leqslant n^2/(1 + k(1 - r_\Pi))$.

3 Différence entre populations finie et infinie

Il convient de remarquer que la différence entre les lois q_∞ et p_∞ n'est pas uniquement due au fait qu'on tirerait les gènes d'un individu de q_∞ dans le k-uplet π_0 avec remise. En effet, dans ce cas, la probabilité que deux gènes d'un individu de π_∞ proviennent du même individu de π_0 serait exactement $1/k$, or nous avons vu que c'est plutôt $2/(k+2)$ (pour le croisement uniforme), qui est supérieur.

Montrons une borne inférieure :

THÉORÈME 8 – *Pour tout $n \geqslant 2$, pour k assez grand, il existe une population initiale telle que*

$$|q_\infty - p_\infty| \geqslant \frac{n}{32k}$$

Le k assez grand à partir duquel la proposition sera valable dépend a priori de n (sinon $n/32k$ pourrait être plus grand que 1). En étudiant la preuve, on obtient facilement une évaluation très grossière : le résultat est valable au moins pour $k \geqslant 32(2n)^{n^2+2}/(1 - r_\Pi)^{n^2}$...

DÉMONSTRATION – On va considérer un individu au temps t, et on va regarder de quels individus de la génération 0 ses n bits proviennent. On va s'intéresser de près à la répartition de ces n individus.

On va tenir le même type de raisonnement que dans la partie 1 : on va montrer qu'avec une certaine probabilité de l'ordre de n^2/k, exactement deux bits ont un ancêtre commun, et que cela introduit une déviation de l'ordre de $1/n$.

On va d'abord évaluer la probabilité que deux bits exactement aient un ancêtre commun au temps 0. Cette probabilité est supérieure à la probabilité que deux bits exactement aient un ancêtre commun au temps 0 et que de plus, tous les bits aient été séparés au temps 1.

Dans la démonstration du théorème précédent, on a vu que la probabilité que deux bits d'un individu de π_∞ aient un ancêtre commun au temps 1 est inférieure à $\dfrac{n^2}{1 + k(1 - r_\pi)}$. Par conséquent, la probabilité qu'ils soient tous séparés au temps 1 est supérieure à $1 - \dfrac{n^2}{1 + k(1 - r_\pi)}$.

Maintenant, sachant que tous les bits sont séparés au temps 1, leurs parents au temps 0 sont simplement choisis uniformément et indépendamment parmi k. D'après le lemme 5, la probabilité que deux d'entre eux tombent ensemble est supérieure à $n^2/4k\ (1 - n^2/2k)$.

Par conséquent, la probabilité (inconditionnelle) qu'au temps 0, exactement deux bits aient un ancêtre commun est supérieure à

$$\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)\left(1-\frac{n^2}{1+k(1-r_\pi)}\right)$$

lui-même supérieur à $\dfrac{n^2}{8k}$ dès que k est assez grand, mettons $k \geqslant 2n^2/(1-r_\Pi)$.

Sous l'hypothèse qu'il existe deux bits ayant le même ancêtre, on va trouver une déviation entre les probabilités d'un événement sous p_∞ et sous q_t. On va prendre pour p_0, évidemment, la loi de probabilité sur $\{0,1\}^n$ composée pour moitié de l'individu $111\ldots1$ et pour moitié de $000\ldots0$. L'événement dont les probabilités sous q_t et p_∞ vont différer sera le nombre de 1 présents dans un individu de la génération t.

On va raisonner comme à la section 1. Pour cela, il faut d'abord établir que les cas où exactement deux bits d'un individu de la génération t ont le même ancêtre à la génération 0 est dominant sur les cas où on a plus de coïncidences. C'est l'objet du lemme suivant, qui établit que la distribution des n ancêtres (distincts ou non) des bits d'un individu a en gros la même asymptotique quand $k \to \infty$ que si ces ancêtres étaient choisis uniformément et indépendamment parmi les k individus de la population initiale.

En particulier, les cas où exactement deux bits ont un même ancêtre auront une probabilité de l'ordre de $1/k$ tandis que les cas où il se produit plus de coïncidences auront un poids d'un ordre inférieur à $1/k^2$. On mesure le nombre de coïncidences à l'aide du nombre d'individus distincts dont proviennent les n bits d'un individu de la génération t. Ce lemme peut avoir un intérêt indépendant.

LEMME 9 – *Il existe des constantes $C_{n,\Pi}$ et $C'_{n,\Pi}$ telles que la probabilité que les n bits d'un individu de la génération $t=\infty$ proviennent de m individus distincts de la population 0 est comprise entre $\dfrac{C_{n,\Pi}}{k^{n-m}}$ et $\dfrac{C'_{n,\Pi}}{k^{n-m}}$, pour k assez grand.*

(On se convainc aisément que parler d'un individu de la génération $t=\infty$ a bien un sens : le processus est markovien sur l'espace des populations à k individus.)

DÉMONSTRATION DU LEMME – Fixons un individu de la génération t, $t \approx \infty$. On a déjà vu que pour $m=n$, la probabilité que tous ses bits aient des ancêtres distincts au temps 0 est supérieure à $1-O(1/k)$, lorsque t est grand.

L'idée est de considérer la chaîne de Markov composée des positions (dans la population à k individus) des ancêtres des n bits de l'invididu considéré, au temps $t-t'$ (chaîne de Markov en t'). On va découper cette chaîne de Markov en classes, la classe m étant composée des situations où les n bits sont répartis

entre $m \leqslant n$ individus de la génération $t - t'$. On va considérer les coefficients de communication entre ces classes, et on va étudier le poids de ces classes à l'équilibre lorsque t' tend vers l'infini (plutôt vers t, mais on prend t très grand).

Notons $m(t')$ le nombre d'invididus distincts dont proviennent les n bits au temps $t - t'$, et $s(m)$ la probabilité que $m(\infty) = m$. On veut démontrer que $s(m) = O(1/k^{n-m})$. On sait déjà que pour $m < n - 1$, $s(m) = O(1/k)$.

Maintenant, estimons la distribution de $m(t' + 1)$ connaissant $m(t')$.

Pour passer de la génération $t - t'$ à $t - t' - 1$, on considère les $m(t')$ individus contenant les n bits. On décompose le processus en deux étapes. Dans la première, on considère les $m(t')$ blocs de bits, et on applique l'opérateur de croisement Π pour trouver $2m(t')$ parents les ayant engendrés. Parmi ces $2m(t')$, seuls m', où $m(t') \leqslant m' \leqslant n$, sont porteurs d'un bit. Dans la seconde étape, on « recolle » ces m' parents dans la population au temps $t - t' - 1$, qui comporte k individus. Le recollement consiste à choisir pour chacun des m' parents, de quel individu il s'agit parmi les k. Ces individus sont choisis indépendamment et uniformément parmi k (sauf qu'on doit faire attention au fait que deux parents d'un même individu de la génération $t - t'$ sont distincts, auquel cas on choisit parmi $k - 1$ seulement, ce qui ne change pas grand-chose).

La probabilité que ces m' parents soient répartis sur $m'' \leqslant m'$ individus de la génération $t - t' - 1$ est en $C_{m'}/k^{m'-m''}$ pour k grand. Maintenant, connaissant $m(t')$, on sait qu'on a $m' \geqslant m(t')$ et que de plus, si $m(t') < n$, alors avec probabilité supérieure à $1 - r_\Pi$, on a $m' > m(t')$.

Autrement dit, la première de nos deux étapes ne peut pas réduire $m(t')$, et l'augmente avec une probabilité supérieure à $1 - r_\Pi$ (si $m(t) < n$) ; la seconde réduit le résultat obtenu avec une probabilité contrôlée, passant de m' à m'' avec une probabilité en $O(1/k^{m'-m''})$. Au total, $m(t' + 1) < m(t')$ avec une probabilité en $O(1/k^{m(t')-m(t'+1)})$, $m(t' + 1) = m(t')$ avec une probabilité inférieure à $r(\pi) + O(1/k)$, et $m(t' + 1) > m(t')$ dans les autres cas : en général, le nombre de blocs de bits augmente, et il n'est réduit qu'avec des probabilités contrôlées par des puissances de k.

Passons à la preuve proprement dite. On raisonne par récurrence descendante sur m.

Supposons qu'on ait démontré que pour tous les $m' \leqslant m$, on a $s(m') = O(1/k^{n-m})$, et que pour $m \leqslant m' \leqslant n$ on a $s(m') = O(1/k^{n-m'})$. Maintenant, la probabilité $s(1)$ qu'au temps 0 ($t' \approx \infty$), tous les bits soient regroupés, vérifie $s(1) \leqslant r_\Pi\, s(1) + O(1/k)\, s(2) + O(1/k^2)\, s(3) + \cdots + O(1/k^{n-1})\, s(n)$ (à l'équilibre). D'après notre hypothèse de récurrence, et comme $r_\Pi < 1$, ceci est en $O(1/k^{m+1})$.

De même $s(2) \leqslant s(1) + r_\Pi\, s(2) + O(1/k)\, s(3) + \cdots + O(1/k^{n-2})\, s(n)$, qui est en $O(1/k^{m+1})$, par hypothèse de récurrence et puisque $r_\Pi < 1$.

De proche en proche jusqu'à $m' = m-1$, on obtient ainsi que pour $m' \leqslant m-1$, on a $s(m) = O(1/k^{n-m+1})$, ce qui clôt notre récurrence et démontre la borne supérieure dans le lemme.

Si l'on veut surveiller la dépendance en Π, on doit ajouter un facteur $1/(1 - r_\Pi)$ qui provient de la forme $s(m) \leqslant r_\Pi\, s(m) + \cdots$ de nos équations. Comme

on a fait moins de n^2 usages de cette équation, apparaît le facteur $1/(1-r_{\Pi})^{n^2}$ annoncé.

Pour démontrer la borne inférieure du lemme, il suffit de constater que $s(n) = 1 - O(1/k)$ et de voir que les coefficients de transition $n \to m$ à partir de l'état $m(t') = n$ sont de l'ordre de $1/k^{n-m}$. □

On a démontré que d'une part, exactement deux bits avaient un ancêtre commun avec une probabilité supérieure à $n^2/8k$; d'autre part, que la situation où strictement plus d'un couple de bits avaient un ancêtre commun avait une probabilité en $O(1/k^2)$. Il suffit alors d'appliquer le lemme 4 pour conclure. □

4 Vitesse de coalescence en population finie

Nous rappelons ici le résultat très classique que dans une population finie, la diversité génétique ne peut que décroître, au point d'obtenir une population-clone.

PROPOSITION 10 – *Pour tout $\varepsilon > 0$, pour*

$$t \geqslant 4k\,(\ln n - \ln \varepsilon + \ln 2)$$

alors le k-uplet π_t est constitué de k fois le même élément avec probabilité supérieure à $1 - \varepsilon$.

Si k est assez grand, ce temps est donc grand devant le temps de convergence vers l'équilibre (qui est de l'ordre de $2 \log n$).

DÉMONSTRATION – Soit $1 \leqslant a \leqslant n$ une position sur le génome. Soit, pour un certain t, i le nombre d'éléments de π_t qui contiennent un 1 à la position a. On va calculer la loi du nombre d'éléments de π_{t+1} contenant un 1 en position a. La probabilité qu'un enfant d'un couple tiré au hasard dans π_t ait un 1 en position a est i/k. Connaissant π_t, la probabilité que j individus de π_{t+1} aient un 1 en position a est donc $\mathrm{C}_k^j (i/k)^j (1 - i/k)^{k-j}$.

Le nombre d'individus de π_t ayant un 1 en position a constitue ainsi une chaîne de Markov sur $\{0 \ldots k\}$ dont la probabilité de transition est $p_{ij} = \mathrm{C}_k^j (i/k)^j (1 - i/k)^{k-j}$. (C'est le modèle de Wright-Fisher.) 0 et k y sont absorbants et tous les éléments de $\{1 \ldots k-1\}$ sont transients. Avec probabilité 1, on atteint donc 0 ou k.

Soit $T_{ki} < \infty$ l'espérance du temps d'absorption sachant que la chaîne démarre à $i \in \{0 \ldots k\}$, et $T_k = \sup_i T_{ki}$. D'après l'inégalité de Markov, la probabilité de ne pas être absorbé au temps $2T_k$ est inférieure à $1/2$. Comme la chaîne est markovienne, la probabilité de ne pas être absorbé au temps t est inférieure à $1/2^{\lfloor t/(2T_k) \rfloor} \leqslant 2/2^{t/(2T_k)}$.

Considérons maintenant simultanément l'ensemble des positions $1 \leqslant a \leqslant n$. La probabilité que la chaîne de Markov associée à l'une des positions ne soit pas absorbée au temps t est inférieure à $n\,2/2^{t/2T_k}$. Que toutes les chaînes soient

absorbées signifie que dans la population π_t, tous les individus ont les mêmes allèles 0 ou 1 aux mêmes endroits, i.e. que tous les individus sont identiques.

Ceci démontre la proposition quand on connaît le résultat, classique dans le modèle de Wright-Fisher (voir par exemple [Ewe]), que $T_k \sim 2k \ln 2$. □

Le même raisonnement est valable (avec des probabilités de transition différentes) si chaque couple engendre deux enfants par répartition aléatoire des allèles des parents entre les deux enfants.

En corollaire, si on sait que l'unique individu composant π_t a une loi proche de p_∞, on voit en revanche que le k-uplet obtenu nous donne une bien mauvaise image d'un k-uplet tiré selon p_∞ :

COROLLAIRE 11 – *Pour $t \geqslant 4k(\ln n - \ln \varepsilon + \ln 2)$, la distance $\left|\text{loi de } \pi_t - p_\infty^{\otimes k}\right|$ est supérieure à*

$$1 - \varepsilon - \prod_{1 \leqslant i \leqslant n} \left(a_i^k + (1 - a_i)^k\right)$$

On rappelle que a_i est la proportion d'individus de la population initiale ayant un 1 en position i du génome.

DÉMONSTRATION – En effet, $\prod_{1 \leqslant i \leqslant n} \left(a_i^k + (1 - a_i)^k\right)$ est la masse que $p_\infty^{\otimes k}$ donne aux k-uplets composés d'éléments tous identiques. □

5 Approximation d'une population à temps moyen

Ainsi on voit qu'aux temps grands, le k-uplet aléatoire π_t n'est pas une bonne image d'un k-uplet tiré selon p_t ou p_∞, pour des raisons de coalescence. Cependant, quand k est grand, le temps caractéristique de coalescence est beaucoup plus grand que le temps caractéristique de convergence de q_t vers p_∞. On peut donc supposer qu'aux temps moyens, on peut extraire de π_t plusieurs individus, mettons $m \leqslant k$ individus, dont la loi jointe serait proche de $p_\infty^{\otimes m}$.

C'est bien le cas :

THÉORÈME 12 – *Soit $m \leqslant k$. Soit q_t^m la loi jointe dans $(\{0,1\}^n)^m$ des m premiers individus du k-uplet π_t. Alors*

$$\left|q_t^m - p_\infty^{\otimes m}\right| \leqslant \frac{m^2 n^2}{1 + k(1 - r_\Pi)} + \frac{m^2 n}{k}\, t + mn^2\, r_\Pi^t$$

Bien évidemment, on peut remplacer « les m premiers individus » par n'importe quel m-uplet choisi a priori du k-uplet d'individus.

Le premier terme correspond au biais intrinsèque dû à la population finie, même aux temps longs, déjà étudié ci-dessus. Le deuxième terme correspond à la coalescence. Le troisième terme traduit la convergence vers p_∞.

Notons qu'on doit prendre k de l'ordre de $(mn)^2$ si l'on veut une évaluation non triviale.

L'optimum en t est atteint pour $t \approx \log_{1/r_\Pi} \frac{nk}{m}$ et est de l'ordre de

$$\frac{m^2n^2}{1+k(1-r_\Pi)} + \frac{nm^2}{k} \log_{1/r_\Pi} \frac{nk}{m}$$

On pourrait, en utilisant les mêmes techniques que précédemment (évaluer la loi du nombre de 1 parmi les mn bits dans le cas où deux des bits ont le même ancêtre), obtenir une borne inférieure qui serait la même à un facteur $1/mn$ (et des constantes) près, valable à t fixé pour k assez grand.

DÉMONSTRATION – On va considérer l'ascendance des mn bits des m premiers individus de π_t. Si ces mn bits proviennent d'individus tous différents de π_0 (ce qui impose $mn \leqslant k$), alors la distribution obtenue sera celle de $p_\infty^{\otimes m}$.

Considérons donc deux bits parmi ces mn. Si ces deux bits sont deux bits différents d'un même individu, rien ne change par rapport à l'étude précédente, et la probabilité qu'ils ne soient pas séparés dans π_0 est inférieure à $\frac{1}{1+k(1-r_\Pi)} + r_\Pi^t$.

Si ce sont deux bits situés à des positions différentes dans deux individus différents de π_t, alors la chaîne de Markov décrivant leur séparation est la même, mais comme au départ ils sont séparés, on peut retirer le terme r_Π^t : la valeur propre dominante est $r_\Pi - 1/k$, et partant de 0 la probabilité qu'ils ne soient pas séparés tend géométriquement vers $\frac{1}{1+k(1-r_\Pi)}$; la probabilité qu'ils soient séparés dans π_0 est donc inférieure à $\frac{2}{1+k(1-r_\Pi)}$.

Par contre, la situation est bien différente si l'on considère deux bits situés à la même position dans deux individus de π_t : en effet, si au cours de l'ascendance ces deux bits sont réunis dans un même individu, c'est qu'il s'agit en fait du même bit hérité de cet individu. Alors en remontant dans la généalogie, ces deux bits seront donc définitivement réunis jusqu'à l'origine dans π_0.

Ces deux bits étant fixés, ceci se produit à chaque génération avec une probabilité $1/k$; la probabilité que cela se produise en t générations est donc inférieure à t/k (et cette évaluation est correcte pour k grand).

On a $mn(n-1)/2$ couples de bits différents d'un même individu ; $m(m-1)n(n-1)/2$ couples de bits de positions différentes dans deux individus différents ; et $m(m-1)n/2$ couples de bits situés à la même position dans deux individus différents ; d'où le résultat. □

Conclusion

La convergence de l'opérateur de croisement vers la population d'équilibre est donc assez bien contrôlée, en population finie comme infinie.

L'étape suivante de ce travail, si on veut aller dans la direction des algorithmes génétiques ou de la théorie de l'évolution, consiste à étudier comment l'ajout d'un mécanisme de sélection modifie ces résultats.

Cependant, même en restant dans ce cadre, demeure le problème de la qualité des estimations : tout au long du développement, que ce soit à population finie ou infinie, on a obtenu des bornes supérieures dépendant du carré du nombre de bits, et des bornes inférieures qui ont la même dépendance en temps mais où ce carré est simplement remplacé par le nombre de bits. Même dans le cas le plus simple du croisement uniforme en population infinie, on ne sait pas quelle est l'évaluation correcte.

Références

[RRS] Y. Rabani, Y. Rabinovich et A. Sinclair, *A computational view of population genetics*, Random Structures and Algorithms **12** (1998), No. 4, 313–334.

[ARV] S. Arora, Y. Rabani, U. Vazirani, *Simulating quadratic dynamical systems is PSpace-complete*, Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing (1994), 459–467.

[Ewe] W. J. Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin (1979).

Table des matières

Un algorithme génétique dans l'espace des arbres

Résumé

Nous présentons une méthode de recherche d'arbres phylogénétiques parcimonieux à l'aide d'un algorithme génétique sur l'espace des arbres, utilisant des croisements de deux arbres. La contribution principale est une définition de mélange probabiliste de deux arbres (ayant des feuilles marquées par le même ensemble) produisant un arbre ressemblant à ses deux parents.

1 Introduction et notations

1.1 Le problème

Le problème de la reconstitution phylogénétique est le suivant : étant donné différents êtres vivants dont on connaît certaines caractéristiques anatomiques ou génétiques, il s'agit de deviner quel a été, le plus vraisemblablement, l'arbre évolutif menant à ces créatures. Pour une introduction aux problèmes mathématiques sous-jacents, on pourra consulter le tutoriel [KW].

Un premier sous-problème, de nature biologique, consiste à rassembler des données comparables entre les différentes espèces (par exemple, à repérer les gènes homologues et à aligner les séquences génétiques correspondantes). Un deuxième problème, de nature combinatoire, consiste à effectuer ensuite une recherche efficace parmi tous les arbres évolutifs possibles entre les espèces, pour trouver le plus « vraisemblable » suivant un certain critère. Choisir ce critère de jugement des différents arbres (ce qui revient à se donner un modèle de l'évolution) constitue en soi un troisième problème.

Nous nous intéressons ici au second problème : l'exploration efficace de l'espace des arbres possibles entre des espèces données (le nombre d'arbres à N feuilles ayant une croissance surexponentielle en N, un examen individuel de tous les arbres possibles est exclu). Nous supposons que le premier problème a été résolu par la connaissance biologique du système. Quant au troisième, nous prendrons comme critère de jugement d'un arbre le critère dit de *parcimonie*, qui consiste à adopter comme plus vraisemblable l'arbre qui minimise le nombre de

mutations à effectuer entre les différentes espèces (voir la définition ci-dessous). Ce choix de la parcimonie est contestable, mais il est couramment retenu pour sa simplicité (voir [RW] pour une discussion de ce choix).

1.2 Les données et le but

Les données sont donc les suivantes : un nombre N d'espèces, et, pour chacune, un nombre k de caractères à valeurs dans un espace métrique (X, dist) qui est, en général, $\{0, 1\}$ (pour des données morphologiques) ou bien $\{A, C, T, G\}$ (pour des données génétiques), avec la métrique discrète. Autrement dit, on se donne N suites $s^1, \ldots, s^N$ de k éléments $s^i_j \in X$ ($1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant k$), qu'on appellera *génome* des espèces étudiées (ce qui est bien sûr un abus de langage dans le cas de données morphologiques).

On cherche donc un arbre $\mathcal{A}$ à N feuilles, chaque nœud interne et chaque feuille $f \in \mathcal{A}$ étant décoré par une suite $s^f \in X^k$ (qu'on appellera *génome* du nœud), de sorte que les décorations des N feuilles de l'arbre soient les N suites $s^1, \ldots, s^N$ codant les caractères des espèces étudiées.

Suivant une convention courante en phylogénie, *tous nos arbres seront enracinés* (l'arbre de l'évolution l'est), la racine étant au voisinage immédiat d'une des espèces données, spécifiée par l'utilisateur (éventuellement une espèce hypothétique, « *outgroup* »).

On définit la distance entre deux suites $s, s' \in X^k$ par

$$\mathrm{dist}(s, s') = \sum_{j=1}^{k} \mathrm{dist}(s_j, s'_j)$$

On définit la longueur totale de l'arbre comme la somme, sur toutes les arêtes de l'arbre, de la distance entre les génomes des deux extrémités de l'arête :

$$\ell(A) = \sum_{\{f,f'\} \text{ arête de } \mathcal{A}} \mathrm{dist}(s^f, s^{f'})$$

Étant donné les génomes des espèces étudiées, le but sera de trouver un arbre décoré $\mathcal{A}$ de longueur minimale. La complexité du problème tient à l'explosion combinatoire du nombre d'arbres possibles avec N.

1.3 L'approche

Le problème se résume à trouver la topologie de l'arbre : en effet, étant donné un arbre, on sait comment trouver efficacement les génomes des nœuds internes minimisant la longueur de l'arbre (voir [H]).

Différentes heuristiques sont utilisées pour ne parcourir qu'une portion « intéressante » de l'espace des arbres. Les techniques usuelles consistent en la construction d'un arbre initial par des méthodes revenant à regrouper les plus

proches voisins, d'une manière ou d'une autre. Puis on applique diverses modifications à cet arbre, par exemple en détacher un sous-arbre pour le rattacher à un autre endroit et regarder si cela diminue la longueur totale.

Notons qu'il n'existe aucune garantie que ces heuristiques produisent bien un arbre de longueur minimale. Elles semblent cependant donner de bons résultats au vu des arbres qu'un humain aurait tendance à produire.

Notre approche s'assimile à un algorithme génétique. On maintient une « population » d'arbres qu'on fait évoluer de génération en génération par un mécanisme de sélection naturelle : plus la longueur d'un arbre est faible, plus la probabilité que cet arbre laisse des descendants à la génération suivante est élevée. Nous utilisons bien sûr des « mutations », qui sont identiques aux modifications que les approches habituelles font subir à chaque génération à l'arbre qu'elles étudient (par exemple, déplacer un sous-arbre). Mais nous utilisons surtout un algorithme génétique à reproduction sexuée : chaque arbre d'une génération est un mélange probabiliste (décrit plus bas) de deux arbres de la génération précédente. Cet aspect de mélange de différents arbres est absent des algorithmes les plus utilisés à ce jour.

Disons d'emblée que les résultats sont plutôt bons, mais cependant non comparables (en qualité mais surtout en temps de calcul, le coût du maintien d'une population entière d'arbres étant lourd) à ceux des logiciels habituels. Deux explications à cela : il est tout à fait possible que les avantages du croisement de deux arbres soient inférieurs à l'inconvénient du surcoût de calcul ; mais, par ailleurs, il est difficile de comparer un logiciel amateur écrit en trois mois avec des logiciels commerciaux développés et optimisés sur des années.

L'idée de croisement d'arbres pourrait donc peut-être avantageusement, avec plus de travail, être intégrée aux logiciels existants, bien que le programme ici présenté ne puisse pas à l'heure actuelle rivaliser avec ceux-ci.

2 Description du programme

Après la lecture des données, le programme produit une population initiale de m arbres selon un procédé décrit plus loin ; puis cette population évolue de génération en génération ; à chaque génération, le (ou les) meilleur arbre trouvé jusqu'alors est affiché. Le programme s'arrête lorsqu'un certain seuil est atteint.

Une nouvelle génération de la population d'arbres survient de la manière suivante. On part d'une génération de m arbres $\mathcal{A}_1, \ldots, \mathcal{A}_m$. Puis on opère comme suit (les détails sont donnés ci-dessous) :

- En utilisant un opérateur de sélection, on attribue à chaque arbre $\mathcal{A}_i$ un poids p_i (avec $p_i \geqslant 0$ et $\sum p_i = 1$). Plus court est l'arbre, plus grand est p_i.
- On crée une population fille $\mathcal{A}'_1, \ldots, \mathcal{A}'_m$ ainsi : m fois de suite indépendamment, on tire au hasard deux « parents » $\mathcal{A}_{i_1}$ et $\mathcal{A}_{i_2}$ dans la population $\mathcal{A}_1, \ldots, \mathcal{A}_m$, en utilisant les probabilités p_i ; puis on place un nouvel arbre $\mathcal{A}' = \mathcal{A}_{i_1} * \mathcal{A}_{i_2}$ où $*$ est un opérateur (aléatoire) de croisement des arbres.
- Chacun des arbres $\mathcal{A}'_i$ subit ensuite un opérateur de mutation aléatoire.

La nouvelle population $(\mathcal{A}'_i)$ remplace l'ancienne ; on évalue la longueur des arbres obtenus, et on répète l'opération autant que souhaité.

Cet algorithme dépend de beaucoup de paramètres : la taille de la population, la construction de la population initiale, la spécification des opérateurs de sélection, de croisement, de mutation, la condition d'arrêt.

2.1 Comment les arbres sont sélectionnés

Les probabilités de sélection p_i sont calculées par la formule

$$p_i = \exp -\beta \ell(\mathcal{A}_i)$$

où $\ell(\mathcal{A})$ est la longueur (nombre de mutations) de l'arbre $\mathcal{A}$, et où β est une constante.

La valeur de β est déterminée de la manière suivante : intuitivement, pour que $\beta\ell(A_i)$ ait un sens, β doit être l'inverse d'un nombre de mutations. L'effet de la sélection doit déjà se faire sentir lorsque, dans l'arbre, on change l'ordre de branchement de trois feuilles voisines ; si ces feuilles sont « quelconques » on s'attend à ce que la variation de la longueur de l'arbre soit du même ordre de grandeur que la distance moyenne entre les génomes de deux feuilles. On pose donc

$$\frac{1}{\beta} = \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \operatorname{dist}\left(s^{i_1}, s^{i_2}\right)$$

avec les notations ci-dessus.

Insistons sur le fait que la valeur de β est extrêmement importante pour le fonctionnement du programme, et que ce raisonnement par homogénéité a donné des résultats étonnants. Lors des premiers essais, une valeur de β estimée à la main avait été utilisée, et le programme fonctionnait mal. De l'instant où β a été déterminé selon cette règle, le programme s'est mis à donner des résultats très satisfaisants.

Comme dans beaucoup d'algorithmes d'optimisation utilisant une température, il a semblé utile de faire décroître la température au cours du temps (cf. le recuit simulé). L'effet est léger, mais on a cru observer empiriquement qu'utiliser, à l'étape t, la valeur $\beta\sqrt{t}$ plutôt que β (i.e. faire décroître la température en $1/\sqrt{t}$) donnait des résultats légèrement meilleurs.

On aurait pu utiliser des variantes sur le même principe d'homogénéité, comme, par exemple, prendre à l'étape t une température égale à la longueur moyenne des arêtes des arbres de la population, ou à la longueur moyenne des arêtes du meilleur arbre, etc. (l'effet de recuit simulé apparaîtrait alors spontanément). Par manque de temps ces variantes n'ont pas été testées.

2.2 Comment les arbres font des enfants

Comment mélanger deux arbres ayant le même ensemble de feuilles de manière à obtenir un arbre enfant « ressemblant » à ses parents ? Un arbre (en-

raciné) sur un ensemble donné de feuilles est uniquement caractérisé par la donnée d'une liste de nœuds, un nœud étant lui-même défini par l'ensemble des feuilles placées sous ce nœud dans l'arbre. Ainsi l'arbre (enraciné) sur les quatre espèces A, B, C, D, dans lequel se séparent deux groupes, l'un contenant A et B, l'autre contenant C et D, peut être décrit par l'ensemble de nœuds $\{\{A, B, C, D\}, \{A, B\}, \{C, D\}, \{A\}, \{B\}, \{C\}, \{D\}\}$ (on peut éventuellement omettre le nœud racine contenant toutes les feuilles, ainsi que les feuilles elles-mêmes, pour écourter la description).

Inversement, un ensemble E de nœuds définit bien un arbre si et seulement si pour tous $A, B \in E$, on a soit $A \subset B$, soit $B \subset A$, soit $A \cap B = \varnothing$ (condition de compatibilité).

L'ensemble des arbres sur un ensemble donné de feuilles possède alors une relation d'ordre naturelle : celle de l'inclusion des nœuds, un arbre étant plus fin qu'un autre si son ensemble de nœuds contient celui de l'autre.

Qui plus est, cette relation d'ordre admet des bornes inférieures. Si les arbres $\mathcal{A}_i$ sont définis par les ensembles de nœuds E_i, l'inf de la famille d'arbres $\mathcal{A}_i$ est l'arbre « consensus strict » défini par l'intersection des E_i (cette intersection vérifie bien la condition de compatibilité ci-dessus).

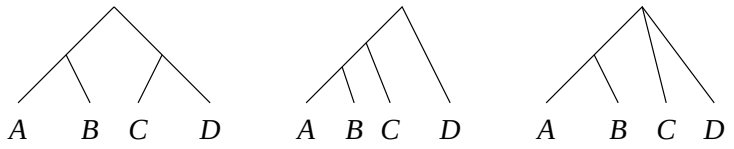


FIG. 1 – Deux arbres et leur borne inférieure

Cet inf semblerait un bon candidat pour le croisement de deux arbres. L'un de ses inconvénients est que l'inf de deux arbres a tendance à avoir un grand nombre de nœuds dégénérés (ayant plus de deux nœuds-fils). L'analogue pour un algorithme génétique standard opérant sur des suites de 0 et de 1 serait de ne conserver lors du croisement que les loci où les génomes des deux parents sont en accord, et de ne rien spécifier pour les autres (on imagine la catastrophe pour des êtres vivants en chair et en os). Habituellement, le croisement conserve bien sûr les gènes communs aux deux parents, mais, en cas de désaccord, complète avec l'un ou l'autre de manière plus ou moins aléatoire.

On peut définir un croisement aléatoire de deux arbres de la manière suivante. On part de deux arbres $\mathcal{A}$ et $\mathcal{A}'$ décrits par des ensembles de nœuds E et E'. On forme la réunion $E'' = E \cup E'$; en général cette réunion ne définit pas un arbre car elle ne vérifie pas la condition de compatibilité. On ordonne au hasard les éléments de E''. Puis on les parcourt un à un selon cet ordre ; et, dès qu'un élément ne satisfait pas la condition de compatibilité avec les éléments déjà parcourus, on le supprime. Arrivé au dernier élément, on est certain d'avoir un ensemble de nœuds E''' vérifiant la condition de compatibilité ; il définit donc bien un arbre.

Ce croisement est une opération aléatoire (comme il est d'usage pour les algorithmes génétiques) car le résultat dépend en général de l'ordre (aléatoire) choisi pour parcourir les éléments de E''. Il vérifie la propriété naturelle que $E \cap E' \subset E''' \subset E \cup E'$. Il est bien sûr idempotent (le croisement de $\mathcal{A}$ avec $\mathcal{A}$ est $\mathcal{A}$ avec probabilité 1) et semble constituer une bonne définition d'un « mélange » probabiliste de deux arbres. Les arbres qui ont une probabilité non nulle d'être obtenus par cette méthode semblent liés aux arbres qui se trouvent sur une géodésique joignant deux arbres dans l'espace des arbres métriques décrit dans [BHV], à condition de remplacer, dans cet espace, la distance euclidienne sur chaque orthant par la distance L^1.

2.3 Comment les arbres mutent

L'opérateur de mutation d'un arbre se compose de plusieurs opérations successives. Elles dépendent d'un taux de mutation τ.

Tout d'abord, chaque nœud de l'arbre est « supprimé » avec probabilité τ. Supprimer un nœud revient à attacher directement tous les nœuds-fils de ce nœud à son nœud-père, ou, dans la description de l'arbre par la liste des contenus de ses nœuds, à supprimer le nœud en question. Cette opération augmente la dégénérescence de l'arbre.

Une fois qu'un nœud a été sélectionné (avec probabilité τ) pour être supprimé, si ce nœud a n fils, chaque fils est lui-même supprimé avec probabilité $1/n$; et, pour chaque fils ainsi supprimé, on parcourt les n' fils de ce fils, chacun étant supprimé avec probabilité $1/n'$; etc., jusqu'à ce que les probabilités ne suppriment plus aucun fils (ce qui se produit certainement, puisque ces probabilités modélisent un processus de branchement de moyenne 1).

Ensuite, pour compenser ces dégénérescences, on effectue sur chaque nœud ayant au moins trois fils l'opération suivante : tant qu'il reste au moins trois nœuds-fils, on en prend deux au hasard, uniformément, et on les regroupe. Cela revient à résoudre la dégénérescence par un arbre binaire choisi au hasard.

Les mutations effectuées jusqu'ici sont locales.

Éventuellement, à cette étape, la racine de l'arbre n'est plus la feuille spécifiée comme racine au départ. On ré-enracine l'arbre dans ce cas.

Ensuite, on calcule les génomes des nœuds internes de l'arbre qui minimisent la longueur de l'arbre. Ceci sert pour la suite.

Afin de ne pas conserver de structure non pertinente, on écrase toute arête de l'arbre qui serait de longueur 0.

Enfin, on effectue des mutations « globales » : p fois de suite, on sélectionne un nœud de l'arbre au hasard. Avec probabilité $1 - \tau$, on ne fait rien. Sinon, on détache le sous-arbre $\mathcal{A}'$ attaché à ce nœud. On choisit un nœud aléatoire x dans $\mathcal{A}'$. On cherche un nœud y de l'arbre original dont le génome soit le plus proche du génome de x. Puis on cherche le nœud y' de $\mathcal{A}'$ dont le génome est le plus proche de celui de x. On ré-enracine $\mathcal{A}'$ en y', obtenant un nouvel arbre $\mathcal{A}''$. On rattache cet arbre $\mathcal{A}''$ à l'arbre original sous le nœud x.

On choisit les valeurs de τ et p comme suit. Si l'arbre considéré comprend n nœuds (avec $N+1 \leqslant n \leqslant 2N-1$), le nombre moyen de mutations locales sera $n\tau$. À l'approche du meilleur arbre, la plupart des mutations sont mauvaises et seules quelques-unes sont bonnes. On prend donc $\tau = 1/2N$, de sorte qu'il y a environ une mutation locale par arbre. Ainsi une bonne mutation ne se verra pas perdue à cause d'une mauvaise survenue dans le même arbre. Pour ce qui est de p, on fait le même genre de raisonnement : le nombre de mutations globales par arbre doit être d'environ 1. Sachant que $\tau = 1/2N$ on prend $p = 2N$.

Par ailleurs, dans certains algorithmes génétiques le taux de mutation est une propriété des individus eux-mêmes. Nous avons donc tenté de favoriser des taux de mutation variables (en assignant au départ le taux τ à tous les individus, puis en modifiant au hasard à chaque pas le taux de mutation d'un individu d'un facteur 1.1 en plus ou en moins au hasard, le taux de mutation du croisement de deux individus étant la moyenne géométrique de ceux des parents). Les taux de mutation n'ont pas semblé varier beaucoup, et aucune amélioration de performance n'a été perceptible.

2.4 Genèse de la population initiale et choix de sa taille

La population initiale est obtenue de la manière suivante : m fois de suite (m étant la taille de la population), on construit d'une part un arbre au hasard, d'autre part un arbre utilisant un algorithme « naïf », puis on croise ces deux arbres en utilisant l'opérateur de croisement décrit plus haut.

L'arbre au hasard est construit de la manière suivante : on initialise une liste d'arbres contenant, au départ, N arbres d'une seule feuille correspondant aux N espèces à analyser. Puis, $N-1$ fois, on choisit au hasard deux arbres de la liste et on les remplace par un arbre ayant ces deux arbres comme sous-arbres.

L'arbre « naïf » est construit en utilisant l'un (au hasard) de deux algorithmes faciles. Le premier algorithme est le suivant : calculer la matrice de distance entre les génomes des N espèces étudiées. Rechercher les deux plus proches voisins, construire un arbre contenant ces deux feuilles. Puis, à chaque étape, rechercher les deux plus proches feuilles n'appartenant pas encore au même arbre ; regrouper les deux arbres contenant ces feuilles. Ceci jusqu'à obtenir un arbre contenant toutes les feuilles. Le second algorithme est le suivant. On classe les espèces dans un ordre aléatoire. On construit un arbre contenant les deux premières espèces et on calcule le (un) génome minimisant la longueur de cet arbre. Ensuite, à chaque étape, on considère l'espèce suivante dans la liste. On l'insère sous le nœud de l'arbre en construction dont son génome est le plus proche. On recalcule les génomes optimaux des nœuds internes du nouvel arbre obtenu.

On prend pour chaque arbre de la population initiale un croisement d'un arbre déterministe naïf et d'un arbre aléatoire ; si la population initiale était, au contraire, composée pour moitié d'arbres entièrement aléatoires et pour moitié d'arbres déterministes, les arbres aléatoires seraient immédiatement éliminés.

Cette méthode d'introduction de diversité est inspirée par la volonté de ne pas tomber dans les mêmes optima locaux que les algorithmes naïfs. Elle semble augmenter un peu le nombre de générations nécessaires pour obtenir un bon arbre (les premiers arbres n'étant pas très bons). Elle semble en tout cas ne pas empirer le résultat final, même si l'on n'a pas encore isolé de situation où de nouveaux arbres seraient découverts par cette méthode.

Quant à la taille de la population, elle est déterminée par le critère suivant : on veut que l'algorithme soit capable d'effectuer correctement l'analogue d'une descente de gradient qui ne mettrait en cause que des modifications locales de l'arbre (mutations locales décrites ci-dessus). Sachant que de l'ordre d'une mutation locale se produit pour chaque arbre à chaque génération, et que le nombre de mutations locales différentes est de l'ordre du nombre de nœuds de l'arbre, pour que chaque mutation locale soit testée à chaque génération, il faut que le nombre d'arbres dans la population soit de l'ordre du nombre de nœuds de l'arbre. Empiriquement, l'on a choisi $m = 6N$.

2.5 Critère d'arrêt

Le critère d'arrêt retenu est le suivant : on s'arrête lorsqu'aucune amélioration n'est intervenue au cours d'un nombre de générations égal à N. Par le même type de raisonnement ci-dessus par lequel on choisissait la taille de la population, laisser un nombre de générations de l'ordre de N autorise probablement chaque *couple* de mutations locales à avoir été testé une fois environ.

3 Performances

3.1 Complexité de l'algorithme

On rappelle que k est la longueur des génomes étudiés, N le nombre d'espèces et m la taille de la population.

La complexité de chacune des étapes utilisées est la suivante.

- **Sélection des arbres.** Évaluer la longueur d'un arbre (calcul du remplissage optimal) peut se faire en $O(kN)$. Le coût total est en $O(mkN)$.
- **Croisement de deux arbres.** Vérifier si deux nœuds de tailles n_1 et n_2 sont soit inclus l'un dans l'autre, soit d'intersection vide (critère de compatibilité pour qu'ils puissent figurer dans le même arbre) peut se faire en $O(n_1 + n_2)$ (en utilisant une représentation des nœuds où les feuilles sont arbitrairement ordonnées). Vérifier si un nœud de taille n_1 est compatible avec tous les nœuds d'un arbre à n_2 feuilles peut donc se faire en $O((n_1+n_2)n_2)$ (et même, s'il se trouve que l'arbre est bien équilibré, plutôt en $O((n_1 + n_2)\log n_2)$, mais nous n'utiliserons pas cette estimation heuristique). Notre croisement de deux arbres commence par fusionner leurs listes de nœuds et ensuite élimine ceux qui sont incompatibles. Au pire cet

algorithme fonctionne en $O(N^3)$. Le coût total est en $O(mN^3)$. Noter que la longueur du génome n'intervient absolument pas.

– **Mutation d'un arbre.** Une mutation locale, telle que décrite plus haut, prend un temps $O(d)$ où d est le nombre maximal de nœuds-fils d'un même nœud de l'arbre, soit $O(N)$ en mettant les choses au pire (plutôt $O(1)$ empiriquement car nos arbres sont presque toujours binaires). Il y a en moyenne une mutation locale par arbre d'après notre choix du taux de mutation, et le coût total des mutations locales est donc $O(mN)$. Pour les mutations globales (cf. ci-dessus), la taille du génome intervient : le coût d'une mutation globale est en $O(kN)$. Avec notre choix du taux de mutation, le coût moyen total des mutations globales est donc $O(mkN)$.
– **Préparation de la population initiale.** Le coût des algorithmes « naïfs » utilisés pour la préparation de la population initiale est en $O(kN^2)$, à répéter m fois.

Le coût total pour t générations est donc $O(kN^2 + tmkN + tmN^3)$. Pour les tailles de données examinées, supprimer l'étape la plus longue (le croisement des arbres) ne changeait pas fondamentalement les temps d'exécution.

3.2 Résultats, comparaison avec PAUP*

L'algorithme a été testé en utilisant quelques données réelles disponibles sur le site Cladestore maintenu par M. J. Benton (de l'université de Bristol), à l'adresse `http://palaeo.gly.bris.ac.uk/cladestore/` (ce site regroupe des données déjà publiées dans des revues). Nous n'avons choisi que des données binaires (l'implémentation des autres formats n'ayant pas été totalement menée à bien), ce qui a beaucoup restreint les possibilités. Les références pour les données sont disponibles sur les pages citées.

Trois jeux de données, en particulier, ont été examinés.

– Archosauria (ancêtres des crocodiles, dinosaures et oiseaux) d'après Benton, voir `http://palaeo.gly.bris.ac.uk/cladestore/Archosauria/Archosauria6.html` : 21 taxons, 134 caractères.
– Eutheria (mammifères placentaires) d'après Novacek et Wyss, voir `http://palaeo.gly.bris.ac.uk/cladestore/Mammalia/Eutheria2.html` : 21 taxons, 68 caractères.
– Ungulata (mammifères à sabots) d'après Prothero et al., voir `http://palaeo.gly.bris.ac.uk/cladestore/Mammalia/Ungulates1.html` : 13 taxons, 28 caractères.

Il avait été envisagé de faire une analyse d'un jeu de données très gros (« rbcL 500 ») de 500 taxons de spermatophytes (plantes à graines), afin de tester si le croisement permettait vraiment une exploration efficace de l'espace des arbres. Le temps de calcul prohibitif a empêché de mener sérieusement l'analyse. (L'analyse initiale avait coûté près de 12 mois-processeurs en 1995).

Les performances ont été comparées à celui du programme de référence PAUP* version 4(beta) , décrit dans [S] et qui peut être acheté sur `http://paup.`

`csit.fsu.edu/`.

Sur ces trois problèmes, le programme testé a fourni des arbres aussi parcimonieux que PAUP* (recherche heuristique). Le temps de calcul est par contre différent (quasi-instantané pour PAUP*, quelques secondes pour le programme testé).

Les jeux de données en question sont malheureusement très petits : il est probable que l'arbre trouvé soit réellement optimal.

À titre d'illustration, afin qu'un texte sur les arbres ne s'achève pas sans en mentionner, donnons un arbre optimal correspondant au plus gros jeu de données testé (Archosauria d'après Benton) :

```
-+
 |-Youngina
 |-Prolacerta
 '-+
   |-+
   | |-Trilophosaurus
   | '-Hyperodapedon
   '-+
     |-Proterosuchus
     '-+
       |-Erythrosuchus
       '-+
         |-+
         | |-Chanaresuchus
         | '-Doswellia
         '-+
           |-Euparkeria
           '-+
             |-Parasuchus
             '-+
               |-+
               | |-Ornithosuchus
               | '-+
               |   |-Lagosuchus
               |   '-Plateosaurus
               '-+
                 |-Gracilisuchus
                 '-+
                   |-+
                   | |-Stagonolepis
                   | '-+
                   |   |-Ticinosuchus
                   |   '-Saurosuchus
                   '-+
                     |-Postosuchus
                     '-+
                       |-Protosuchus
                       '-+
                         |-Terrestrisuchus
                         '-Crocodylus
```

À noter que plusieurs arbres optimaux coexistent dans ce cas : le programme étudié aussi bien que PAUP* donnent des variantes de cet arbre, où, par exemple, les dinosaures (représentés par *Plateosaurus*) sont plus proches des crocodiliens (représentés par *Crocodylus*) que les aétosaures (*Stagonolepis*).

4 Perspectives et conclusion

L'intérêt d'utiliser un croisement d'arbres n'est donc pas démontré par ces expériences (pas plus que son inintérêt). Le ralentissement de performance est lié (outre la qualité de la programmation) à l'utilisation d'une population large et d'algorithmes stochastiques n'allant pas droit au but en testant les possibilités (de mutation d'un arbre) de manière systématique. Peut-être gagnerait-on à utiliser l'idée de croisement non pas dans un cadre d'algorithme génétique mais dans un cadre moins aléatoire, en l'incorporant à un programme de type PAUP*.

Le croisement peut avoir d'autres applications. Ainsi lorsqu'on travaille avec des modèles plus complexes que la parcimonie, par exemple, des modèles de maximum de vraisemblance, qui demandent que chaque branche de l'arbre porte un ou plusieurs paramètres (tels qu'un taux de mutation), ces paramètres devant eux-mêmes faire l'objet d'une optimisation difficile. L'utilisation d'un algorithme génétique avec croisement (où par exemple les paramètres des enfants seraient la moyenne de ceux des parents) pourrait permettre de traiter en même temps la recherche de la meilleure topologie et la recherche des meilleures valeurs de paramètres. (Idée discutée avec C. Kéribin.)

Enfin, de nombreuses améliorations ponctuelles du programme seraient à effectuer : meilleur choix de la population initiale (meilleure répartition entre aléa et algorithmes « naïfs »), meilleures mutations (en particulier, la mutation globale, qui coupe un sous-arbre pour l'attacher à l'endroit portant le génome le plus proche, est beaucoup trop déterministe et cela a semblé poser problème en particulier pour les espèces ayant des données manquantes ; on devrait accorder une certaine probabilité d'attache en d'autres points ayant un génome un peu moins proche), etc. On pourrait aussi s'inspirer des mécanismes de diploïdie : lors du croisement, plutôt que d'éliminer purement et simplement les nœuds de l'arbres incompatibles, on pourrait les conserver à titre de « gènes récessifs » ce qui permettrait à des nœuds ayant eu un succès évolutif à un moment d'être réutilisés plus tard.

Quoi qu'il en soit, l'intérêt du croisement d'abres n'est pas établi, mais intégrer le principe du croisement à des programmes déjà connus pour fonctionner est sans doute une voie intéressante.

Références

[BHV] L.J. Billera, S.P. Holmes, K. Vogtmann, *Geometry of the Space of Phylogenetic Trees*, Adv. Appl. Math. **27** (2001), 733–767.

[H] J.A. Hartigan, *Minimum mutation fits to a given tree*, Biometrics **29** (1973), 53–65.

[KW] J. Kim, T. Warnow, *Tutorial on Phylogenetic Tree Estimation*, Proc. Intelligent Systems for Molecular Biology Conference, Heidelberg (1999).

[RW] K. Rice, T. Warnow, *Parsimony is hard to beat!*, Proc. Third Annual Int. Conf. on Computing and Combinatorics (COCOON), Shangai (1997), eds. T. Jiang, D.T. Lee, 124–133.

[S] D. Swofford, *PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Sinauer Associates, Sunderland, Massachusetts (2001).

Table des matières

La démographie du PRA

Résumé

Nous présentons une preuve du fait que le PRA utilisé en (très) grande population a, dans un premier temps, une convergence superexponentielle, contrôlée par le trou spectral du groupe. Nous présentons aussi une borne inférieure pour cette convergence superexponentielle. Ces résultats, inspirés des algorithmes génétiques, ne peuvent cependant pas montrer que le PRA fait mieux que la marche aléatoire simple.

1 Le Product Replacement Algorithm

Le Product Replacement Algorithm (PRA) est un algorithme stochastique dont le but est le suivant. Soit G un groupe fini donné sous la forme suivante : un système générateur et trois fonctions renvoyant respectivement l'élément neutre, l'inverse d'un élément, le produit de deux éléments (par exemple, un groupe de matrices). Le cardinal de G est supposé connu. Le but de l'algorithme stochastique est de renvoyer un élément du groupe dont la loi soit proche de la loi uniforme sur le groupe. De tels éléments aléatoires uniformément répartis sont utiles pour un grand nombre d'algorithmes sur les groupes (voir les nombreuses références données dans [Pak1]).

Le PRA a été introduit dans [CLMNO]. C'est un algorithme qui s'est révélé extrêmement efficace en apparence, sans pour autant que son efficacité puisse être prouvée. Nous renvoyons à [Pak1] pour une histoire et une justification du problème, ainsi qu'un bilan (qui n'est plus à jour) des connaissances. Depuis, Pak a prouvé (cf. [Pak2]) un très intéressant théorème de vitesse de convergence.

Une explication conceptuelle mais conjecturale de ces performances a été avancée par Lubotzky et Pak dans [LP].

Le PRA consiste en l'itération de l'opération suivante. Étant donné un système générateur $(s_1, \ldots, s_m)$ d'un groupe G (système non nécessairement minimal, et qui peut, par exemple, contenir l'élément neutre, plusieurs fois le même élément, etc.), on produit aléatoirement un autre système générateur ainsi : on tire au hasard (uniformément) deux indices distincts $1 \leqslant i, j \leqslant m$; puis, dans le m-uplet générateur, on remplace l'élément s_j soit par $s_j s_i$, soit par $s_j s_i^{-1}$, soit par $s_i s_j$, soit par $s_i^{-1} s_j$, en choisissant au hasard entre ces quatre possibilités. Il

est immédiat de vérifier que le nouveau m-uplet obtenu est encore un m-uplet générateur de G.

On part d'un m-uplet donné de G, on itère cette transformation aléatoire un certain nombre de fois, puis on sélectionne l'élément s_1 (ou un autre, d'ailleurs) du m-uplet obtenu à la dernière étape.

Nous étudions ici le comportement du PRA en très grande population m. Nous montrons que dans ce cadre, la convergence de l'algorithme, au début, est superexponentielle (Proposition 1), en $\lambda_1^{e^{t/m}}$, mais avec des constantes faibles qui font qu'en temps de calcul on ne peut pas prouver par cette méthode que le PRA fait mieux que la marche aléatoire simple.

Nous donnons aussi des invariants curieux (mais sans application) du PRA sur les groupes cycliques. Nous montrons enfin une sorte de borne inférieure (proposition 2).

2 Comportement en très grande population

2.1 Notations

À chaque étape, deux individus choisis au hasard sont croisés, l'un des parents est remplacé par le produit de ce croisement, et les autres individus sont maintenus dans la population.

On étudie ici la loi du nombre de parents d'un individu au cours du temps.

On regardera le processus en remontant le temps, i.e. on part de la génération 0 et on suppose le processus défini aux temps négatifs. Si on regarde le processus à l'envers, on voit qu'à chaque génération, avec probabilité $1/m$, un individu a deux parents, et qu'il n'en a qu'un avec probabilité $1 - 1/m$.

On fixe une fois pour toutes un individu de la génération 0, que nous noterons X, dont on étudie l'ascendance.

Soit N_t, pour $t \geqslant 0$, le nombre d'ancêtres (aléatoire) au temps $-t$ de X ; les ancêtres sont comptés avec multiplicités, c'est-à-dire que si un individu de la génération t est ancêtre par deux voies différentes, il compte double. Cette quantité ne dépend que des choix d'ancêtres parmi m individus et pas du groupe dans lequel on travaille.

Soit T_k, pour $k \geqslant 1$, le nombre d'étapes où l'individu de la génération 0 a k ancêtres (on ne compte pas la génération 0). Ainsi, si $\sum_{i=1}^{k} T_i = t$, on a $N_t = k$ et $N_{t+1} = k + 1$.

Si p et q sont deux mesures de probabilité sur le même ensemble, on note $|p - q|_{TV} = \frac{1}{2} \sum_x |p(x) - q(x)| \leqslant 1$.

Notons que l'analyse faite ici serait triviale si on modifiait la définition du PRA de sorte qu'à chaque étape, tous les individus soient croisés entre eux, comme dans un algorithme génétique classique. Ce qui nous intéresse ici est donc uniquement la démographie résultant du fait de croiser un seul couple à chaque génération.

2.2 Loi de N_t

On va d'abord s'intéresser à l'espérance de N_t, qui est indépendante du groupe sur lequel on travaille.

Faisons fonctionner le PRA sur $\mathbb{N}$ en partant du m-uplet $(1, 1, \ldots, 1)$, sans faire apparaître les signes $-$, et notons $(a_1^t, \ldots, a_m^t)$ le m-uplet obtenu au temps t. On voit que la loi de N_t est celle de a_i^t pour un i quelconque. Calculons son espérance.

Supposons qu'au temps t, on ait $\sum_i a_i^t = s$. Calculons l'espérance de $\sum_i a_i^{t+1}$ sachant s :

$$\begin{aligned}
\mathbb{E}\sum_i a_i^{t+1} &= \frac{1}{m}\left(\sum_i \left(a_i + \frac{1}{m-1}\sum_{j\neq i} a_j + \sum_{j\neq i} a_j\right)\right) \\
&= \left(1+\frac{1}{m}\right)\sum_i a_i \\
&= \left(1+\frac{1}{m}\right) s
\end{aligned}$$

On en déduit que l'espérance de la somme des a_i^t vaut $m\,\left(1+\frac{1}{m}\right)^t$. Comme cette espérance est la somme des espérances des a_i^t et que par symétrie toutes ces espérances sont égales, on en déduit que pour tout i, on a $\mathbb{E}N_t = \mathbb{E}a_i^t = \left(1+\frac{1}{m}\right)^t$, qui vaut environ $e^{t/m}$.

On va maintenant estimer les moments d'ordre deux des a_i. Pour cela, deux quantités vont entrer en jeu : l'espérance de $(a_i^t)^2$, et celle de $a_i^t a_j^t$ pour $j \neq i$. Connaissant tous les (a_j^t) au temps t pour $j = 1 \ldots m$, on a :

$$\begin{aligned}
\mathbb{E}\left(\left(a_i^{t+1}\right)^2 \mid \left(a_j^t\right)_j\right) &= \left(1-\frac{1}{m}\right)\left(a_i^t\right)^2 + \frac{1}{m(m-1)}\sum_{j\neq i}\left(a_i^t + a_j^t\right)^2 \\
&= \left(a_i^t\right)^2 + \frac{1}{m(m-1)}\sum_{j\neq i}\left(a_j^t\right)^2 + \frac{2}{m(m-1)}\sum_{j\neq i} a_i^t a_j^t
\end{aligned}$$

Pour $i \neq j$, par un calcul similaire on obtient :

$$\mathbb{E}\left(a_i^{t+1}a_j^{t+1} \mid \left(a_k^t\right)_k\right) = a_i^t a_j^t + \frac{\left(a_i^t\right)^2 + \left(a_j^t\right)^2}{m(m-1)} + \frac{1}{m(m-1)}\sum_{k\neq i,j}(a_i^t + a_j^t)a_k^t$$

Par conséquent, en notant $A_t = \mathbb{E}\left(a_i^t\right)^2 = \mathbb{E}N_t^2$ et $B_t = \mathbb{E}a_i^t a_j^t$, indépendants de i et j par symétrie, on obtient

$$\begin{aligned}
A_{t+1} &= \left(1+\frac{1}{m}\right)A_t + \frac{2}{m}B_t \\
B_{t+1} &= \frac{2}{m(m-1)}A_t + \left(1+\frac{2(m-2)}{m(m-1)}\right)B_t
\end{aligned}$$

Par diagonalisation de la matrice ainsi obtenue, on peut montrer que sa plus grande valeur propre est inférieure à $1 + 2/m + 8/m^2$ et que la composante correspondante pour A_t est inférieure à 4, d'où :

$$\mathbb{E}N_t^2 \leqslant 4\left(1 + \frac{2}{m} + \frac{8}{m^2}\right)^t \leqslant 4e^{2t/m+6t/m^2}$$

2.3 Probabilité des collisions

On dit qu'il y a une collision à la génération t s'il existe un individu de cette génération qui est ancêtre de X par au moins deux voies différentes. (Ceci est fonction de la généalogie des éléments du k-uplet et est indépendant de la valeur de ces éléments.)

Connaissant N_t, la probabilité qu'il y ait une nouvelle collision dans le passage du temps t au temps $t+1$ est inférieure à $\frac{N_t(N_t-1)}{m(m-1)}$. La probabilité qu'il se soit produit au moins une collision entre le temps 0 et le temps t est donc inférieure à

$$\begin{aligned}\sum_{i=1}^{t-1} \mathbb{E}\frac{N_t(N_t-1)}{m(m-1)} &\leqslant \sum_{i=1}^{t-1} \frac{4}{m(m-1)}\left(1 + \frac{2}{m} + \frac{8}{m^2}\right)^i \\ &\leqslant \frac{4e^{2t/m+6t/m^2}}{m}\end{aligned}$$

2.4 Loi des T_k

L'évaluation que nous avons de $\mathbb{E}N_t$ et $\mathbb{E}N_t^2$ ne suffit pas à majorer la probabilité que N_t soit petit (car l'écart-type de N_t est du même ordre que sa moyenne). Pour cela, on va raisonner sur le logarithme de N_t en utilisant les T_k définis ci-dessus : T_k est le temps passé dans l'état $N_t = k$.

On va estimer les T_k en l'absence de collision. Dans tout ce paragraphe on suppose qu'à l'étape où on se trouve, il n'y a pas encore eu de collision. En particulier, on prend $m \geqslant k$.

Le processus N_t est markovien. T_k est le temps passé à l'état k par ce processus. Les probabilités de transition sont les suivantes. Il y a k chances sur m que l'un des individus ancêtres de X soit celui qui a deux parents à la génération $-t-1$. Par conséquent, dans l'état k, avec probabilité k/m on passe à l'état $k+1$, et avec probabilité $1 - k/m$ on reste à l'état k.

Le temps T_k passé à l'état k suit donc une loi géométrique (qui démarre à 1 et non à 0 d'après nos conventions) : la probabilité que $T_k = i$, pour $i \geqslant 1$, vaut $p(1-p)^{i-1}$ avec $p = k/m$. En particulier, l'espérance de T_k vaut $1/p$. Sa variance vaut $(1-p)^2/p^2 + (1-p)/p$.

On s'intéresse à la somme $S_k = \sum_{i=1}^{k} T_i$; on a $N_{S_k} = k$ et $N_{S_{k+1}} = k+1$. Les T_i sont indépendants.

L'espérance de S_k est $\mathbb{E}S_k = \sum_{i=1}^{k} m/i$ et en particulier

$$m \log k \leqslant \mathbb{E}S_k \leqslant m + m\log(k-1)$$

La variance de S_k est

$$\sigma S_k = \sum_{i=1}^{k} \left(\frac{1-i/m}{i/m} \right)^2 + \frac{1-i/m}{i/m} \leqslant m^2\pi^2/6 + m + m\log k$$

T_k suit une loi géométrique sur $\mathbb{N}$, démarrant à 1 et de raison $1-k/m$. Pour la simplicité des expressions, on travaillera sur les variables réduites $t_k = T_k - 1$ qui ont l'avantage de suivre des lois géométriques commençant à 0. Cela donnera simplement une translation de k sur la loi de S_k.

Calculons la transformée de Fourier de la loi de t_k. Pour $x \in [0; 2\pi]$ on a :

$$\sum_{n\in\mathbb{N}} \frac{k}{m} (1-k/m)^n \, e^{-ixn} = \frac{k/m}{1-(1-k/m)e^{-ix}}$$

Alors, la loi de $S_k - k$ est la convolée des lois des t_j, $j = 1 \ldots k$. On l'obtient par transformée de Fourier inverse du produit des transformées de Fourier de la loi des t_j, soit

$$\begin{aligned}
\mathbb{P}(S_k = t+k) &= \frac{1}{2\pi} \int_{x=0}^{2\pi} \frac{e^{itx}\,\mathrm{d}x}{\prod_{j=1}^{k} 1-(1-j/m)e^{-ix}} \prod_{j=1}^{k} \frac{j}{m} \\
&= \frac{k!}{m^k} \frac{1}{2i\pi} \oint_{z\in S^1} \frac{z^t\,\mathrm{d}z/z}{\prod_{j=1}^{k} 1-(1-j/m)/z}
\end{aligned}$$

et d'après la formule de Cauchy,

$$\begin{aligned}
\mathbb{P}(S_k = t+k) &= \frac{k!}{m^k} \sum_{j=1}^{k} \frac{(1-j/m)^{t+k-1}}{\prod_{j'\neq j} j'/m - j/m} \\
&= \frac{k!}{m^k} \sum_{j=1}^{k} \frac{(1-j/m)^{t+k-1}\,(-1)^{j+1}}{(j-1)!(k-j)!/m^{k-1}} \\
&= \frac{k}{m} \sum_{j=0}^{k-1} \mathrm{C}_{k-1}^{j} \left(1 - \frac{j+1}{m}\right)^{t+k-1} (-1)^j
\end{aligned}$$

On s'intéresse à la déviation de S_k au-dessus de sa moyenne qui est environ $m\log k$; prenons donc $t+k-1 \geqslant m\log k$. Montrons qu'alors, en valeur absolue, le premier terme de la somme est dominant.

On a $\mathrm{C}_{k-1}^{j+1}/\mathrm{C}_{k-1}^{j} \leqslant k$. Par ailleurs, $(1-(j+2)/m)/(1-(j+1)/m) \leqslant 1-1/m$. Par conséquent, si $t+k-1 \geqslant m\log k$, ce rapport à la puissance $t+k-1$ est inférieur à $(1-1/m)^{m\log k} \leqslant 1/k$. Autrement dit, les termes de la somme sont décroissants en valeur absolue. Comme en outre la somme est alternée, on peut la majorer par son premier terme :

$$\mathbb{P}(S_k = t+k) \leqslant \frac{k}{m}(1-1/m)^{t+k-1}$$

Posant $t + k = m \log k + m\delta$, il vient :

$$\mathbb{P}(S_k = m \log k + m\delta) \leqslant \frac{2}{m}\, e^{-\delta}$$

et par sommation

$$\mathbb{P}(S_k \geqslant m \log k + m\delta) \leqslant 4e^{-\delta}$$

soit encore, sachant que $N_{S_k} = k$ et posant $k = e^{t/m-\delta}$

$$\mathbb{P}(N_t \leqslant e^{t/m-\delta}) \leqslant 4e^{-\delta}$$

qui est le résultat qui nous intéresse.

À noter que quand k est grand, l'erreur relative est de l'ordre de $1/\log k$.

Pour la déviation vers les valeurs inférieures (dont nous n'avons pas besoin), on pourrait remplacer T_k par une variable aléatoire t_k égale à T_k si $T_k \leqslant \mathbb{E}T_k = m/k$, et m/k si $T_k > m/k$.

2.5 Dispersion des N_t

Étudions maintenant la dispersion de N_t autour de sa moyenne.

On a évidemment $N_t > k \Leftrightarrow S_k < t$. Posant successivement $t = m \log k + m\delta$ et $t = m \log k - m\delta$ dans l'évaluation ci-dessus de la loi de S_k, et négligeant les problèmes de parties entières, on obtient

$$e^{t/m-\delta} \leqslant N_t \leqslant e^{t/m+\delta}$$

avec probabilité supérieure à $1 - 8/(\delta - 1)^2$, sous les conditions $\delta > 1$ et $t < m^2 - m\delta$ (condition issue de $\log k < m$).

À noter que N_t est ainsi connu à un facteur constant près.

On va donc s'intéresser à $\frac{1}{t} \log N_{mt}$. Comme $\mathbb{P}\left(\frac{1}{t} \log N_{mt} \geqslant 1 + \delta\right) \leqslant \frac{4}{(t\,\delta - 1)^2}$, on a

$$\begin{aligned} \mathbb{E}\left(\frac{1}{t} \log N_{mt}\right) &\leqslant 1 + \delta + \int_\delta^\infty \frac{4\mathrm{d}u}{(tu-1)^2} \\ &\leqslant 1 + \delta + \frac{4}{t(t\delta - 1)} \end{aligned}$$

Inversement, $\mathbb{P}\left(\frac{1}{t} \log N_{mt} \leqslant 1 - \delta\right) \leqslant \frac{4}{(t\,\delta - 1)^2}$ et donc

$$\begin{aligned} \mathbb{E}\left(\frac{1}{t} \log N_{mt}\right) &\geqslant 1 - \delta - \int_\delta^1 \frac{4\mathrm{d}u}{(tu-1)^2} \\ &\geqslant 1 - \delta - \frac{4}{t(t\delta - 1)} \end{aligned}$$

Prenant $\delta = 3/t$ dans ces expressions, on obtient

$$1 - \frac{5}{t} \leqslant \mathbb{E}\left(\frac{1}{t} \log N_{mt}\right) \leqslant 1 + \frac{5}{t}$$

et de plus, tout $\varepsilon > 0$,

$$\lim_{t\to\infty}\lim_{m\to\infty}\mathbb{P}\left(\frac{1}{t}\log N_{mt}\in[1-\varepsilon;1+\varepsilon]\right)=1$$

3 PRA et convolution

Supposons que la population au temps 0 est obtenue par m tirages successifs d'éléments du groupe selon une loi p symétrique. En population assez grande, donc sans collisions, un élément à l'étape t est le produit (avec des exposants ± 1 pris au hasard) de N_t éléments de la population au temps 0.

On sait que $|p^{*t}-U|_{TV}\leqslant\sqrt{|G|}\lambda_1^t$ où λ_1 est la plus grande valeur propre de l'opérateur de la marche aléatoire sur G (égale à 1 moins la première valeur propre du laplacien).

Soit q_t la loi d'un individu tiré au temps t par le PRA. Conditionnellement au fait qu'il n'y a pas eu de collision et que N_t est supérieur à N, on a

$$|q_t-U|_{TV}\leqslant\sqrt{|G|}\lambda_1^N$$

On utilise ensuite d'une part l'évaluation de la probabilité de collision par $\mathbb{E}N_t^2$, et d'autre part que $\mathbb{P}\left(N_t\leqslant e^{t/m-\delta}\right)\leqslant 4e^{-\delta}$.

On obtient ainsi

PROPOSITION 1 – *Pour tout* $\varepsilon>0$

$$|q_t-U|_{TV}\leqslant\varepsilon+\frac{4e^{2t/m+6t/m^2}}{m}+\sqrt{|G|}(\lambda_1)^{e^{t/m-\log(4/\varepsilon)}}$$

Pour rendre cette quantité plus petite qu'un certain seuil δ, connaissant le λ_1 de la marche aléatoire sur G, on pose $\varepsilon=\delta/3$, puis on fixe t/m de sorte que $\sqrt{|G|}\lambda_1^{e^{t/m-\log(4/\varepsilon)}}\leqslant\delta/3$, ce qui peut se faire en prenant

$$t/m=C(\delta)\log\left(\log|G|\,/\log(1/\lambda_1)\right)$$

Ensuite, on fixe m tel que $4e^{2t/m+6t/m^2}/m$ soit inférieur à $\delta/3$, soit

$$m=C(\delta)\left(\log|G|\,/\log(1/\lambda_1)\right)^2$$

Puis on calcule t par $t=m\,t/m$, ce qui donne

$$t=C(\delta)\left(\log|G|\,/\log(1/\lambda_1)\right)^2\log\left(\log|G|\,/\log(1/\lambda_1)\right)$$

(La marche aléatoire simple nécessite un temps en $C(\delta)\left(\log|G|\,/\log(1/\lambda_1)\right)$; comme notre résultat utilise qu'une partie du PRA simule une marche aléatoire, on ne peut pas montrer mieux par cette méthode.)

L'utilisation de ce résultat nécessite de très grandes populations : si $e^{t/m}$ est grand, alors $e^{2t/m+6t/m^2}$ l'est plus encore, et m doit être grand devant cette dernière quantité. Autrement dit, à m fixé, cette estimation n'est intéressante que si $t \ll m \log m$, et en particulier l'exposant de λ_1 sera petit devant m.

Ce résultat, bien que sans intérêt pratique (il nécessite des très grandes populations), peut au moins fournir une explication conceptuelle de la rapidité de l'algorithme : en effet, on obtient une convergence superexponentielle (avec une constante de temps $1/m$). Il serait extrêmement intéressant que des résultats similaires soient valables pour des populations plus petites.

Pour évaluer ce résultat dans quelques situations, on peut par exemple utiliser une borne bien connue sur le λ_1, à savoir $1 - \lambda_1 \geqslant \frac{1}{m\,\eta\,\mathrm{diam}\,G}$, où m est le nombre d'éléments d'un système générateur, $\mathrm{diam}\,G$ le diamètre du graphe de Cayley de G pour ce système générateur, et η le nombre maximal de fois qu'un générateur donné doit être utilisé pour écrire minimalement un élément quelconque du groupe (au plus $\mathrm{diam}\,G$), on obtient

$$t = C(\delta)\,(m\,\eta\,\mathrm{diam}\,G \log|G|)^2 \log(m\,\eta\,\mathrm{diam}\,G \log|G|)$$

ce qui est à comparer au $(\log|G|)^9\,(\log\log|G|)^5$ obtenu par I. Pak.

Par exemple, si on considère $\mathbb{Z}/n\mathbb{Z}$, avec comme système de générateurs les puissances de 2 jusqu'à $2^{\log_2 n}$, on a $m = \log_2 n$, $\mathrm{diam}\,G = \log_2 n$ et $\eta = 1$, et l'évaluation est en $(\log n)^6 \log\log n$. Ce résultat est à comparer à celui obtenu par Lubotzky et Pak[1] pour des groupes commutatifs, qui donne $(\log n)^7$ dans ce même cas, pour obtenir un m-uplet uniforme (ce qui est légèrement différent du fait d'obtenir un élément uniforme, mais pour m de l'ordre de $\log|G|$ comme ici, les éléments d'un m-uplet uniforme sont uniformes).

La dépendance en δ que nous obtenons est assez bonne (polynomiale en $\log 1/\delta$). On se soucie plus de la dépendance du résultat en le groupe sur lequel on travaille.

Répétons que, bien évidemment, comme notre technique de démonstration est fondée sur le fait qu'une sous-partie du PRA fonctionne comme la marche aléatoire, on ne peut pas obtenir par cette voie directe une meilleure estimation que pour la marche aléatoire. Les économies de calcul du PRA par rapport à la marche aléatoire ne peuvent provenir que des collisions, qui sont justement ce que nous avons négligé dans cette étude (et qui force la grande taille de population à utiliser). L'étape suivante de l'étude est donc naturellement l'évaluation du biais introduit par les collisions.

4 Le PRA sur $\mathbb{Z}$ et $\mathbb{Z}/n\mathbb{Z}$

Supposons qu'on fait fonctionner le PRA sur $\mathbb{Z}$, et que la génération 0 est obtenue par m tirages indépendants selon une loi p sur $\mathbb{Z}$. On va s'intéresser

[1] On se base sur l'estimation en $m^6 \log|G|$ démontrée dans leur article [LP] pour les groupes commutatifs, bien qu'ils annoncent $m^5 \log|G|$ en utilisant un lemme non démontré.

à la loi d'un individu obtenu au temps t, dont on va calculer l'espérance et la variance. Soit μ l'espérance de la loi p et σ^2 sa variance.

Par un calcul similaire à celui mené plus haut, si on note a_i^t le i-ième élément du m-uplet obtenu au temps t, on a :

$$\begin{aligned}\mathbb{E}\left(a_i^{t+1}|\left(a_j^t\right)_j\right) &= \left(1-\frac{1}{m}\right)a_i^t + \frac{1}{2m(m-1)}\sum_{j\neq i}(a_i^t+a_j^t) \\ &\quad + \frac{1}{2m(m-1)}\sum_{j\neq i}(a_i^t-a_j^t) \\ &= a_i^t\end{aligned}$$

et par conséquent, l'espérance d'un élément tiré au temps t est l'espérance de la loi p : l'espérance est conservée par le processus.

À noter que ceci donne aussi un invariant non trivial sur $\mathbb{Z}/n\mathbb{Z}$ si 2, m et $m-1$ sont premiers avec n, auquel cas nos divisions ont un sens.

Pour le moment d'ordre deux :

$$\begin{aligned}\mathbb{E}\left(\left(a_i^{t+1}\right)^2|\left(a_j^t\right)_j\right) &= \left(1-\frac{1}{m}\right)\left(a_i^t\right)^2 + \frac{1}{2m(m-1)}\sum_{j\neq i}(a_i^t+a_j^t)^2 \\ &\quad + \frac{1}{2m(m-1)}\sum_{j\neq i}(a_i^t-a_j^t)^2\end{aligned}$$

d'où l'on tire que $\mathbb{E}\left(a_i^t\right)^2 = \left(1+\frac{1}{m}\right)^t(\sigma^2+\mu^2)$.

Par conséquent, si σ_t est la variance d'un élément obtenu au temps t, on a $\sigma_t = \left(1+\frac{1}{m}\right)^t(\sigma^2+\mu^2)-\mu^2$.

Un calcul similaire montre que le moment $\mathbb{E}a_i^t a_j^t$ pour $i\neq j$ est constant, égal à μ^2.

On peut aussi s'intéresser à la variance statistique, notons-la v_t^2, du m-uplet $(a_1^t,\ldots,a_m^t)$ obtenu au temps t. On peut évaluer :

$$\begin{aligned}\mathbb{E}v_t^2 &= \mathbb{E}\left(\frac{1}{m}\sum\left(a_i^t\right)^2\right) - \mathbb{E}\left(\frac{1}{m}\sum a_i^t\right)^2 \\ &= \left(1-\frac{1}{m}\right)\left(\left(1+\frac{1}{m}\right)^t(\sigma^2+\mu^2)-\mu^2\right)\end{aligned}$$

Un calcul d'assez de moments d'ordre supérieur permettrait peut-être de trouver, dans $\mathbb{Z}/n\mathbb{Z}$, le nombre de fois modulo n que le PRA donne chaque élément $x\in\mathbb{Z}/n\mathbb{Z}$. Bien sûr, le nombre de fois modulo n est assez éloigné de la proportion de fois...

5 Une borne inférieure

Montrons désormais une sorte de borne inférieure. On suppose que

PROPOSITION 2 – *Pour tout m, pour tout $t \geqslant \log_2 m + \log_2 \log m + 2$, pour tout $k \geqslant 2$ il existe un groupe fini G à k générateurs tel que*

$$\left| q_t - p_0^{*2^t} \right|_{TV} \geqslant 1/2$$

où q_t est la loi d'un élément donné par le PRA au temps t, et p_0 la mesure uniforme sur l'ensemble $S = \{e, a_1^{\pm 1}, \ldots, a_k^{\pm 1}\}$ où les a_i sont les générateurs de G, et où le m-uplet initial du PRA est constitué d'éléments de S.

DÉMONSTRATION – Posons $S = \{e, a_1^{\pm 1}, \ldots, a_k^{\pm 1}\}$ et $s = |S| = 2k + 1$. La démonstration s'appuie sur l'existence, pour tout r, d'un groupe fini sur les générateurs $a_1, \ldots a_k$ où la plus petite relation non triviale entre générateurs est de longueur supérieure à r. Cela signifie que les mots de longueur $r/2$ sur les générateurs et leurs inverses qui ne contiennent pas $\ldots a_i a_i^{-1} \ldots$ sont des éléments distincts du groupe. Par exemple, $\mathrm{PSL}_2(\mathbb{Z})$ contient un sous-groupe libre d'indice 6, et donc un sous-groupe de $\mathrm{PSL}_2(\mathbb{Z}/n\mathbb{Z})$ pour n assez grand répond à la question.

Dans un tel groupe, on peut construire $(s-1)(s-2)^{r/2-1}$ mots distincts de longueur inférieure à $r/2$: pour la première lettre, on doit choisir un élément de S distinct du neutre, ensuite pour chaque lettre on doit faire attention à ne pas prendre l'inverse de la lettre précédente. Ces mots représentent tous des éléments distincts du groupe.

La mesure $p_0^{*2^t}$ charge tous les mots en les générateurs de longueur inférieure à 2^t. Par conséquent, si $2^t \leqslant r/2$, chacun des $(s-1)(s-2)^{2^t-1}$ mots construits précédemment est chargé par $p_0^{*2^t}$, d'un poids $1/s^{2^t}$.

Soit π_t le m-uplet issu du PRA au temps t.

Si m est petit, le PRA n'a accès qu'à une petite partie de ces mots. En effet, si on écrit $t = t_1 + t_2$, à l'étape t_1, π_{t_1} est constitué de m mots de longueur inférieure à 2^{t_1}. Ensuite, un mot de $\pi_{t_1+t_2}$ sera formé en prenant le produit d'au plus 2^{t_2} mots de π_{t_1} ; il sera donc formé à partir de seulement m « motifs de base » en les générateurs de départ.

Précisément, il y a $m^{2^{t_2}}$ mots de longueur 2^{t_2} en les m motifs de π_{t_1}. Maintenant, ces motifs sont des mots de longueur 2^{t_1} sur les générateurs de base, ce qui laisse moins de $s^{m\,2^{t_1}}$ possibilités pour le m-uplet π_{t_1}. Au total, cela laisse $m^{2^{t_2}}\, s^{m\,2^{t_1}}$ possibilités pour un mot de $\pi_{t_1+t_2}$. Ceci est à comparer aux $(s-1)(s-2)^{2^{t_1+t_2}-1}$ mots de longueur $2^{t_1+t_2}$.

Chacun de ces $(s-1)(s-2)^{2^{t_1+t_2}-1}$ est chargé par $p_0^{*2^t}$ d'un poids $1/s^{2^t}$. La distance $\left| q_t - p_0^{*2^t} \right|$ est donc supérieure à $1/s^{2^t}$ fois le nombre de ces mots qui ne peuvent pas être atteints par π_t :

$$\left| q_t - p_0^{*2^t} \right| \geqslant \frac{1}{s^{2^{t_1+t_2}}} \left((s-1)(s-2)^{2^{t_1+t_2}-1} - m^{2^{t_2}}\, s^{m\,2^{t_1}} \right)$$

(Afin de ne pas alourdir les notations, on n'a pas écrit le facteur inférieur à s qui s'introduit entre « le nombre de mots de longueur 2^t » et « le nombre de mots de longueur au plus 2^t ». On laisse les modifications au lecteur.)

On vérifie que $t_2 \geqslant \log_2 m+1$, $t_1 \geqslant \log_2 \log m+1$ conviennent largement pour rendre cette expression supérieure à $1/2$. Ce qui donne $t \geqslant \log_2 m+\log_2 \log m+2$ comme annoncé ; en outre, nous devons prendre un groupe où la plus petite relation est de longueur $r \geqslant 2^t$, ce qui donne $r \geqslant 4m \log m$ pour ce choix de t. Notons aussi qu'on a le libre choix du nombre de générateurs du groupe pourvu qu'il y en ait au moins deux (le groupe libre à un générateur étant un peu trop exigu). □

Références

[CLMNO] F. Celler, C. R. Leedham-Green, S. Murray, A. Niemeyer, E. A. O'Brien, *Generating random elements of a finite group*, Comm. Alg. **23** (1995), 4931–4948.

[LP] A. Lubotzky, I. Pak, *The product replacement algorithm and Kazhdan's property (T)*, J. Amer. Math. Soc. **52** (2000), No. 12, 5525–5561.

[Pak1] I. Pak, *What do we know about the product replacement algorithm?*, in *Groups and Computation III*, eds. W. Kantor, A. Seress, de Gruyter, Berlin (2001), 301–347.

[Pak2] I. Pak, *The product replacement algorithm is polynomial*, Proc. 41st Ann. Symp. Found. Comp. Sc. (Redondo Beach, 2000), IEEE Comput. Soc. Press, Los Alamitos (2000), 476–485.

Table des matières

Remerciements

VOICI, par ordre alphabétique, quelques personnes qui ont joué un rôle dans l'élaboration de cette thèse, ou qui simplement m'ont accompagné et soutenu amicalement dans ce travail.

Thomas Delzant a manifesté un intérêt constant pour mes travaux sur les groupes aléatoires. C'est lui qui m'a orienté vers l'étude de la condition de petite simplification dans le modèle à graphes des groupes aléatoires. Sa compréhension de la théorie de la petite simplification relative a été très utile, ainsi que ses connaissances sur certains points techniques. Il a en outre eu le courage de se proposer spontanément au poste peu envieux de rapporteur de cette thèse...

Damien Gaboriau s'est très vivement intéressé aux problématiques qui m'occupent. Il a été à l'origine de mon long séjour à l'École normale supérieure de Lyon au printemps 2003, l'une des périodes les plus productives (et les plus agréables) de ma thèse. Ses encouragements amicaux ont été particulièrement bienvenus.

Étienne Ghys suit mes travaux très en profondeur et m'encourage depuis longtemps. Nous avons eu de nombreuses conversations extrêmement enrichissantes, que ce soit pour acquérir un peu du recul qu'il possède sur la signification des thèmes abordés, pour trancher un point technique, pour obtenir des conseils et encouragements lors de moments délicats, ou simplement pour partager sa grande culture exprimée de manière si truculente. Il a été l'initiateur, avec D. Gaboriau, de mon séjour à l'École normale supérieure de Lyon au printemps 2003 ; il a bien voulu consacrer du temps à rapporter cette thèse. Depuis encore plus longtemps, son excellent ouvrage (avec P. de la Harpe) sur les groupes hyperboliques a été mon livre de chevet ou presque.

Je voudrais particulièrement remercier Misha Gromov qui a tout simplement créé la plupart des objets mathématiques étudiés ici : groupes hyperboliques, groupes aléatoires... Ses écrits contiennent toutes les intuitions à la base des développements ultérieurs, au point que l'on doute toujours si le théorème que l'on démontre n'était pas déjà bien connu de lui depuis des années (ce qu'il nie toujours). Il a accepté de diriger ma thèse et de me consacrer un peu de son temps : quand il approuve une conjecture on est sûr de ne pas se fourvoyer...

C'est un exposé de Claire Kenyon qui a inspiré mon étude des opérateurs de croisement dans les algorithmes génétiques. Je voudrais la remercier pour cette impulsion ainsi que pour des relectures attentives des parties du manuscrit qui relevaient de sa compétence.

Richard Kenyon a déterminé mon intérêt pour les groupes aléatoires : c'est lui qui, à la suites d'expériences numériques, a semble-t-il le premier détecté l'erreur dans la démonstration de Gromov du premier théorème de transition de phase, qui a piqué ma curiosité. De proche en proche il a ainsi suscité ce qui constitue désormais la part principale de ce recueil.

A.Yu. Ol'shanskiĭ a détecté une erreur importante dans le traitement de la torsion dans une première version du théorème de transition de phase des quotients aléatoires. C'est cette erreur qui m'a conduit à la condition de « torsion inoffensive » nécessaire et suffisante pour la validité du théorème. Il apporte

depuis constamment de nombreuses corrections de détail au manuscrit.

Pierre Pansu n'a pas compté son temps, me consacrant plusieurs heures chaque semaine depuis plus de quatre ans désormais. Il a supporté pendant tout ce temps, et su canaliser, mon discours tumultueux. Mes idées en étaient toujours plus claires à la sortie qu'à l'arrivée. Je lui suis également très reconnaissant de la liberté qu'il m'a laissée dans le choix de mes thèmes de recherche.

Panos Papasoglu a organisé à Orsay en 2002 un groupe de travail sur la géométrie des groupes, cadre de discussions agréables et très informatives sur le sujet. Sa compréhension du théorème local-global hyperbolique a été fort utile.

Frédéric Paulin m'a été d'une aide irremplaçable par sa disponibilité constante pour mes questions de tout ordre. Que celles-ci fussent techniques, portant sur un résultat classique, et la démonstration était claire ou bien l'orientation bibliographique judicieuse. Qu'elles relevassent de l'attitude à adopter dans des moments délicats, les conseils étaient toujours nets et bien inspirés. On mesure mieux sa tâche lorsqu'on sait qu'il fait de même avec un grand nombre des élèves en mathématiques à l'École normale...

L'intérêt d'Alain Valette pour mon travail remonte à longtemps ; nous avons été en contact régulier sur les groupes aléatoires et il m'a prodigué de nombreux conseils. Il a aussi eu l'excellente idée d'organiser des rencontres sur le sujet à l'université de Neuchâtel en décembre 2002, réunissant tous les protagonistes de l'affaire, rencontres très fructueuses scientifiquement et particulièrement agréables humainement.

Andrzej Żuk travaille lui-même, entre autres, sur les groupes aléatoires et m'a invité à un premier séjour à l'École normale supérieure de Lyon au printemps 2002. Séjour au cours duquel m'est venue une des idées principales de la démonstration du théorème sur les quotients aléatoires. Il a toujours manifesté beaucoup d'enthousiasme pour mon travail.

Je voudrais aussi saluer toute l'équipe du laboratoire de mathématiques de l'ENS-Lyon, pour la chaleur de leur accueil lors de mes deux séjours et nombreuses visites, ainsi que pour les nombreuses conversations amicales sur tous sujets (y compris mathématiques !), qui ont fait de ces séjours des moments parmi les plus agréables humainement et féconds scientifiquement de ma thèse.

Bien sûr, l'équipe de topologie à Orsay n'est pas en reste ! en particulier les discussions philosophiques du déjeuner... mais aussi la largeur de vue mathématique des membres du laboratoire.

Enfin, une mention particulière à Martine Justin et Florence Koch, secrétaires d'équipe à Orsay et à l'ENS-Lyon, pour leur gentillesse, leur dynamisme et leur efficacité.

Mais je voudrais surtout remercier tous mes amis pour leur courage à supporter jour après jour mes discours théoriques déraisonnables, mes manières envahissantes et les oscillations quotidiennes de mon humeur suivant l'avancement de mes mathématiques.

Table analytique