Groups Geom. Dyn. (Online first)
DOI 10.4171/GGD/820

Published by EMS Press

The rates of growth in an acylindrically hyperbolic group

Koji Fujiwara

Abstract. Let G be an acylindrically hyperbolic group on a δ-hyperbolic space X. Assume there exists M such that for any finite generating set S of G, the set S^M contains a hyperbolic element on X. Suppose that G is equationally Noetherian. Then we show the set of the growth rates of G is well ordered. The conclusion was known for hyperbolic groups, and this is a generalization. Our result applies to all lattices in simple Lie groups of rank 1, and more generally, relatively hyperbolic groups under some assumption. It also applies to the fundamental group, of exponential growth, of a closed orientable 3-manifold except for the case that the manifold has Sol-geometry.

Dedicated to the memory of Eliyahu Rips with admiration

Contents

1. Introduction . . . 1
2. Lower bound of a growth rate . . . 6
3. Well-orderedness . . . 19
4. Continuity of the growth rate . . . 20
5. Examples . . . 35
6. The set of growth of subgroups . . . 40
7. Finiteness . . . 43
References . . . 57

1. Introduction

1.1. Definitions and results

Let G be a finitely generated group with a finite generating set S. We always assume that $S = S^{-1}$ unless we say otherwise. Let $B_n(G,S)$ be the set of elements in G whose word lengths are at most n with respect to the generating set S. We also denote S^n instead of $B_n(G,S)$. Let $\beta_n(G,S) = |B_n(G,S)|$. The *exponential growth rate* of (G,S) is defined to be

$$e(G,S) = \lim_{n\to\infty} \beta_n(G,S)^{\frac{1}{n}}.$$

Mathematics Subject Classification 2020: 20F65.
Keywords: growth rates, acylindrically hyperbolic groups, equationally Noetherian groups.

A finitely generated group G has *exponential growth* if there exists a finite generating set S such that $e(G, S) > 1$. The group G has *uniform exponential growth* if there exists $c > 1$, such that for every finite generating set S, $e(G, S) \geq c$.

Given a finitely generated group G, we define

$$e(G) = \inf_{|S|<\infty} e(G, S),$$

where the infimum is taken over all the finite generating sets S of G. Since there are finitely generated groups that have exponential growth but do not have uniform exponential growth [29], the infimum, $e(G)$, is not always realized by a finite generating set for a finitely generated group G.

Given a finitely generated group G, we further define the following set in $\mathbb{R}$:

$$\xi(G) = \{e(G, S) \mid |S| < \infty\},$$

where S runs over all the finite generating sets of G. The set $\xi(G)$ is always countable.

A non-elementary hyperbolic group contains a non-abelian free group; hence, it has exponential growth. In fact, a non-elementary hyperbolic group has uniform exponential growth [18]. Recently it is proved that $\xi(G)$ of a non-elementary hyperbolic group G is well ordered (hence, in particular, has a minimum) [13]. It was new even for free groups.

In this paper, we deal with larger classes of groups. We state a main result. See Definition 1.8 for the definition of acylindricity. See Definition 1.9 for the definition of equational Noetherianity.

Theorem 1.1 (Well-orderedness for acylindrical actions). *Suppose G acts on a δ-hyperbolic space X acylindrically, and the action is non-elementary. Assume that there exists a constant M such that for any finite generating set S of G, the set S^M contains a hyperbolic element on X. Assume that G is equationally Noetherian. Then, $\xi(G)$ is a well-ordered set.*

In particular, $\inf \xi(G)$ is realized by some S, that is, $e(G) = e(G, S)$.

The theorem holds under a weaker assumption, namely, we may replace the acylindricity of the action with the condition that S^M contains a hyperbolic and weak proper discontinuity (WPD) element (Theorem 3.1). See Definition 2.1 for the definition of WPD. Theorem 1.1 is an immediate consequence of Theorem 3.1 by Lemma 2.3. See the explanation at the beginning of Section 3.

We give some applications.

Theorem 1.2 (Theorem 5.4). *Let G be a group that is hyperbolic relative to a collection of subgroups $\{P_1, \dots, P_n\}$. Suppose G is not virtually cyclic, and not equal to P_i for any i. Suppose each P_i is finitely generated and equationally Noetherian. Then $\xi(G)$ is well ordered.*

As examples of this theorem, we prove the following.

Theorem 1.3 (Rank 1 lattices, Theorem 5.5). *Let G be one of the following groups:*

(1) *A lattice in a simple Lie group of rank* 1.

(2) *The fundamental group of a complete Riemannian manifold M of finite volume such that there exist $a, b > 0$ with $-b^2 \leq K \leq -a^2 < 0$, where K denotes the sectional curvature.*

Then $\xi(G)$ is well ordered.

Another family of examples are 3-manifold groups.

Theorem 1.4 (Theorem 5.7). *Let M be a closed orientable* 3*-manifold, and $G = \pi_1(M)$. If M is one of the following, then G has exponential growth and $\xi(G)$ is well ordered:*

(1) *M is not irreducible and G is not isomorphic to $\mathbb{Z}_2 * \mathbb{Z}_2$.*

(2) *M is irreducible, not a torus bundle over a circle, and M has a non-trivial JSJ-decomposition.*

(3) *M admits hyperbolic geometry.*

(4) *M is Seifert fibered such that the base orbifold is hyperbolic.*

Potential examples of Theorem 1.1 are mapping class groups. We discuss this class in Section 5.3. See Example 1.12 for non-examples.

We also show some finiteness result as follows. This was known for hyperbolic groups too [13].

Theorem 1.5 (Theorem 7.1). *Suppose the same assumption holds for G as in Theorem 1.1. Then for any $\rho \in \xi(G)$, up to the action of* Aut(G)*, there are at most finitely many finite generating sets S such that $e(G, S) = \rho$.*

As a part of the proof of the main theorem, we show a basic result on the growth of a group, generalizing a result known for hyperbolic group in [1]. Given a group G that satisfies all the assumptions in Theorem 1.1 except the one that G is equationally Noetherian, there exists a constant $A > 0$ such that for any finite generating set S of G, we have

$$e(G, S) \geq A|S|^A.$$

The constant A depends only on δ and the acylindricity constants. See Proposition 2.10 for the statement. Examples include mapping class groups and rank 1 lattices (see Example 2.11).

We also discuss the set of growth of subgroups in a finitely generated group G. Define

$$\Theta(G) = \{e(H, S) \mid S \subset G, |S| < \infty, H = \langle S \rangle, e(H, S) > 1\}.$$

The set $\Theta(G)$ is countable and contains $\xi(G)$. If G is a hyperbolic group, it is known by [13, Section 5] that $\Theta(G)$ is well ordered. Similarly, we prove the following.

Theorem 1.6 (Theorem 6.6). *Suppose G is one of the groups in Theorem 1.3. Then $\Theta(G)$ is a well-ordered set.*

We also prove the following.

Theorem 1.7 (Finiteness, Theorem 7.15). *Let G be one of the groups in Theorem 1.3. Then for each $\rho \in \Theta(G)$, there are at most finitely many (H, S), up to isomorphism of H, such that S is a finite generating set of $H < G$ with $e(H, S) = \rho$.*

This kind of finiteness is known for hyperbolic groups [13], and we generalize it (Theorem 7.1), which implies the above theorem as examples.

Some more definitions are in order in the following section.

1.2. Acylindrical actions

To generalize the properness of a group action, Bowditch [4] introduced the following definition.

Definition 1.8 (Acylindrical action). An action by isometries of a group G on a metric space X is *acylindrical* if for any $\varepsilon > 0$, there exist $R = R(\varepsilon) > 0$ and $N = N(\varepsilon) > 0$ such that for all $x, y \in X$ with $d(x, y) \geq R$, the set

$$\{g \in G \mid d(x, g(x)) \leq \varepsilon, d(y, g(y)) \leq \varepsilon\}$$

contains at most N elements.

A group G is called an *acylindrically hyperbolic* group [24] if it acts on some δ-hyperbolic space X such that the action is acylindrical and non-elementary. Here, we say the action is *elementary* if the limit set of G in the Gromov boundary ∂X has at most two points. If the action is non-elementary, it is known that G contains hyperbolic isometries on X. Non-elementary hyperbolic groups and non-virtually-abelian mapping class groups are examples of acylindrically hyperbolic groups [4]. There are many other examples.

1.3. Limit groups and equational Noetherianity

Let G be a group and Γ a finitely generated (or countable) group. Let $\mathrm{Hom}(\Gamma, G)$ be the set of all homomorphisms from Γ to G.

A sequence of homomorphisms $\{f_n\}$ from Γ to G is *stable* if for each $g \in \Gamma$, either $f_n(g) = 1$ for all sufficiently large n or $f_n(g) \neq 1$ for all sufficiently large n. If the sequence is stable, then the *stable kernel* of the sequence, $\underrightarrow{\ker}(f_n)$, is defined by

$$\underrightarrow{\ker}(f_n) = \{g \in \Gamma \mid f_n(g) = 1 \text{ for all sufficiently large } n\}.$$

We call the quotient $\Gamma/\underrightarrow{\ker}(f_n)$ a *G-limit group*, or the limit group over G, associated with $\{f_n\}$, and the homomorphism $f : \Gamma \to \Gamma/\underrightarrow{\ker}(f_n)$ the *limit homomorphism*. We say the sequence $\{f_n\}$ *converges* to f.

Let G be a group and $F(x_1,\dots,x_\ell)$ the free group on $X = \{x_1,\dots,x_\ell\}$. For an element $s \in F(x_1,\dots,x_\ell)$ and $(g_1,\dots,g_\ell) \in G^\ell$, let $s(g_1,\dots,g_\ell) \in G$ denote the element after we substitute every x_i with g_i and x_i^{-1} by g_i^{-1} in s. Given a subset $S \subset F(x_1,\dots,x_\ell)$, define

$$V_G(S) = \{(g_1,\dots,g_\ell) \in G^\ell \mid s(g_1,\dots,g_\ell) = 1 \text{ for all } s \in S\},$$

where S is called *a system of equations* (with X the set of variables), and $V_G(S)$ is called *the algebraic set* over G defined by S. We sometimes suppress G from $V_G(S)$.

Definition 1.9 (Equationally Noetherian). A group G is *equationally Noetherian* if for every $\ell \geq 1$ and every subset S in $F(x_1,\dots,x_\ell)$, there exists a finite subset $S_0 \subset S$ such that $V_G(S_0) = V_G(S)$.

Remark 1.10. This definition appears in, for example, [15]. There is another version of the definition that considers $S \subset G * F(x_1,\dots,x_\ell)$, which is originally in [2] and also in [28]. They are equivalent (see [28, Lemma 5.1]).

Notable examples of equationally Noetherian groups include finitely generated free groups [17], linear groups [2], hyperbolic groups without torsion [26] and possibly with torsion [28], and hyperbolic groups relative to equationally Noetherian subgroups [15].

What is important for us is the following general principle.

Lemma 1.11 (Basic principle). *Let $\eta : F \to L$ be the limit homomorphism of a sequence of homomorphisms, $f_n : F \to G$. Suppose G is equationally Noetherian. Then for sufficiently large n, f_n factors through η, namely, there exists a homomorphism $h_n : L \to G$ such that $h_n \circ \eta = f_n$.*

This is elementary but for readers' convenience we include a proof.

Proof. Let $X = \{x_1,\dots,x_\ell\}$ and suppose $F = F(X)$. Let $R = \{r_i\} \subset F(X)$ be a set of defining relations for L. In general, this is an infinite set. Each r_i is a word on X, so that we can see R as a system of equations with X the variable set. Since G is equationally Noetherian, there is a finite subset $R_0 \subset R$ such that $V(R) = V(R_0)$, namely, every solution (an element in G^ℓ) for R_0 is a solution for R.

Now, since $\eta(r_1) = 1$ in L for a large enough n, we have $f_n(r_1) = 1$ in G since η is the limit of $\{f_n\}$. By the same reason, since R_0 is a finite set, there exists N such that for every $n \geq N$, we have $f_n(r_i) = 1$ in G for all $r_i \in R_0$. In other words, if $n \geq N$, then $(f_n(x_1),\dots,f_n(x_\ell)) \in V(R_0)$. But since $V(R_0) = V(R)$, this implies that if $n \geq N$, then $(f_n(x_1),\dots,f_n(x_\ell)) \in V(R)$, namely, $f_n(r_i) = 1$ in G for all $r_i \in R$. Since R is a system of defining relations for L, it implies that each f_n with $n \geq N$ factors through $\eta : F \to L$. ∎

Regarding Theorem 1.1, one cannot omit the assumption that G is equationally Noetherian. The following example is pointed out by Ashot Minasyan. A group G is called *Hopfian* if every surjective homomorphism $f : G \to G$ is an isomorphism.

Example 1.12. Take a finitely generated non-Hopfian group, G, for example $BS(2,3)$, (a Baumslag–Solitar group). Put $H = G * \mathbb{Z}$. Then H is non-Hopfian, that is, there exists a surjection $f : H \to H$ that is not an isomorphism. It is a standard fact that a finitely generated equationally Noetherian group is Hopfian; therefore, H is not equationally Noetherian.

But all other assumptions in Theorem 1.1 are satisfied by H. Let T be the Bass–Serre tree for $G * \mathbb{Z}$. The tree T is 0-hyperbolic. The action of H on T is acylindrical and non-elementary. Also, for any finite generating set S of H, it is a well-known lemma (due to Serre, cf. [6]) that S^2 contains a hyperbolic isometry on T.

We verify that $\xi(H)$ is not well ordered. It suffices to argue that there is no finite generating set S with $e(H) = e(H,S) > 1$. First, it is known [25] that a finitely generated group K that is a free product is *growth-tight*, namely, for any surjective homomorphism $h : K \to K$ that is not an isomorphism, and for any finite generating set S of K, $e(K,S) > e(K,h(S))$. Now, it follows that $e(H)$ is not achieved by any S, since if it did, then take such S. But then, the non-isomorphic, surjective homomorphism f in the above would imply $e(H,S) > e(H,f(S))$, a contradiction.

1.4. Family of groups

The whole paper is concerning the set of growth rates of one group, but it is tempting to deal with the set of growth of an infinite family of groups. Some of the key propositions in this paper hold for a family of groups by a straightforward modification of the proofs. In view of that, we state variations of some results for a family of groups. See Propositions 4.9 and 7.12. We are planning papers dealing with the fundamental groups of hyperbolic manifolds and 3-manifolds.

2. Lower bound of a growth rate

Although most statements in the paper are for geodesic spaces X, *we consider a graph for X instead of a geodesic space in the arguments throughout the paper unless we indicate otherwise.* The advantage is that an infimum is achieved for various notions, for example, $L(S)$, in the arguments. But, we do not lose generality by assuming that X is a graph since we can always consider the 1-skeleton of a Rips complex of a geodesic space with a group action. By doing so, the various assumptions we consider (such as the hyperbolicity of the space, acylindricity of the action) remain valid, maybe with slightly difference constants.

2.1. Hyperbolic isometries and axes

Suppose a group G acts on a δ-hyperbolic, geodesic space X by isometries. Choose a base point $x \in X$ and for $g \in G$, put

$$L(g) = \inf_{x \in X} |x - g(x)|, \quad \lambda(g) = \lim_{n \to \infty} \frac{|x - g^n(x)|}{n},$$

where $L(g)$ is called the *minimal displacement*, $\lambda(g)$ is called the *translation length* and does not depend on the choice of x, and for any $n > 0$, we have $\lambda(g^n) = n\lambda(g)$. We also have

$$\lambda(g) \leq L(g) \leq \lambda(g) + 7\delta.$$

The first inequality is trivial and we leave the second as an exercise (e.g., use [1, Corollary 1]). For a finite set $S \subset G$ and $x \in X$, define

$$L(S, x) = \max_{s \in S} |x - s(x)|,$$

then define

$$L(S) = \inf_{x \in X} L(S, x).$$

We recall a few definitions and facts from δ-hyperbolic spaces. An isometry g of a hyperbolic space X is called *elliptic* if the orbit of a point by g is bounded, and *hyperbolic* if there are $x \in X$ and $C > 0$ such that for any $n > 0$, we have $|x - g^n(x)| > Cn$. The element g is hyperbolic iff $\lambda(g) > 0$.

A hyperbolic isometry g is associated with a bi-infinite quasi-geodesic, γ, called an *axis* in X. If there exists a bi-infinite geodesic γ that is invariant by g, that would be an ideal choice for an axis, but that is not always the case.

As a remedy, if $L(g) \geq 10\delta$, take a point $x \in X$ where $L(g)$ is achieved. Then take a geodesic $[x, g(x)]$ between x and $g(x)$ and consider the union of its g-orbit, which defines a g-invariant path (see, e.g., [9]). If $L(g) < 10\delta$, then take $n > 0$ such that $L(g^n) \geq 10\delta$ and apply the construction to g^n, and use this path for g, which is not g-invariant. We denote this axis as $A(g)$ in this paper. Also, for g^n with $g \neq 0$, we may also take $A(g)$ as an axis for g^n.

For g, an axis $A(g)$ is not unique, but uniformly (over all hyperbolic g) quasi-geodesic, such that for any two points $x, y \in A(g)$, the Hausdorff distance between the segment between x, y on $A(g)$ and a geodesic between x, y, $[x, y]$ is at most 10δ. Also, if $Hd(A(g), h(A(g))$ for $h \in G$ is finite, then it is bounded by 10δ, where Hd is the Hausdorff distance. We sometimes call $A(g)$ a 10δ-*axis*. We consider a direction on the 10δ-axis using the action of g.

A hyperbolic isometry g defines two limit points in ∂X, the visual (Gromov) boundary of X, by $g^\infty = \lim_{n \to \infty} g^n(x), g^{-\infty} = \lim_{n \to -\infty} g^n(x)$, where x is a base point. We say two hyperbolic isometries g, h are *independent* if $\{g^{\pm\infty}\}$ and $\{h^{\pm\infty}\}$ are disjoint. If the Hausdorff distance between two axes is finite, then we say they are *parallel*.

2.2. WPD elements

We consider another version of properness of a group action that is weaker than acylindricity.

Definition 2.1 (WPD, uniformly WPD, D-WPD). Let G acts on a δ-hyperbolic space X. Suppose that $g \in G$ is hyperbolic on X. We say g is *WPD* if there is a 10δ-axis, γ, of g such that for any $\varepsilon > 0$, there exists $D = D(\varepsilon) > 0$ such that for any $x, y \in \gamma$ with $|x - y| \geq D\lambda(g)$, the number of the elements in the following set is at most D:

$$\{h \in G \mid |h(x) - x| \leq \varepsilon, |h(y) - y| \leq \varepsilon\}. \tag{2.1}$$

In application we often take $y = g^D(x)$. If we want to make the function D explicit, we say that g is D-WPD, or WPD with respect to D. We say that D is a function for WPD.

If there is a function D such that if a set of hyperbolic elements in G is D-WPD, then we say they are *uniformly WPD*, or *uniformly D-WPD*. If all hyperbolic elements in G are uniformly D-WPD, then we say the action is uniformly (D-)WPD.

Some remarks are in order. The notion of WPD was introduced in [3], where the function D is not used, but the definitions are equivalent.

If g is D-WPD, then it is D'-WPD for any function D' such that $D'(\varepsilon) \geq D(\varepsilon)$ for all ε. So, without loss of generality, we assume that $D(\varepsilon)$ does not decrease when we increase ε. We often use the value $D(100\delta)$ in this paper; for example, see Lemma 2.2. For convenience we also assume that $D(100\delta) \geq 50$, which we use in the proof of Lemma 7.4.

The choice of a 10δ-axis is not important in the definition. Also, one can use C-axis for any $C > 0$. It only changes the function $D(\varepsilon)$. Uniformly WPD is related to but weaker than the notion of weak acylindricity in [9]. See also [9, Example 1] for the difference between acylindricity and variations of WPD.

For an acylindrical action, it is known [12, Lemma 2.1] that there exists $T > 0$ such that for any hyperbolic element g, we have $\lambda(g) \geq T$. This holds for uniformly WPD actions too, and the argument is same, but for the readers' convenience, we prove it.

Lemma 2.2 (Lower bound on $\lambda(g)$). *Suppose G acts on a δ-hyperbolic space X. Let $g \in G$ be hyperbolic with a 10δ-axis γ:*

(1) *If g is D-WPD, then $\lambda(g) \geq \frac{50\delta}{D(100\delta)}$.*

(2) *If the action is acylindrical, then $\lambda(g) \geq \frac{50\delta}{N(150\delta)}$.*

Proof. (1) Set $\mathcal{D} = D(100\delta)$. Let $x, y \in \gamma$ with $|x - y| \geq \mathcal{D}\lambda(g)$. Then there must be some n with $0 \leq n \leq \mathcal{D}$ such that $|x - g^n(x)| > 100\delta$ or $|y - g^n(y)| > 100\delta$. This is because otherwise, all the elements $1, g, \ldots, g^{\mathcal{D}}$, which are $\mathcal{D} + 1$ distinct elements, are contained in the set

$$\{h \in G \mid |h(x) - x| \leq 100\delta, \text{and } |h(y) - y| \leq 100\delta\}.$$

This is impossible since the set contains at most $\mathcal{D}$ elements. Now suppose, say, x satisfies $|x - g^n(x)| > 100\delta$. Then it implies that $\lambda(g^n) \geq 50\delta$. Since $n \leq \mathcal{D}$, we have $\lambda(g) \geq \frac{50\delta}{\mathcal{D}}$.

(2) The argument is similar to (1). Take $x, y \in \gamma$ with $|x - y| = R(150\delta)$. If $|x - g^n(x)| \leq 100\delta$ for some n, then $|y - g^n(y)| \leq 150\delta$. This implies, by acylindricity, there must be n with $1 \leq n \leq N(150\delta)$ such that $|x - g^n(x)| > 100\delta$. It follows $\lambda(g^n) \geq 50\delta$, so that $\lambda(g) \geq \frac{50\delta}{N(150\delta)}$. ∎

For a function $D(\varepsilon)$, put

$$T = \frac{50\delta}{D(100\delta)},$$

then by Lemma 2.2 (1) we have $\lambda(g) \geq T$ for a D-WPD element g.

The following lemma is straightforward.

Lemma 2.3 (Acylindricity implies uniform WPD). *If an action of G on a δ-hyperbolic space X is acylindrical, then it is uniformly WPD.*

Proof. Suppose $g \in G$ is hyperbolic on X, and let γ be a 10δ-axis. Let $R(\varepsilon), N(\varepsilon)$ be the acylindricity constants. Also, let $T > 0$ be a uniform bound for $\lambda(g) \geq T$ by Lemma 2.2 (2). T does not depend on g nor ε. Suppose ε is given. Let $K = K(\varepsilon)$ be a smallest integer with $K \geq \frac{R(\varepsilon)}{T}$. The constant K does not depend on g. Put $D(\varepsilon) = \max\{K(\varepsilon), N(\varepsilon)\}$. We will show that the action is uniformly D-WPD.

Let $x, y \in \gamma$ be such that $|x - y| \geq D\lambda(g)$. Then, the right-hand side satisfies $D\lambda(g) \geq K\lambda(g) \geq KT \geq R(\varepsilon)$. Hence, by the acylindricity, there are at most $N(\varepsilon)$ elements which simultaneously move each of x, y by at most ε. Since $D(\varepsilon) \geq N(\varepsilon)$, the action is D-WPD. ∎

We state a lemma which is useful for us.

Lemma 2.4. *Suppose there are at most D elements in the set defined by (2.1) in Definition 2.1 for $\varepsilon = 100\delta$ if $|x - y| \geq D\lambda(g)$. Then g is WPD, and moreover, there is a function D' with $D'(100\delta) = D$ that depends only on D, δ such that g is D'-WPD.*

Proof. Let γ be a 10δ-axis of g. Suppose $\varepsilon > 0$ is given. We may assume $\varepsilon > 100\delta$. Take $x, y \in \gamma$ such that $|x - y| \geq D\lambda(g) + 2\varepsilon + 1000\delta$. Take $p, q \in \gamma$ between x and y with $|x - p| = |y - q| = \varepsilon + 50\delta$. Then $|p - q| \geq D\lambda(g) + 800\delta$.

Let J be the collection of elements j in G such that $|x - j(x)| \leq \varepsilon$ and $|y - j(y)| \leq \varepsilon$. If $j \in J$, then $|p - j(p)| \leq \varepsilon + 30\delta$ and $|q - j(q)| \leq \varepsilon + 30\delta$, and $j(p), j(q)$ are in the 15δ-neighborhood of γ. It implies that J contains a subset J_0 that contains at most $(2\varepsilon + 200\delta)/(10\delta)$ elements such that for any $j \in J$, one can find $j_0 \in J_0$ with $|p - j^{-1} j_0(p)| \leq 70\delta$. To see it, consider points $p_1, p_2 \in \gamma$ with $|p_1 - p_2| = 2\varepsilon + 200\delta$ such that p is the mid point of the segment between p_1, p_2 on γ, which we denote by $[p_1, p_2]$. Then the points $j(p)$ with $j \in J$ is contained in the 30δ-neighborhood of $[p_1, p_2]$. One should imagine that this neighborhood is a narrow tube around $[p_1, p_2]$.

Now by a pigeon hole argument, one can find a desirable subset J_0. (Notice that $|p - j^{-1} j_0(p)| = |j(p) - j_0(p)|$, so that one needs to find a point $j_0(p)$ near (i.e., at most 70δ) a given point $j(p)$, which is possible.)

But $|p - j^{-1} j_0(p)| \leq 70\delta$ implies $|q - j^{-1} j_0(q)| \leq 100\delta$. This is because $|p - q| \geq 800\delta$ and both $[j(p), j(q)]$ and $[j_0(p), j_0(q)]$ are contained in the 20δ-neighborhood of γ. We have shown that the element $j^{-1} j_0$ moves both p, q by at most 100δ. By our assumption, there are at most D possibilities for such element. In conclusion, J contains at most $D \times (2\varepsilon + 200\delta)/(10\delta)$ elements. We proved that g is WPD.

We compute a WPD-function for g, which we denote by D'. By assumption, we may set $D'(100\delta) = D$. First,

$$D \times \frac{2\varepsilon + 200\delta}{10\delta} = D \times \left(\frac{\varepsilon}{5\delta} + 20\right).$$

Next,

$$D + \frac{2\varepsilon + 1000\delta}{\lambda(g)} \geq D + D \times \frac{2\varepsilon + 1000\delta}{50\delta} = D \times \left(21 + \frac{\varepsilon}{25\delta}\right)$$

by Lemma 2.2 (1). So, if $\varepsilon > 100\delta$, we set

$$D'(\varepsilon) = D \max\left\{20 + \frac{\varepsilon}{5\delta}, 21 + \frac{\varepsilon}{25\delta}\right\} = D \times \left(21 + \frac{\varepsilon}{5\delta}\right). \qquad \blacksquare$$

2.3. Elementary closure

Suppose G acts on a hyperbolic space X and let $g \in G$ be a hyperbolic isometry with an axis γ. The *elementary closure* of g is defined by

$$E(g) = \{h \in G \mid Hd(\gamma, h(\gamma)) < \infty\}.$$

It turns out that $E(g)$ is a subgroup of G. Clearly, $\langle g \rangle < E(g)$.

We denote the a-neighborhood of a subset $Y \subset X$ by $N_a(Y)$.

Lemma 2.5 (Parallel axes). *Suppose G acts on a δ-hyperbolic space X. Let $g \in G$ be hyperbolic with a 10δ-axis γ. Let $h \in G$, then*

(1) *If $h \in E(g)$, then $Hd(\gamma, h(\gamma)) \leq 50\delta$.*

(2) *Assume that g is D-WPD. If $h \notin E(g)$, then the diameter of $h(\gamma) \cap N_{50\delta}(\gamma)$ is at most*

$$2D(100\delta)L(g) + 100\delta.$$

This lemma is well known in slightly different versions, for example [12, Lemma 2.2] for acylindrical actions, so the proof will be brief.

In general, if two axes have finite Hausdorff distance, then we say they are *parallel*.

Proof. (1) By definition, $Hd(\gamma, h(\gamma)) < \infty$. Since both $\gamma, h(\gamma)$ are 10δ-axes, we have a desired bound.

(2) Suppose not. Suppose that the direction of $\gamma, h\gamma$ coincides along the parallel part. Set $\mathcal{D} = D(100\delta)$. Take $x \in h(\gamma)$ near one end of the intersection $h(\gamma) \cap N_{50\delta}(\gamma)$ such that $g^n(x)$ for $0 \le n \le 2\mathcal{D}$ are in the 50δ-neighborhood of the intersection. This is possible since the intersection is long enough. Consider the points $x, g^{\mathcal{D}}(x)$. Then $|x - g^{\mathcal{D}}(x)| \ge \mathcal{D}\lambda(g)$. Letting $(hgh^{-1})^{-n}g^n$ with $0 \le n \le \mathcal{D}$ act on $x, g^{\mathcal{D}}(x)$, we have

$$|x - (hgh^{-1})^{-n}g^n(x)| \le 100\delta, \quad |g^{\mathcal{D}}(x) - (hgh^{-1})^{-n}g^n(g^{\mathcal{D}}(x))| \le 100\delta$$

for all $0 \le n \le \mathcal{D}$.

But since g is D-WPD, there are at most $\mathcal{D}$ such elements, so that it must be that for some $n \ne m$, we have $(hgh^{-1})^{-n}g^n = (hgh^{-1})^{-m}g^m$, so that g^{n-m} and h commute. It implies that γ and $h(\gamma)$ are parallel, a contradiction.

If the direction for γ and $h\gamma$ are opposite, we consider $(hgh^{-1})^n g^n$ instead of $(hgh^{-1})^{-n}g^n$, and the rest is same. ■

We quote a fact.

Proposition 2.6 (Elementary closure). *Suppose G acts on a δ-hyperbolic space X. Let $g \in G$ be a hyperbolic element on X. Assume g is D-WPD. Then,*

(1) *$E(g)$ is virtually $\mathbb{Z}$, and contains $\langle g\rangle$ as a finite index subgroup.*

(2) *If $h \in G$ is a hyperbolic element such that g and h are independent, then $E(g) \cap E(h)$ is finite.*

Proof. (1) This is [8, Lemma 6.5].

(2) If $E(g) \cap E(h)$ is infinite, then it contains $\langle g^N\rangle$ for some $N > 0$. But since g and h are independent, for a sufficiently large m, we have $g^m \notin E(h)$, impossible. ■

Note that under the assumption of the proposition, the action of G is elementary if and only if G is virtually $\mathbb{Z}$, which is equivalent to that $G = E(g)$ in this case.

Several comments are in order. We will refer to some of them later. In general, if a group H is virtually $\mathbb{Z}$, then there exists an exact sequence

$$1 \to F \to H \to C \to 1,$$

where C is either $\mathbb{Z}$ or $\mathbb{Z}_2 * \mathbb{Z}_2$, and F is finite. In the case that H is $E(g)$ in the above, we denote the finite group F by $F(g)$. If $C = \mathbb{Z}$, $F(g)$ is the set of elements of finite order in $E(g)$.

The axis γ defines two points in the ideal boundary of X, which we denote $\{\gamma(\infty), \gamma(-\infty)\}$. $E(g)$ is exactly the set of elements that leaves this set invariant. If $h \in E(g)$ swaps those two points, we say it *flips* the axis since it flips the direction of the axis γ.

If the action of G on X is D-uniform WPD, then for any hyperbolic g, we have $|F(g)| \le 2D(100\delta)$. Moreover, if $C = \mathbb{Z}$, then $|F(g)| \le D(100\delta)$.

Note that $\langle g \rangle < E(g)$. If the action is non-elementary, then $E(g) \neq G$, which is equivalent to that G is not virtually $\mathbb{Z}$, so that G contains two independent hyperbolic isometries.

We say a hyperbolic element g is *primitive* if $C = \mathbb{Z}$ and each element $h \in E(g)$ is written as $h = fg^n$ for some $f \in F(g)$ and $n \in \mathbb{N}$.

2.4. About the constants

From now on, there will be many constants to make the argument concrete and precise. In the argument we consider a sequence of generators S_n. It would be a good idea to keep in mind that there are two kinds of constants.

The first kind are those that are fixed once the constants δ, $D(100\delta)$, M are given by the action:

$$\delta,\ D(100\delta),\ M; \quad T,\ k,\ m,\ b.$$

The constants k, m will appear in Lemma 2.8 and b in the proof of Lemma 4.2.

The second kind are those that depend on a generating set S_n of G:

$$L(S_n), \quad \lambda(g), \quad L(g), \quad \lambda(u), \quad L(S_n^{2MD}), \quad \Delta_n.$$

If $L(S_n)$ is bounded, then all of the constants in the second will be bounded, but if not, then, roughly they all diverge in the same order as $L(S_n)$.

We remark that if $L(S_n)$ diverges, then one way to argue is to rescale X by $\frac{1}{L(S_n)}$, then go to the limit, which is a tree, and use the geometry of the tree. This approach is the one taken in [13], where only this case happens. But in our setting, the new feature is that possibly, $L(S_n)$ is bounded. In this paper, we use a unified approach.

For convenience, we assume $\delta > 0$ from now on.

2.5. Lower bound of a growth rate

Lemma 2.7 (Hyperbolic element of large displacement). *Let X be a δ-hyperbolic space and S a finite set of isometries of X. Suppose that $L(S) \geq 30\delta$. Let $x \in X$ be such that $L(S) = L(S, x)$. Then there is a hyperbolic element $g \in S^2$ such that*

$$L(S) - 8\delta \leq |x - g(x)|$$

and

$$|x - g(x)| - 16\delta \leq L(g).$$

This is exactly same as a part of [1, Lemma 7] (in that paper, our element g is denoted as b in the proof), although their setting is that X is a Cayley graph of a hyperbolic group G and $S \subset G$. As we said in the beginning of this section, we only consider a graph for X, so that a point $x \in X$ in the lemma always exists. The proof is same verbatim after a suitable translation of notions, so we omit it (cf. [6, Theorem 1.4] for somewhat similar result).

Note that if $s \in S$ then $|x - s(x)| \leq L(S)$, and if $s \in S^2$ then $L(s) \leq |x - s(x)| \leq 2L(S)$ by the definition of $L(S)$ and the choice of x.

We summarize the properties of g we use later:

- $L(S) - 24\delta \leq L(g) \leq 2L(S)$.
- The distance between x and the 10δ-axis of g is at most 20δ.

The second one follows from $L(g) = \min_{y \in X} |y - g(y)| \geq |x - g(x)| - 16\delta$.

Lemma 2.8 (Free subgroup with primitive hyperbolic elements). *Let X be a δ-hyperbolic geodesic space and G a group acting on X. Assume G is not virtually cyclic. Let $D(\varepsilon)$ be a function for WPD. Set*

$$k = 60D(200\delta), \quad m = 66k + 4 = 3960D(200\delta) + 4.$$

Let S be a finite set that generates G with $L(S) \geq 50\delta$, and $x \in X$ with $L(S) = L(S, x)$. Suppose $g \in S^2$ is a hyperbolic element with a 10δ-axis γ with $d(x, \gamma) \leq 20\delta$ such that

$$L(g) \geq |x - g(x)| - 16\delta \geq L(S) - 24\delta.$$

Assume that g is D-WPD.

Then there exists $s \in S$ such that $g^k, sg^k s^{-1}$ are independent and freely generate a rank 2 free group $F < G$ with the following property. There is a WPD-function $D'(\varepsilon)$ with

$$D'(100\delta) = D(200\delta)$$

such that every non-trivial element $h \in F$ is hyperbolic on X and D'-WPD. Also, h satisfies:

$$\lambda(h) \geq 10(2D(200\delta)L(g) + 100\delta) \geq 10L(S).$$

Moreover, there is an element $u \in F$ that satisfies:

(1) *u has a 10δ-axis α with $d(x, \alpha) \leq 50\delta$.*

(2) *u is primitive, namely, there exists*

$$1 \to F(u) \to E(u) \to \mathbb{Z} \to 1$$

such that any element $h \in E(u)$ is written as $h = fu^p$ with $f \in F(u)$ and $p \in \mathbb{Z}$.

(3) *$|F(u)| \leq D'(100\delta) = D(200\delta)$.*

(4) *$u \in S^m$.*

By $L(S) \geq 50\delta$, we have $L(g) \geq 26\delta$, $2L(g) \geq L(S)$, and $2\lambda(g) \geq L(g)$.

Remark 2.9. In the proof of Proposition 2.10, we will apply Lemma 2.8 to $S^{MD(200\delta)}$ instead of a generating set S itself. In that case, the element $s \in S^{MD(200\delta)}$ that appears in the above lemma can be chosen from S itself. This is because we choose such s in the beginning of the proof of the lemma and the rest of the argument is exactly same.

Proof. Set $\mathcal{D} = D(200\delta)$ and $T = \frac{50\delta}{\mathcal{D}}$. Since g is D-WPD, by Lemma 2.2 (1), we have $\lambda(g) \geq T$. Remember that by our assumption, we always have $D(100\delta) \leq D(200\delta)$, so that the estimate in the lemma holds for $D(200\delta)$ as well.

Since G is not virtually cyclic, there is an element $s \in S$ with $s \notin E(g)$. By Lemma 2.5 (1), the diameter of the intersection $s(\gamma) \cap N_{50\delta}(\gamma)$ is at most

$$2\mathcal{D}L(g) + 100\delta.$$

In view of this, set

$$k = 10\Big(4\mathcal{D} + \frac{100\delta}{T}\Big) = 60\mathcal{D} \geq 10.$$

Then

$$\lambda(g^k) \geq k\lambda(g) \geq 10(4\mathcal{D}\lambda(g) + 100\delta) \geq 10(2\mathcal{D}L(g) + 100\delta).$$

Here, we used $2\lambda(g) \geq L(g), \lambda(g) \geq T$.

Note that sgs^{-1} is hyperbolic and D-WPD with a 10δ-axis $s(\gamma)$. Since $\lambda(g^k)$ is at least 10 times longer than the above intersection, g^k and sg^ks^{-1} freely generate a free group, F. Also, its non-trivial elements h are hyperbolic and

$$\lambda(h) \geq \lambda(g^k) \geq 10(2\mathcal{D}L(g) + 100\delta) \geq 10L(S).$$

For the last inequality, we used $2L(g) \geq L(S)$.

We argue that there is a function $D'(\varepsilon)$ with $D'(100\delta) = D(200\delta)$ such that every non-trivial $h \in F$ is uniformly D'-WPD. We only need to argue for $\varepsilon = 100\delta$ to check WPD by Lemma 2.4. Let γ be a 10δ-axis of h. We will show that for any $y, z \in \gamma$ with $|y - z| \geq 3\lambda(h)$, there are at most $D(200\delta)$ elements $j \in G$ satisfying

$$|j(y) - y| \leq 100\delta \quad \text{and} \quad |j(z) - z| \leq 100\delta.$$

We denote the collection of those elements j by J. We will use the fact that the axis of h is, roughly speaking, a concatenation of some translates of the segment $[x, g^k(x)]$ by elements in F (cf. a more precise description of the axis of the element u below.) Also, we point out that without loss of generality, one may take a conjugate of h in G in the argument. Since h is a word on g^k and sg^ks^{-1}, by taking a conjugate of h in G, one may assume that the segment $[x, g^k(x)]$ is contained in the 30δ-neighborhood of the segment $[y, z]$ except for some small neighborhood of x and $g^k(x)$. Taking a further conjugate of h if necessary, one may assume that the segment $[x, g^{k/2}(x)]$ is contained in the 30δ-neighborhood of $[y, z]$. Here, we used $|y - z| \geq 3\lambda(h)$.

That implies that one can find points $p, q \in [x, g^{k/2}(x)]$ with $|p - q| \geq \frac{k}{4}\lambda(g) = 15\mathcal{D}\lambda(g)$ such that $|p - j(p)| \leq 150\delta$ and $|q - j(q)| \leq 150\delta$ if $j \in J$.

But since g is D-WPD where $\mathcal{D} = D(200\delta)$, the set J contains at most $D(200\delta)$ elements. (Strictly speaking the points p and q are maybe not exactly on a 10δ-axis of g, but one can choose nearby points of p and q on the axis and argue.) We showed that h is uniformly WPD.

Now we argue for the moreover part. We define an element u by

$$u = g^k (sg^{10k} s^{-1}) g^{20k} (sg^k s^{-1}) g^k.$$

(1) To construct an axis α of u, consider the following (see Figure 1):

$$[x, g^k(x)] \cup g^k s[x, g^{10k}(x)] \cup g^k(sg^{10k}s^{-1})[x, g^{20k}(x)] \\ \cup g^k(sg^{10k}s^{-1})g^{20k}s[x, g^k(x)] \cup g^k(sg^{10k}s^{-1})g^{20k}(sg^k s^{-1})[x, g^k(x)],$$

which consists of five geodesic segments, which we call *pieces*. We name them as $P_1, \dots, P_5$. Their length are $U, 10U, 20U, U, U$ for some constant $U > 0$ maybe with some error up to 100δ. Remember that

$$U \geq \lambda(g^k) \geq 10L(S).$$

Note that P_1 is contained in γ, which is the axis of g. Put $\alpha_1 = \gamma$. Similarly, $P_2, \dots, P_5$ are contained in the axes of the conjugates of g by $g^k s$, $g^k(sg^{10k}s^{-1})$, $g^k(sg^{10k}s^{-1})g^{20k}s$, $g^k(sg^{10k}s^{-1})g^{20k}(sg^k s^{-1})$, respectively. We will call those axes $\alpha_2, \alpha_3, \alpha_4, \alpha_5$.

There are four short ($\leq L(S)$) gaps between P_i and P_{i+1}. We put (short) geodesics between the gaps and obtain a path, $\overline{\alpha}$. Then take the union of the u-orbit of $\overline{\alpha}$, which we call $\hat{\alpha}$. This is an axis for u since g and sgs^{-1} are independent and their 10δ-axes stay close ($\leq 50\delta$) to each other along a short (compared to U) segment.

We call the image of a piece P_i by $u^p (p \in \mathbb{Z})$ also a piece. The axis $\hat{\alpha}$ is a sequence of pieces (with short gaps in between).

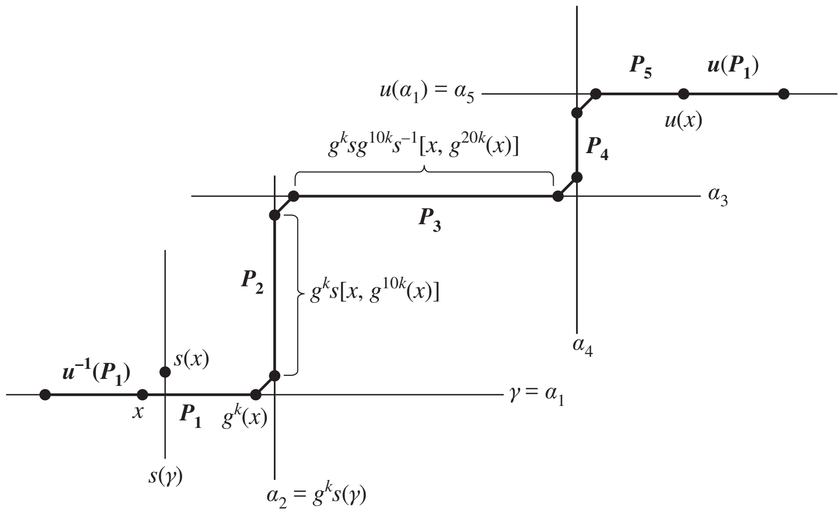


Figure 1. The thick line is $\overline{\alpha}$. It is a broken geodesic from x to $u(x)$ with four short "gaps." The first gap is $[g^k(x), g^k s(x)]$.

By construction, $x \in \widehat{\alpha}$, but maybe $\widehat{\alpha}$ is not exactly a 10δ-axis for u, so we take a 10δ-axis, α. One can check that $d(x, \alpha) \leq 20\delta$. (This is the reason we put g^k at the end of u. Without g^k at the end, u is still hyperbolic.) Also, $\widehat{\alpha}$ and α stay close to each other in the sense that most part of each piece in $\widehat{\alpha}$, except for some short parts near the two ends, is in the 20δ-neighborhood of α.

(2) If two axes β, β' are parallel, we write $\beta \sim \beta'$. Since $s \notin E(g)$, $\alpha_1 \not\sim \alpha_2$. Also, $\alpha_2 \not\sim \alpha_3, \alpha_3 \not\sim \alpha_4$, and $\alpha_4 \not\sim \alpha_5$.

We first argue that $E(u)$ maps to $\mathbb{Z} = C$ in the exact sequence after Proposition 2.6. Suppose $h \in E(u)$. By definition, $\alpha \sim h(\alpha)$. It suffices to show that h does not flip the direction of α. For that, we examine the sequence of the lengths of the pieces on α and on $h(\alpha)$. Here, we say that $h(P)$ is a piece if P is a piece in α.

The sequence of the length for the part $P_1, \ldots, P_5$ is $U, 10U, 20U, U, U$, so that on α, the sequence is (from the left to the right on the top in Figure 2):

$$\cdots ; U, 10U, 20U, U, U; \quad U, 10U, 20U, U, U; \cdots$$

Now, if the direction of $h(\alpha)$ was opposite to α, then the sequence on $h(\alpha)$ would be (from the left to the right in the figure):

$$\cdots ; U, U, 20U, 10U, U; \quad U, U, 20U, 10U, U; \cdots$$

From those two sequences we observe that one of the following three cases must hold:

(i) There is a piece P on $h(\alpha)$ that intersects both $N_{50\delta}(P_1)$ and $N_{50\delta}(P_2)$ at least $\frac{U}{2}$ in diameter, or

(ii) There is a piece P on $h(\alpha)$ that intersects both $N_{50\delta}(P_2)$ and $N_{50\delta}(P_3)$ at least $\frac{U}{2}$ in diameter, or

(iii) There is a piece P on $h(\alpha)$ that intersects both $N_{50\delta}(P_3)$ and $N_{50\delta}(P_4)$ at least $\frac{U}{2}$ in diameter.

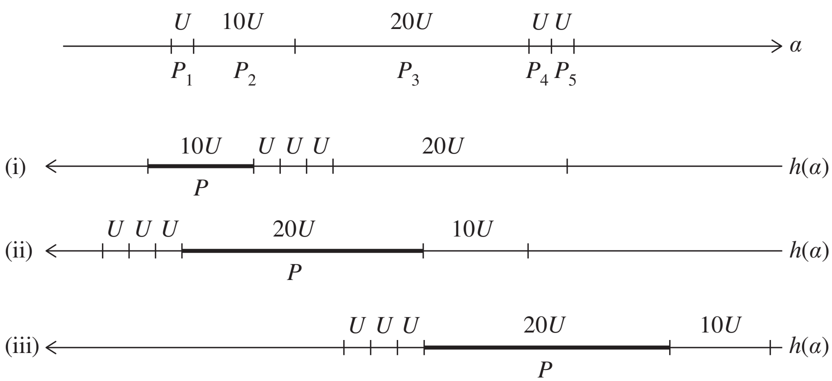


Figure 2. The direction of α is to the right, and the direction of $h(\alpha)$ is to the left. The figure indicates the three positions of P in bold line.

In the above, P has length (approximately) $5U$ or $10U$.

Let β be the axis that contains the piece P. Then, since $\frac{U}{2}$ is at least $5(2\mathfrak{D}L(g) + 100\delta)$, Lemma 2.5 (2) would imply that (i) $\alpha_1 \sim \beta \sim \alpha_2$, (ii) $\alpha_2 \sim \beta \sim \alpha_3$, or (iii) $\alpha_3 \sim \beta \sim \alpha_4$, respectively. In either case, we obtain a contradiction since $\alpha_1 \not\sim \alpha_2$, $\alpha_2 \not\sim \alpha_3$, and $\alpha_3 \not\sim \alpha_4$. We showed that h does not flip α.

By the same reason, that is, the constrain from the combinatorics, $h(x)$ must be close to $u^p(x)$ for some $p \in \mathbb{Z}$, namely, the distance is at most $2\mathfrak{D}L(g) + 100\delta$. This implies that $\alpha_1 \sim (hu^{-p})(\alpha_1)$ and also, $\alpha_2 \sim (hu^{-p})(\alpha_2)$. It then follows that

$$hu^{-p} \in E(g), \quad hu^{-p} \in E(g^n sgs^{-1} g^{-n}),$$

which implies $g^k(hu^{-p})g^{-k} \in (E(g) \cap E(sgs^{-1}))$. But the right-hand side is a finite group, so that hu^{-p} has finite order; therefore, $hu^{-p} \in F(g)$. And, of course, $hu^{-p} \in F(u)$ namely, there is $f \in F(u)$, with $h = fu^p$.

(3) This is a consequence of the D'-uniform WPD, and is mentioned in the paragraphs following Proposition 2.6.

(4) Since $g \in S^2$ and $s \in S$, we have $u \in S^{14k+4}$. We are done since $m = 66k + 4$. ■

Proposition 2.10 (Lower bound of growth). *Suppose G acts on a δ-hyperbolic space X. Let $D(\varepsilon)$ be a function for WPD. Assume that there exists a constant M such that for any finite generating set S of G, the set S^M contains a hyperbolic element h that is D-WPD.*

Set

$$A = \frac{1}{79220M(D(200\delta))^3} > 0.$$

Then for any finite generating set S of G, we have $e(G, S) \geq A|S|^A$.

This result is a generalization of the result on hyperbolic groups by [1]. We adapt their argument to our setting, which is straightforward.

Proof. Set $\mathfrak{D} = D(200\delta)$. Since the element $h \in S^M$ is D-WPD, we have $\lambda(h) \geq \frac{50\delta}{\mathfrak{D}}$ by Lemma 2.2 (1). It implies that $\lambda(h^{\mathfrak{D}}) \geq 50\delta$. Since $h^{\mathfrak{D}} \in S^{M\mathfrak{D}}$, we have $L(S^{M\mathfrak{D}}) \geq 50\delta$.

Lemma 2.7 applies to $S^{M\mathfrak{D}}$ since $L(S^{M\mathfrak{D}}) \geq 50\delta$. By Lemma 2.7 applied to $S^{M\mathfrak{D}}$ with x such that $L(S^{M\mathfrak{D}}) = L(S^{M\mathfrak{D}}, x)$, there is $g \in S^{2M\mathfrak{D}}$ such that

$$L(S^{M\mathfrak{D}}) - 8\delta \leq |x - g(x)|$$

and the 10δ-axis γ of g satisfies $d(x, \gamma) \leq 20\delta$.

Then Lemma 2.8 applies to $S^{M\mathfrak{D}}$ and $g \in S^{2M\mathfrak{D}}$. By Lemma 2.8 applied to $S^{M\mathfrak{D}}$ and g, there exists $s \in S^{M\mathfrak{D}}$ such that $\langle g^k, sg^k s^{-1}\rangle = F$ and F contains u that is primitive such that

$$|F(u)| \leq \mathfrak{D}; \quad u \in S^{M\mathfrak{D}m}; \quad L(u) \geq 10L(S^{M\mathfrak{D}}),$$

where $k = 60\mathfrak{D}$, $m = 3960\mathfrak{D} + 4$, and that u is D'-WPD, where $D'(100\delta) = D(200\delta)$.

We note that the element s can be chosen from S since G is not virtually cyclic. This can be easily seen in the proof of Lemma 2.8 since we only need to choose s such that $s \notin E(g)$ (see Remark 2.9).

Now, take a maximal subset $W \subset S$ such that any two distinct elements $w, v \in W$ are in different $F(u)$-(right) cosets. Then, $|W| \geq \frac{|S|}{\mathfrak{D}}$.

Let α be a 10δ-axis of u with $d(x, \alpha) \leq 50\delta$.

Claim 1. For distinct $v, w \in W$, $v\alpha, w\alpha$ are not parallel. Indeed, if they were parallel, then $w^{-1}v \in E(u)$. Moreover, we will argue $w^{-1}v \in F(u)$, which is a contradiction since w, v are in distinct $F(u)$-cosets. The reason for $w^{-1}v \in F(u)$ is that since $w^{-1}v \in S^2 \subset S^{M\mathfrak{D}}$, we have

$$L(w^{-1}v) \leq |w^{-1}v(x) - x| \leq L(S^{M\mathfrak{D}}).$$

But on the other hand, since u is primitive, if $w^{-1}v \notin F(u)$, then

$$L(w^{-1}v) \geq L(u) - 100\delta \geq 6L(S^{M\mathfrak{D}}).$$

The last inequality is from $L(u) \geq 10L(S^{M\mathfrak{D}})$. Those two estimates contradict. We showed the claim.

It implies that for distinct $v, w \in W$, the intersection of $v\alpha$ and the 50δ-neighborhood of $w\alpha$ is bounded by

$$2D'(100\delta)L(u) + 100\delta$$

by Lemma 2.5 (2) since u is D'-WPD. Remember that $D'(100\delta) = D(200\delta) = \mathfrak{D}$. So, this bound is $2\mathfrak{D}L(u) + 100\delta$.

Set $U = u^{20\mathfrak{D}}$. Then $U \in S^{20M\mathfrak{D}^2 m}$ and $L(U) \geq 19\mathfrak{D}L(u)$ (maybe not quite $20\mathfrak{D}L(u)$). α is an axis for U as well. Set

$$B = \{wUw^{-1} \mid w \in W\}.$$

Claim 2. We have $|B| = |W|$ and B freely generates a free group of rank $|W|$.

This is because for any $w \in W$, we have $|x - w(x)| \leq L(S^{M\mathfrak{D}})$ since $W \subset S^{M\mathfrak{D}}$, which means the axis of wUw^{-1}, $w\alpha$, is close to x. To be precise, close means that the distance is much smaller than $L(u)$ since $L(u) \geq 10L(S^{M\mathfrak{D}})$. Also, for the axes $w\alpha$ and $v\alpha$ of any distinct $v, w \in W$, the intersection of one with the 50δ-neighborhood of the other is nine times shorter than $L(U)$ since $L(U) \geq 19\mathfrak{D}L(u)$. In this setting, the usual ping-pong argument shows the claim.

Since $w \in S$, we have $B = \{wUw^{-1} \mid w \in W\} \subset S^{20M\mathfrak{D}^2 m+2}$. It follows that for any $n \in \mathbb{N}$,

$$|S^{(20M\mathfrak{D}^2 m+2)n}| \geq |B^n| \geq |B|^n = |W|^n \geq \frac{|S|^n}{\mathfrak{D}^n}.$$

It implies that

$$e(G, S) \geq \mathfrak{D}^{-\frac{1}{20M\mathfrak{D}^2 m+2}} |S|^{\frac{1}{20M\mathfrak{D}^2 m+2}}.$$

Since

$$\min\left\{\mathfrak{D}^{-\frac{1}{20M\mathfrak{D}^2 m+2}}, \frac{1}{20M\mathfrak{D}^2 m+2}\right\} = \frac{1}{20M\mathfrak{D}^2 m+2},$$

which is at least $\frac{1}{79220M\mathcal{D}^3}$, since $m = 3960\mathcal{D} + 4$. Setting

$$A = \frac{1}{79220M\mathcal{D}^3},$$

we have $e(G, S) \geq A|S|^A$. This is a desired conclusion since $\mathcal{D} = D(200\delta)$. ■

Example 2.11. Proposition 2.10 applies to the following examples:

(1) Non-elementary hyperbolic groups (the original case in [1]).

(2) The mapping class groups of a compact orientable surface $\Sigma_{g,p}$ with $3g + p \geq 4$. See Section 5.3, where the assumptions are checked.

(3) A lattice in a simple Lie group of rank 1 (see Theorem 5.5).

(4) The fundamental group of a complete Riemannian manifold of finite volume whose sectional curvature is pinched by two negative constants (see Theorem 5.5).

3. Well-orderedness

3.1. Main theorem

We prove the following theorem. Note that Theorem 1.1 immediately follows from this theorem combined with Lemma 2.3, since the lemma says that an acylindrical action is uniformly WPD, so that every hyperbolic element is WPD.

Theorem 3.1 (Well-orderedness for uniform WPD actions). *Suppose G acts on a δ-hyperbolic space X, and G is not virtually cyclic. Let $D(\varepsilon)$ be a function for WPD. Assume that there exists a constant M such that for any finite generating set S of G, the set S^M contains a hyperbolic element on X that is D-WPD. Assume that G is equationally Noetherian. Then, $\xi(G)$ is a well-ordered set.*

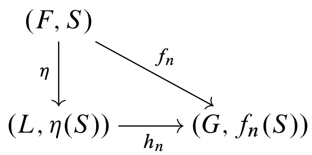
Proof. We will prove that $\xi(G)$ does not contain a strictly decreasing convergent sequence. To argue by contradiction, suppose that there exists a sequence of finite generating sets $\{S_n\}$, such that the sequence $\{e(G, S_n)\}$ is a strictly decreasing sequence and $\lim_{n\to\infty} e(G, S_n) = d$, for some $d > 1$.

By Proposition 2.10, we may assume that the cardinality of the generating sets $|S_n|$ from the decreasing sequence is bounded, and by possibly passing to a subsequence we may assume that the cardinality of the generating sets is fixed, $|S_n| = \ell$.

Let $S_n = \{x_1^{(n)}, \ldots, x_\ell^{(n)}\}$. Let F be the free group of rank ℓ with a free generating set: $S = \{s_1, \ldots, s_\ell\}$. For each index n, we define a map: $f_n : F \to G$, by setting: $f_n(s_i) = x_i^{(n)}$. Since S_n are generating sets, the map f_n is an epimorphism for every n. Note that $e(G, S_n) = e(G, f_n(S))$.

Since F is countable, the sequence $\{f_n : F \to G\}$ subconverges to a surjective homomorphism $\eta : F \to L$. The group L is called a *limit group* over the group G.

By assumption, G is equationally Noetherian. By the general principle (Lemma 1.11), there exists an epimorphism $h_n : L \to G$ such that by passing to a subsequence we may assume that all the homomorphisms $\{f_n\}$ factor through the limit epimorphism: $\eta : F \to L$, that is, $f_n = h_n \circ \eta$.



Notice that since $f_n = h_n \circ \eta$ for every index n, we have $e(G, f_n(S)) \leq e(L, \eta(S))$. We will show the following key result.

Proposition 3.2. *Suppose G satisfies the assumption in Theorem 3.1. Let $(L, \eta(S))$ be the limit group over G of a sequence $f_n : (F, S) \to (G, f_n(S))$, where F is a free group with a free generating set S and $f_n(S)$ are generating sets of G. Then* $\lim_{n\to\infty} e(G, f_n(S)) = e(L, \eta(S))$.

We postpone proving this proposition until the next section and finish the proof of the theorem.

We assumed that the sequence $\{e(G, f_n(S))\}$ is strictly decreasing; hence, it cannot converge to an upper bound of the sequence, $e(L, \eta(S))$. But the proposition says that it must converge to $e(L, \eta(S))$, a contradiction. Theorem 3.1 is proved. ∎

4. Continuity of the growth rate

We prove Proposition 3.2. We already know that $e(L, \eta(S)) \geq e(G, f_n(S))$ from the existence of the surjections h_n, which follows from that G is equationally Noetherian.

It suffices to show that given $\varepsilon > 0$, for a large enough n,

$$\log e(L, \eta(S)) - \varepsilon \leq \log e(G, f_n(S)).$$

The strategy of the proof of this is same as [13]. We note that from now on, we do not use that G is equationally Noetherian in the proof.

Since the proof is long and complicated, we first informally describe the idea, which already appeared in [13]. We want to show $e(G, f_n(S))$ is almost equal to $e(L, \eta(S))$ for a large enough n. First of all, if we take a large enough r, then $B_r(L, \eta(S))$ contains elements roughly as many as $e(L, \eta(S))^r$ by the definition. Fix such r. Then if we take n large enough, $B_r(L, \eta(S))$ and $B_r(G, f_n(S))$ are identical via the map h_n since L is a limit group. But it does not mean that $e(G, f_n(S))$ is almost equal to $e(L, \eta(S))$ since the growth of the balls in $(G, f_n(S))$ may decay if we take the radius larger. But it turns out that if we take r large enough, then roughly speaking, the growth of the ball of radius r

in $(G, f_n(S))$ is almost equal to $e(G, f_n(S))$. This is due to the well-known "local-to-global" principle in δ-hyperbolic spaces, and it is implemented by inserting "separators" in our argument. The threshold for the radius is given by m in the proof (see Section 4.5).

We explain the idea more in detail. By the definition of the growth rate, we have for all r,

$$\log e(L, \eta(S)) \leq \frac{1}{r} \log |B_r(L, \eta(S))|.$$

This is because the sequence $\{\log |B_r(L, \eta(S))|\}$ is sub-additive.

Fix r (we will choose r sufficiently large in the argument we will give later). Then choose n large enough such that $h_n : B_r(L, \eta(S)) \to B_r(G, f_n(S))$ is injective. The following map is naturally induced from h_n for each $q \in \mathbb{N}$:

$$B_r(L, \eta(S))^q \to B_{qr}(G, f_n(S)) \subset G$$

by mapping $(w_1, \ldots, w_q)$ to $h_n(w_1 \cdots w_q)$. If there is an r such that this map is injective for all q, then an easy computation would show the desired inequality for the n by letting $q \to \infty$.

But we cannot expect that this map is injective in general. For example, the concatenation of $h_n(w_1), h_n(w_2), \ldots, h_n(w_q)$ may have lots of backtracks at the concatenation points, and $h_n(w_1 \cdots w_q)$ is maybe the trivial element. As a remedy, we insert elements u_i of bounded length, called *separators*, and define a new map sending $(w_1, \ldots, w_q)$ to $h_n(w_1 u_1 w_2 u_2 \cdots w_q u_q)$. This map is denoted by Φ_n. The separators are constructed in Lemma 4.2. We arrange that the concatenation of elements, after we insert $h_n(u_i)$s, is a uniform quasi-geodesic in G, so that $h_n(w_1 u_1 w_2 u_2 \cdots w_q u_q)$ is at least not the trivial element. For this part we use the assumption that G acts on a hyperbolic group.

It is possible that, even after this modification, the map $\Phi_n : B_r(L, \eta(S))^q \to B_{q(r+b)}(G, f_n(S))$ is maybe not injective. Here, b is the bound of the length of the separators. What we can actually show is that Φ_n is injective if we restrict it to the q-tuples in some fixed portion of $B_r(L, \eta(S))$ (see Lemma 4.8), which is enough for our purpose. This part is very technical. To argue that Φ_n is injective on the certain fixed portion, we use the action of G on X, that is, we map a tuple by Φ_n to G, then let it act on X. Then we analyze the orbit of a base point in X.

4.1. Separators

We review the setting. X is a δ-hyperbolic space, and G acts on it. We assume $\delta \geq 1$. For each n, S_n is a finite generating set of G such that S_n^M contains a hyperbolic element that is D-WPD. Using the homomorphism $h_n : L \to G$, we let L act on X. We first construct separators as elements in G then pull them back to L by h_n.

Let $g \in S_n^M$ be a hyperbolic element that is D-WPD. Set

$$\mathcal{D} = D(200\delta).$$

Then $\lambda(g) \geq \frac{50\delta}{\mathcal{D}}$ by Lemma 2.2 (1). It implies that

$$100\delta \leq \lambda(g^{2\mathcal{D}}) \leq L(S_n^{2\mathcal{D}M}).$$

Fix n. Let $y_n \in X$ be a point where $L(S_n^{2\mathfrak{D}M})$ is achieved. We call y_n a base point. Put

$$\Delta_n = 100\delta + 4\mathfrak{D}L(S_n^{2\mathfrak{D}M}).$$

We define a germ with respect to the constant Δ_n. Recall that given three points $x, y, z \in X$, the *Gromov product*, $(y,z)_x$, is defined as follows:

$$(y,z)_x = \frac{|x-y| + |x-z| - |y-z|}{2}.$$

Definition 4.1 (Germs, equivalent and opposite germs). Let $[x,y]$ be a (directed) geodesic segment in X. Suppose that $|x-y| \geq 10\Delta_n$. Then, the initial segment of $[x,y]$ of length $10\Delta_n$ is called the *germ* of $[x,y]$ at x, denoted by $\mathrm{germ}([x,y])$. If $|x-y| < 10\Delta_n$, then we define the germ to be empty.

We say two non-empty germs, $[x,y]$, $[x,z]$, at a common point x, are *equivalent* if $(y,z)_x \geq 4\Delta_n$, and *opposite* if $(y,z)_x \leq 2\Delta_n$.

We sometimes call the germ of $[y,x]$ at y as the germ of $[x,y]$ at y. If γ is the germ of $[x,y]$ at x, then for $g \in G$, the segment $g(\gamma)$ is the germ of $[g(x), g(y)]$ at $g(x)$. For $g \in G$, we consider the germ of $[y_n, g(y_n)]$ and call it *the germ of g at y_n*, and write $\mathrm{germ}(g)$.

Recall from Lemma 2.8 that $k = 60\mathfrak{D}$, $m = 3964\mathfrak{D} + 4$.

We consider germs with respect to Δ_n.

Lemma 4.2 (The constant b and separators, cf. [13, Lemma 2.4]). *There exists a constant b with the following property, where b depends only on δ, M and $D(\varepsilon)$. For every n, there exist primitive, hyperbolic elements $u_1, u_2, u_3, u_4 \in S_n^b$, and mutually opposite germs $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ of some elements in $S_n^{2\mathfrak{D}M}$ at the base point y_n that satisfy:*

(i) The germs are all at y_n in the following.
- *The germ of $[y_n, u_1(y_n)]$ is equivalent to γ_1. The germ of $[y_n, u_1^{-1}(y_n)]$ is equivalent to γ_3.*
- *The germ of $[y_n, u_2(y_n)]$ is equivalent to γ_1. The germ of $[y_n, u_2^{-1}(y_n)]$ is equivalent to γ_4.*
- *The germ of $[y_n, u_3(y_n)]$ is equivalent to γ_2. The germ of $[y_n, u_3^{-1}(y_n)]$ is equivalent to γ_3.*
- *The germ of $[y_n, u_4(y_n)]$ is equivalent to γ_2. The germ of $[y_n, u_4^{-1}(y_n)]$ is equivalent to γ_4.*

(ii) *For all i,*

$$\lambda(u_i) \geq 100\Delta_n,$$

and the distance from y_n to the 10δ-axis of u_i is at most Δ_n.

(iii) *For every $w \in G$, and all i, j (possibly $i = j$), if the 20δ-neighborhood of $[y_n, u_i(y_n)]$ intersects the segment $[w(y_n), wu_j(y_n)]$, then the diameter of the* *intersection is bounded by:* $\frac{1}{10}d(y_n, u_i(y_n))$ *and* $\frac{1}{10}d(y_n, u_j(y_n))$*. If* $i = j$*, we assume in addition that* $w \notin F(u_i)$*, where* $F(u_i)$ *is the finite normal subgroup in* $E(u_i)$*.*

Remark 4.3. Some remarks are in order.

(1) Regarding Lemma 4.2 (iii), if $w \in F(u_i)$ then the Hausdorff distance between $[y_n, u_i(y_n)]$ and $[w(y_n), wu_i(y_n)]$ is at most $\Delta_n + 100\delta$ since w moves any point on a 10δ-axis of u_i by at most 50δ since otherwise w would be hyperbolic.

(2) As we will see in the proof, it suffices to take b to be

$$b = 343440M(D(200\delta))^2 + 22. \tag{4.1}$$

The elements u_i are called *separators* and the property (iii) is called the *small cancelation property* of separators.

Note that we have $L(S_n^{2M\mathfrak{D}}) \geq L(S_n)$.

Proof. Since $L(S_n^{2M\mathfrak{D}}) \geq 100\delta$ and y_n is a point where $L(S_n^{2M\mathfrak{D}})$ is achieved, by Lemma 2.7 (also see the properties at the bullets after the lemma), there is a hyperbolic element $g \in S_n^{4M\mathfrak{D}}$ such that

$$2L(S_n^{2M\mathfrak{D}}) \geq L(g) \geq L(S_n^{2M\mathfrak{D}}) - 24\delta \geq \frac{1}{2}L(S_n^{2M\mathfrak{D}})$$

and the 10δ-axis of g, γ, is at distance at most 10δ from y_n.

By Lemma 2.8 applied to $S_n^{2M\mathfrak{D}}$ and g and the 10δ-axis γ, there exists $s \in S_n$ (see Remark 2.9) such that g^k and sg^ks^{-1} are independent hyperbolic elements, which freely generate a free group F whose non-trivial element, h, satisfies

$$\lambda(h) \geq 10(2\mathfrak{D}L(g) + 100\delta) \geq 5\Delta_n.$$

Recall that $k = 60D(200\delta) = 60\mathfrak{D}$. Note that $g^k, sg^ks^{-1} \in S_n^{4M\mathfrak{D}k+2}$. The distance from y_n to $s\gamma$ is at most $40\delta + L(S_n^{2M\mathfrak{D}})$. Also, the intersection of γ and the 50δ-neighborhood of $s\gamma$ is at most $2\mathfrak{D}L(g) + 100\delta$ in length, which is $\leq \Delta_n$, by Lemma 2.5 since γ and $s\gamma$ are not parallel.

Consider the following four germs at y_n with respect to the constant Δ_n:

$$\gamma_1 = \text{germ}(g^k), \quad \gamma_2 = \text{germ}(sg^ks^{-1}), \quad \gamma_3 = \text{germ}(g^{-k}), \quad \gamma_4 = \text{germ}(sg^{-k}s^{-1}).$$

Note that any two of them are mutually opposite.

Then, there exist separators $u_i \in F$ such that $u_i \in S_n^b$, where

$$b = 1431 \cdot 4\mathfrak{D}Mk + 22 = 343440\mathfrak{D}^2M + 22.$$

For example, set $w = g^k, z = sg^ks^{-1}$, and

$$\begin{aligned}
u_1 &= wzw^2zw^3z \cdots w^{19}zw^{20},\\
u_2 &= w^{21}zw^{22}z \cdots w^{39}zw^{40}z,\\
u_3 &= zw^{41}zw^{42}zw^{43}z \cdots w^{59}zw^{60},\\
u_4 &= zw^{61}zw^{62}zw^{63}z \cdots w^{79}zw^{80}z.
\end{aligned}$$

We compute that they are in S_n^b since $w \in S_n^{4\mathfrak{D}Mk}, z \in S_n^{4\mathfrak{D}Mk+2}$. It will be important that this number b does not depend on S_n. (See the proof of Proposition 3.2.)

Because of the combinatorial reason, they are primitive hyperbolic elements. The argument is similar to the one we used to show that u is primitive hyperbolic, using γ and $s\gamma$ are not parallel, in the proof of Lemma 2.8 (2). We omit it.

(i) By definition of u_i,

$$\begin{aligned}
\mathrm{germ}[y_n, u_1(y_n)] = \mathrm{germ}[y_n, u_2(y_n)] &= \mathrm{germ}(w) = \gamma_1;\\
\mathrm{germ}[y_n, u_3(y_n)] = \mathrm{germ}[y_n, u_4(y_n)] &= \mathrm{germ}(z) = \gamma_2;\\
u_1^{-1}\,\mathrm{germ}[u_1(y_n), y_n] = u_3^{-1}\,\mathrm{germ}[u_3(y_n), y_n] &= \gamma_3;\\
u_2^{-1}\,\mathrm{germ}[u_2(y_n), y_n] = u_4^{-1}\,\mathrm{germ}[u_4(y_n), y_n] &= \gamma_4.
\end{aligned}$$

Here, we mean that the four germs in each line are same or equivalent to each other.

(ii) The estimate for $\lambda(u_i)$ is straightforward from the definitions of u_i. The claim on the axes is also shown similarly to the case of u in Lemma 2.8 (1), and we omit it. We remark that the distance estimate on the axes differs since it comes from the fact that some of the separators start or end in w, so that distance becomes larger.

(iii) This follows from the definition of u_i, namely, the combinatorial structure of the words. The argument is similar to show that u in Lemma 2.8 is primitive (Lemma 2.8 (2)), so we will be brief. Suppose $i \neq j$, and the intersection was longer. Then, because of the combinatorial structure of the words u_i and u_j, it follows the axes γ and $s\gamma$ would be parallel, by Lemma 2.5, which is impossible. We are done. Suppose $i = j$, and the intersection was longer. Then since $w \notin F(u_i)$, by the same reason, γ and $s\gamma$ would be parallel, a contradiction. So we are done in this case too. ∎

Remember $u_i \in G$ depend on the index n of S_n, so let us write them as $u_i(n)$. Now, since h_n is surjective, let $\hat{u}_i(n) \in L$ be an element with $h_n(\hat{u}_i(n)) = u_i(n)$. Note that the word length of $\hat{u}_i(n)$ in terms of $\eta(S)$ is also bounded by b. The elements $\hat{u}_i(n) \in L$ are also called the *separators* for h_n.

In the following we may just write u (instead of $\hat{u}$) to denote a separator for h_n to simplify the notation. We note that $|F(h_n(u))| \leq D(100\delta) \leq \mathfrak{D}$ for any separator u for any h_n since $h_n(u)$ is primitive (see the comment after Proposition 2.6).

4.2. Forbidden elements

Given m, choose and fix n large enough such that the map h_n is injective on $B_{2m}(L, \eta(S))$. The following discussion applies to all large enough n, but in each discussion, such n is fixed. This is possible since this set is finite in the limit group L. Remember

$$\Delta_n = 100\delta + 4\mathfrak{D}L(S_n^{2\mathfrak{D}M}).$$

Given $w \in B_m(L, \eta(S))$ we choose one of the separators, $u \in L$, for h_n such that the germ of $[y_n, h_n(w)(y_n)]$ at $h_n(w)(y_n)$ and the germ of $[h_n(w)(y_n), h_n(wu)(y_n)]$ at $h_n(w)(y_n)$ are opposite. Such a separator u exists by the property (i) in Lemma 4.2. Here, if $|y_n - h_n(w)(y_n)| < 10\Delta_n$, then we choose any separator u. We say u is *admissible* for w.

Geometrically, we have the concatenation $[y_n, h_n(w)(y_n)] \cup [h_n(w)(y_n), h_n(wu)(y_n)]$ is almost a geodesic. Note that

$$|y_n - h_n(wu)(y_n)| \geq |y_n - h_n(w)(y_n)| + |h_n(w)(y_n) - h_n(wu)(y_n)| - 4\Delta_n.$$

Furthermore, the Hausdorff distance between $[y_n, h_n(wu)(y_n)]$ and $[y_n, h_n(w)(y_n)] \cup [h_n(w)(y_n), h_n(wu)(y_n)]$ is at most $2\Delta_n + 10\delta$.

Moreover, given $w, w' \in B_m$ then we can choose a separator u such that u is admissible for w and u^{-1} is admissible for w'^{-1}. We say u is *admissible* for w, w'. Note that

$$\begin{aligned}|y_n - h_n(wuw')(y_n)| \geq |y_n - h_n(w)(y_n)| &+ |h_n(w)(y_n) - h_n(wu)(y_n)| \\ &+ |h_n(wu)(y_n) - h_n(wuw')(y_n)| - 8\Delta_n.\end{aligned}$$

Also, the Hausdorff distance between $[y_n, h_n(wuw')(y_n)]$ and $[y_n, h_n(w)(y_n)] \cup [h_n(w)(y_n), h_n(wu)(y_n)] \cup [h_n(wu)(y_n), h_n(wuw')(y_n)]$ is at most $2\Delta_n + 10\delta$.

For an integer $q > 0$, we define a map

$$\Phi_n : B_m(L, \eta(S))^q \to B_{q(m+b)}(G, f_n(S)) \subset G$$

by sending $(w_1, \ldots, w_q)$ to $h_n(w_1 u_1 \cdots w_q u_q)$, where $u_i \in L$ are separators we choose that are admissible for w_i, w_{i+1} for h_n. Remember that $u_i \in B_b(L, \eta(S))$.

We cannot expect that Φ_n is injective, even if we choose separators carefully. But we will argue that on a large portion, called the set of "feasible elements," Φ_n is injective by showing the image of the base point $y_n \in X$ by those elements are all distinct.

For the given m, we first define *forbidden* elements in $B_m(L, \eta(S))$. The definition depends on n. Recall that we chose and fixed a large enough n so that h_n is injective on $B_{2m}(L, \eta(S))$.

Definition 4.4 (Forbidden elements and tails). Given $w \in B_m(L, \eta(S))$, if there exist $w' \in B_m(L, \eta(S))$ and a separator u which is admissible for w such that

$$|h_n(w')(y_n) - h_n(wu)(y_n)| \leq \frac{1}{5}|y_n - h_n(u)(y_n)|, \tag{4.2}$$

then we say that w is *forbidden* with respect to n (or in terms of h_n). We call the segment $[h_n(w)(y_n), h_n(wu)(y_n)]$ a *tail* of w. The tail depends on the choice of u, so if we want to specify it, we say the tail of the pair (w, u). We sometimes use (w, u) to indicate the tail itself if there is no danger of confusion.

We remind that the separators u that appear in the above definition are for h_n.

A couple of remarks are in order. Let w_1, w_2 be two (forbidden) elements and u_1, u_2 admissible separators, respectively. If $u_1 = u_2$ and $h_n(w_1^{-1}w_2) \in F(h_n(u_1))$, then we say that the Hausdorff distance between the two tails $[h_n(w_1)(y_n), h_n(w_1u)(y_n)]$ and $[h_n(w_2)(y_n), h_n(w_2u)(y_n)]$ is at most $\Delta_n + 100\delta$. This is an immediate consequence of Remark 4.3. In this case, we say that the tails of (w_1, u_1) and (w_2, u_2) are *parallel*.

On the other hand, if two tails (w_1, u_1) and (w_2, u_2) are not parallel, then the intersection of one of the tails with the 20δ-neighborhood of the other tail is bounded by the $\frac{1}{10}$ of the length of each tail. This is by the small cancelation property of the separators (Lemma 4.2 (iii)),

We record one immediate consequence we use later.

Lemma 4.5 (Parallel tails). *Assume that h_n is injective on $B_{2m}(L, \eta(S))$. Suppose $w \in B_m(L, \eta(S))$ is forbidden with respect to n, and u is a separator admissible for w. Then there are at most $\mathcal{D}$ possibilities for $w_1 \in B_m$, including $w = w_1$, such that w_1 is forbidden, u is admissible for w_1, and the tails for (w, u) and (w_1, u) are parallel.*

Proof. Since the two tails are parallel, we have $h_n(w_1^{-1}w) \in F(h_n(u))$ by the definition that two tails are parallel. Recall that (see the paragraphs after Proposition 2.6)

$$|F(h_n(u))| \leq \mathcal{D}$$

for all separators u since u is primitive. Therefore, we have at most $\mathcal{D}$ possibilities for $h_n(w_1^{-1}w)$. But since h_n is injective on $B_{2m}(L, \eta(S))$, we have at most $\mathcal{D}$ possibilities for $w_1^{-1}w \in B_{2m}(L, \eta(S))$, and we are done. ∎

4.3. Ratio of forbidden elements

The proof of the following lemma occupies this subsection. If an element is not forbidden, then we say it is *non-forbidden*.

Lemma 4.6. *Assume that h_n is injective on $B_{2m}(L, \eta(S))$. Consider the forbidden/non-forbidden elements in $B_m(L, \eta(S))$ with respect to n. Then*

$$|\{\text{the forbidden elements}\}| \leq \mathcal{D}|\{\text{the non-forbidden elements}\}|.$$

The lemma says that the ratio of the non-forbidden elements in $B_m(L, \eta(S))$ is at least $1/(1 + \mathcal{D})$ for all m and large enough n, depending on m. Recall that $\mathcal{D}$ does not depend on n (nor m).

A similar estimate appears in [13, Lemma 2.7], but the proof is significantly different, and unfortunately more complicated in our case.

From now on, we denote $B_m(L, \eta(S))$ as B_m. The strategy of the proof is to show that if $w \in B_m$ is forbidden, it will force some other elements in B_m to be non-forbidden.

Proof. The proof has three parts.

Part I. We first construct a subset $C(w)$ in B_m and a tree-like graph $T(w)$ in X. The construction is inductive. In the step k, a subset $C_k(w)$ and a tree-like graph $T_k(w)$ are defined, which will end in finite steps, and we obtain $C(w)$ and $T(w)$ at the end.

Suppose $w \in B_m$ is forbidden. We explain the inductive steps to construct $C(w)$ and $T(w)$.

Step 0. Set $C_0(w) = w$ and $T_0(w) = [y_n, h_n(w)(y_n)]$.

Step 1. Since w is forbidden, there exist $w' \in B_m$ and a separator u that is admissible for w satisfying (4.2). Let $w' = s_1 \cdots s_r, s_i \in \eta(S), r \leq m$ be a shortest representative with respect to the word metric by $\eta(S)$. Then this defines a sequence of points in X, which we call a path from y_n to $h_n(w')(y_n)$ as follows:

$$y_n, h_n(s_1)(y_n), h_n(s_1 s_2)(y_n), \ldots, h_n(s_1 \cdots s_r)(y_n).$$

We denote this path by γ. The distance between any two adjacent points on γ is at most $L(S_n^{2MD})$ since y_n is moved by at most $L(S_n^{2MD})$ by any element in S_n^{2MD}, and $h_n(s_i) \in S_n \subset S_n^{2MD}$.

Consider the nearest point projection in X, denoted by π, from a point $x \in \gamma$ to the tail $[h_n(w)(y_n), h_n(wu)(y_n)]$. The nearest points may not be unique, but we choose one as $\pi(x)$. Then the distance between the projection of any two adjacent points on γ is at most $L(S_n^{2MD}) + 100\delta$, which is at most $\frac{\Delta_n}{4}$. Note that $|\pi(y_n) - h_n(w)(y_n)| \leq 2\Delta_n + 150\delta$ since u is admissible for w. Also, $|\pi\big(h_n(w')(y_n)\big) - h_n(wu)(y_n)| \leq \frac{1}{5}|h_n(w)(y_n) - h_n(wu)(y_n)| + 10\delta$ by (4.2).

By Lemma 4.2 (ii), the length of the tail is at least $100\Delta_n$. Let P, Q be the two points on the tail that trisect the tail into three pieces of equal length, where P is closer to $h_n(w)(y_n)$ than Q is. Each of the three pieces has length at least $33\Delta_n$. Let $h_n(s_1 \cdots s_p)(y_n)$ be a point on γ whose projection is closest to P, and $h_n(s_1 \cdots s_q)(y_n)$ a point whose projection is closest to Q. Then

$$|\pi(h_n(s_1 \cdots s_p)(y_n)) - P| \leq L(S_n^{2MD}) \leq \frac{\Delta_n}{4},$$
$$|\pi(h_n(s_1 \cdots s_q)(y_n)) - Q| \leq L(S_n^{2MD}) \leq \frac{\Delta_n}{4}.$$

We denote $s_1 \cdots s_p$, and $s_1 \cdots s_q \in B_m$ by w_0 and w_1, respectively, and call them the *candidates* for non-forbidden elements obtained from w. They depend on u, w' too (see Figure 3).

Consider the union of the three segments as follows:

$$\Pi(w) = [h_n(w)(y_n), h_n(wu)(y_n)] \cup [h_n(w_0)(y_n), \pi(h_n(w_0)(y_n))]$$
$$\cup [h_n(w_1)(y_n), \pi(h_n(w_1)(y_n))].$$

The tree $\Pi(w)$ is topologically embedded in X. This tree depends not only w, but also u, w'. We call $[h_n(w)(y_n), h_n(wu)(y_n)]$ the *tail* in $\Pi(w)$. We call $\pi(h_n(w_0)(y_n))$

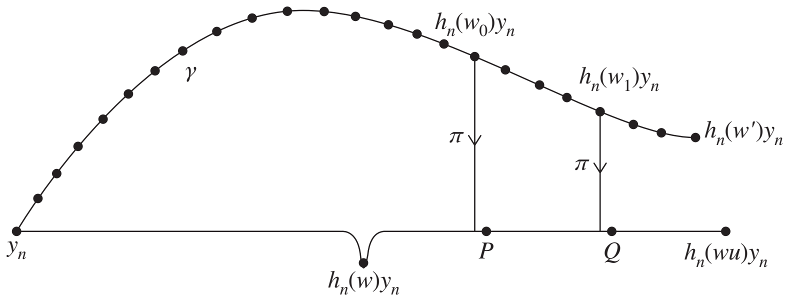


Figure 3. If w is forbidden, then we have two candidates w_0, w_1 for non-forbidden elements.

and $\pi(h_n(w_1)(y_n))$ the *branch points* of $\Pi(w)$, and P, Q the *trisecting points* of (the tail of) $\Pi(w)$. This union is a disjoint union except for the two branch points. The distance between the branch points is at least $32\Delta_n$. This in particular implies that $w_0 \neq w_1$.

We set $C_1(w) = \{w, w_0, w_1\} \subset B_m$, and $T_1(w) = [y_n, h_n(w)(y_n)] \cup \Pi(w)$. If both w_0, w_1 are non-forbidden, this is the end of the construction of $C(w)$ and $T(w)$, and put $C(w) = C_1(w), T(w) = T_1(w)$. Otherwise we go to the second step.

The graph $T_1(w)$ is a tree-like graph in the sense that it is the union of the segment $[y_n, h_n(w)(y_n)]$ and a tree $\Pi(w)$ attached at the point $h_n(w)(y_n)$. The intersection of the two sets is contained in the $(2\Delta_n + 100\delta)$-neighborhood of this point since u is admissible for w.

Step 2. By assumption, at least one of w_0 and w_1 is forbidden. We start with the case that w_1 is forbidden. Then w_1 has a separator u_1 admissible for w_1, and an element $w_1' \in B_m$ that satisfies (4.2). Let π_1 denote the nearest point projection to the tail $[h_n(w_1)(y_n), h_n(w_1u_1)(y_n)]$.

Then as in the first step, there is a path, which is a sequence of points obtained from a shortest expression of w_1' in $\eta(S)$, between y_n and $h_n(w_1')(y_n)$ in X. From this path, we obtain two elements, $w_{10}, w_{11} \in B_m$, using the projection π_1 from the path to the tail $[h_n(w_1)(y_n), h_n(w_1u_1)(y_n)]$. Also, we obtain a tree embedded in X as follows:

$$\begin{aligned}\Pi(w_1) = [h_n(w_1)(y_n), h_n(w_1u_1)(y_n)] &\cup [h_n(w_{10})(y_n), \pi(h_n(w_{10})(y_n))] \\ &\cup [h_n(w_{11})(y_n), \pi(h_n(w_{11})(y_n))].\end{aligned}$$

Let $T_1(w)$ and $\Pi(w_1)$ be connected at the point $h_n(w_1)(y_n)$. We note that the intersection of $T_1(w)$ and $N_{10\delta}(\Pi(w_1))$ is contained in a "small" ball centered at the point $h_n(w_1)(y_n)$. To be precise, the radius of the ball is at most either (i) $(2\Delta_n + 100\delta)$ or (ii) $(4\Delta_n + \frac{1}{10}$ of the minimum of the length of the tail in $\Pi(w)$ and the tail in $\Pi(w_1)$). We explain the reason. We consider two cases. The first one is that the two sets $N_{10\delta}(\Pi(w_1))$ and the tail of $\Pi(w)$, namely, $[h_n(w)(y_n), h_n(wu)(y_n)]$, do not intersect. (For example, if the length of $[h_n(w_1)(y_n), \pi(h_n(w_1)(y_n))]$ is at least $3\Delta_n$, then they do not intersect.) In this case, we have the bound given in (i) since u is admissible to w, and u_1 is admissible to w_1, while the tail of (w, u) and the tail of (w_1, u_1) are both at least $100\Delta_n$ long. On the other hand, let us assume that those two sets intersect. It implies that the 10δ-neighborhood of the tail $[h_n(w_1)(y_n), h_n(w_1u_1)(y_n)]$ and the tail $[h_n(w)(y_n), h_n(wu)(y_n)]$ intersect. But in this case, by the small cancelation property of the separators, we have the desired bound on the radius of the ball given in (ii) since the two tails cannot be parallel. This is because the two points $h_n(w_1)(y_n)$ and $h_n(w)(y_n)$ are sufficiently far (see Remark 4.3). We add $4\Delta_n$ to the bound to take into account the length of the segment $[h_n(w_1)(y_n), \pi(h_n(w_1)(y_n))]$, which is at most $3\Delta_n$ in this case.

Also, the branch points on the tail $[h_n(w_1)(y_n), h_n(w_1u_1)(y_n)]$, and the trisecting points P_1, Q_1 are out of the $10\Delta_n$-neighborhood of the tail $[h_n(w)(y_n), h_n(wu)(y_n)]$.

It follows that w, w_1, w_{10}, w_{11} are distinct elements in B_m. If w_0 is not forbidden, then we put $C_2(w) = C_1(w) \cup \{w_{10}, w_{11}\}$ and

$$T_2(w) = T_1(w) \cup \Pi(w_1).$$

If w_0 is forbidden, then we do the same construction as we did to w_1 and obtain two candidates: $w_{00}, w_{01} \in B_m$, and a tree $\Pi(w_0)$ which intersects the previously obtained tree-like graph $[y_n, h_n(w)(y_n)] \cup \Pi(w) \cup \Pi(w_1)$ only in a ball of small radius at $h_n(w_0)(y_n)$. To be precise, the radius is at most $(2\Delta_n + 100\delta)$, or $(4\Delta_n + \frac{1}{10}$ of the minimum of the length of the tail in $\Pi(w)$ and the tail in $\Pi(w_0))$. In particular, $w, w_0, w_{00}, w_{01}, w_1, w_{10}, w_{11}$ (if they exist) are all distinct (see Figure 4).

We set $C_2(w) = C_1(w) \cup \{w_{00}, w_{01}, w_{10}, w_{11}\}$ and

$$T_2(w) = T_1(w) \cup \Pi(w_0) \cup \Pi(w_1).$$

We are left with the case that w_1 is not forbidden and w_0 is forbidden. This case is treated in the same way as in the case that w_1 is forbidden and w_0 is not forbidden. We obtain $C_2(w) = C_1(w) \cup \{w_{00}, w_{01}\}$ and $T_2(w) = T_1(w) \cup \Pi(w_0)$.

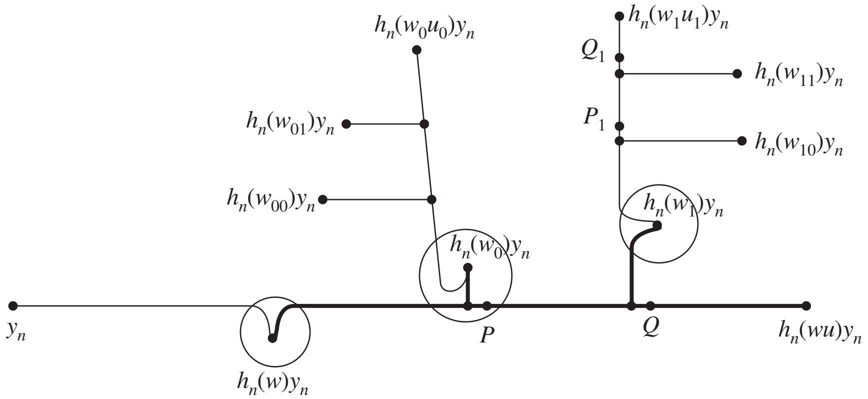


Figure 4. This is $T_2(w)$ in the case that both w_0, w_1 are forbidden. $\Pi(w)$ is in thick lines. $T_2(w)$ contains the segment $[y_n, h_n(w)(y_n)]$ and two other trees $\Pi(w_0), \Pi(w_1)$. They intersect each other only in balls of "small" radius at the points where they are connected. Here, small is compared to Δ_n and the length of the tails in $\Pi(w), \Pi(w_0), \Pi(w_1)$. Those balls are indicated in the picture.

The construction of $C_2(w)$ and $T_2(w)$ is finished. If all of the elements w_{00}, w_{01} and also $w_{10}, w_{11} \in B_m$ (some of them may not exist) are non-forbidden, we stop here and set $C(w) = C_2(w), T(w) = T_2(w)$; otherwise, we go to the third step.

Step 3. In the third step, we do the same construction to the forbidden elements among $w_{00}, w_{01}, w_{10}, w_{11} \in B_m$. For example, if w_{00} is forbidden, then we obtain two elements w_{000}, w_{001} and a tree $\Pi(w_{00})$. This tree intersects the tree-like graph $T_2(w)$ in the ball of small radius at $h_n(w_{00})(y_n)$. Here, small is in the same sense as in Step 2. We define $C_3(w)$ from $C_2(w)$ by adding a pair of elements that are obtained from each forbidden element w_{**} in this step. Also, we define $T_3(w)$ be adding the trees $\Pi(w_{**})$ for forbidden elements w_{**}.

Step-$(N+1)$. Up to Step-N, we have $C_N(w), T_N(w)$. If all of the elements in $C_N(w)\backslash C_{N-1}(w)$ are non-forbidden, then we stop and put $C(w) = C_N(w), T(w) = T_N(w)$. Otherwise, we go to the $(N+1)$-th step.

By assumption, at least one element in $C_N(w)\backslash C_{N-1}(w)$, $w_{i_1\cdots i_N}$, is forbidden. Then we do the same construction as in Step 1, and obtain the tree $\Pi(w_{i_1\cdots i_N})$ – two elements, $w_{i_1\cdots i_N 0}$ and $w_{i_1\cdots i_N 1}$. We add those two elements to $C_N(w)$, and also add $\Pi(w_{i_1\cdots i_N})$ to $T_N(w)$. We do this all forbidden elements in $C_N(w)\backslash C_{N-1}(w)$, and get $C_{N+1}(w)$ and $T_{N+1}(w)$. Note that the trees $\Pi(w_I)$ we added in this step are disjoint from each other; moreover, the distance between two $\Pi(w_I), \Pi(w_{I'})$ with $I \neq I'$ is at least $30\Delta_n$, where I, I' are multi-subscripts of length N.

Since $C_n(w) \subset B_m$ for all n, the set $C_n(w)$ gets bigger at least by 2 as n increases by 1, so the process must end in finite steps since B_m is finite. At the end of the process, we have a set $C(w)$ in B_m, and a tree-like graph $T(w)$. We remark that $C(w), T(w)$ depend on the choice of separators in the construction, but it is not important for our purpose.

Note that by construction, in $C(w)$, we have

$$|\{\text{forbidden elements}\}| + 1 = |\{\text{non-forbidden elements}\}|.$$

Also, the orbit of y_n by the elements in $C(w)$ are vertices of $T(w)$.

Part II. Let $w_1, w_2 \in B_m$ be two forbidden elements, and we analyze how the sets $C(w_1)$ and $C(w_2)$ are related. In the construction of $C(w)$, for each forbidden element, v, we chose a separator u and an element $v' \in B_m$. The pair (v, u) defines the tail $[h_n(v)(y_n), h_n(vu)(y_n)]$, which is an arc of $T(w)$.

Claim. Let $w_1, w_2 \in B_m$ be forbidden elements and u_2 the separator we chose for w_2 to construct $T(w_2), C(w_2)$. Assume that the tail, τ, of the pair (w_2, u_2) is not parallel to any of the tail of a pair (v, u) that appears to construct $T(w_1), C(w_1)$. Then, $C(w_1) \cap C(w_2)$ is empty or $\{w_2\}$.

Let $P, Q \in \tau$ be the trisecting points of $\Pi(w_2)$. Then neither of them is in the $10\Delta_n$-neighborhood of any trisecting point that appears in $T(w_1)$. Indeed, suppose it was, and let R be a trisecting point of the tail, σ, in $T(w_1)$ such that $|P - R|$ or $|Q - R|$ is at most $10\Delta_n$. Then, since $T(w_1), T(w_2)$ have y_n as the common "root" vertex that they start from, the intersection of σ and the 20δ-neighborhood of τ is at least $\frac{1}{5}$ of σ, which contradicts the small cancelation property.

This implies that not only the branch points on τ, but also all branch points of $T(w_2)$ are outside of the $9\Delta_n$-neighborhood of $T(w_1)$. This is because both $T(w_1), T(w_2)$ are tree-like graphs. It then follows that $C(w_1) \cap C(w_2)$ is empty or just w_2 because all other points in $C(w_2)$ appear on $T(w_2)$ after the first two branch points on τ, that are close to P, Q. We showed the claim.

Part III. It follows from the claim in Part II that there is a finite collection of forbidden elements w in B_m such that the $C(w)$s are mutually disjoint, and that any forbidden element $v \in B_m$ is either contained in the union of those $C(w)$s, or v has an admissible separator u such that the tail of the pair (v, u) is parallel to one of the tails that appears in one of the $T(w)$s.

To see that, order the forbidden elements in B_m as $w_1, w_2, \ldots$ Choose admissible separators $u_1, u_2, \ldots$ First, construct $C(w_1)$ using the separators we chose. Next, if w_2 is already contained in $C(w_1)$, then disregard it. Also, if (w_2, u_2) is parallel to one of the tails in $T(w_1)$, then disregard w_2 also. Otherwise, construct $C(w_2)$ and keep it. If w_3 is contained in $C(w_1)$ or $C(w_2)$, or (w_3, u_3) is parallel to one of the tails of $T(w_1)$ or $T(w_2)$, then disregard w_3. Otherwise, construct $C(w_3)$ and keep it, and so on. By the claim we have shown that those $C(w)$s are mutually disjoint such that for any forbidden element $v \in B_m$, either v is contained in one of the $C(w)$s we have, or (v, u) is parallel to one of the tails of the $T(w)$s.

To finish the proof, by Lemma 4.5, the union of those $C(w)$s contains at least $\frac{1}{\mathcal{D}}$ of the forbidden elements in B_m. But since in each $C(w)$ there are more non-forbidden elements than forbidden elements (by 1), we conclude that in B_m:

$$|\{\text{forbidden elements}\}| \leq \mathcal{D}|\{\text{non-forbidden elements}\}|.$$

Lemma 4.6 is proved. ∎

4.4. Feasible elements

By Lemma 4.6, at least $\frac{1}{\mathcal{D}+1}$ of $B_m(L, \eta(S))$ consists of non-forbidden elements.

We choose a maximal subset in the set of non-forbidden elements in $B_m(L, \eta(S))$ such that for any two distinct elements w, w' in the set, $h_n(w), h_n(w')$ are not in the same coset with respect to $F(h_n(u_i))$ for any separator u_i. We call an element in this set an *adequate* element. Then, in $B_m(L, \eta(S))$,

$$\frac{|\{\text{non-forbidden elements}\}|}{\mathcal{D}^4} \leq |\{\text{adequate elements}\}|.$$

This is because there are only four separators, u, and $|F(h_n(u))| \leq \mathcal{D}$ for each u. Combining this with Lemma 4.6, we get

$$\frac{|B_m(L, \eta(S))|}{\mathcal{D}^4(\mathcal{D}+1)} \leq |\{\text{adequate elements in } B_m(L, \eta(S))\}|. \tag{4.3}$$

Definition 4.7 (Feasible elements). Fix m, then n. An element of the form $w_1 u_1 \cdots w_q u_q$ with $w_i \in B_m(L, \eta(S))$ are called *feasible* of *type* q if all w_i are adequate and each u_i is admissible for w_i, w_{i+1}. We define $q = 0$ for the empty element.

For this feasible elements, we consider the following broken geodesic:

$$\alpha = [y_n, h_n(w_1)(y_n)] \cup [h_n(w_1)(y_n), h_n(w_1 u_1)(y_n)] \cup \cdots \\ \cup [h_n(w_1 u_1 \cdots w_q)(y_n), h_n(w_1 u_1 \cdots w_q u_q)(y_n)].$$

The Hausdorff distance between α and the geodesic $[y_n, h_n(w_1 u_1 \cdots w_q u_q)(y_n)]$ is at most $2\Delta_n + 100\delta$.

Lemma 4.8 (cf. [13, Lemma 2.6]). *For every q, the map Φ_n is injective on the set of q-tuples of adequate elements in $B_m(L, \eta(S))$.*

We recall the definition of Φ_n from Section 4.2. Given a q-tuples of adequate elements, we choose admissible separators (they are not unique) and form a feasible element of type q using them, then Φ_n maps it to an element in $B_{q(m+b)}(G, f_n(S)) \subset G$.

Proof. We argue by induction on the type q. If $q = 0$ then nothing to prove since the element is empty.

Assume the conclusion holds for $q \geq 0$. Suppose $(q+1)$-tuples of adequate elements $w_1, w_2, \ldots, w_{q+1}$ and $w'_1, w'_2, \ldots, w'_{q+1}$ are given. Let $W = w_1 u_1 w_2 u_2 \cdots w_{q+1} u_{q+1}$ and $W' = w'_1 u'_1 w'_2 u'_2 \cdots w'_{q+1} u'_{q+1}$ be two feasible elements of type $q+1$. We assume that $h_n(W) = h_n(W')$ and want to show $w_i = w'_i$ for all i.

Between y_n and $h_n(W)(y_n) = h_n(W')(y_n)$, we state that the elements W, W' define two broken geodesics α, α'. Look at the initial parts of α, α': $[y_n, h_n(w_1)(y_n)] \cup [h_n(w_1)(y_n), h_n(w_1 u_1)(y_n)]$ and $[y_n, h_n(w'_1)(y_n)] \cup [h_n(w'_1)(y_n), h_n(w'_1 u'_1)(y_n)]$. We first show that $u_1 = u'_1$. Suppose not. Then it would imply that either w_1 or w'_1 is forbidden. We explain the reason. First, by the small cancelation property of the separators (Lemma 4.2 (iii)), either $[y_n, h_n(w_1)(y_n)] \cup [h_n(w_1)(y_n), h_n(w_1 u_1)(y_n)]$ is contained in the $10\Delta_n$-neighborhood of $[y_n, h_n(w'_1)(y_n)]$, or $[y_n, h_n(w'_1)(y_n)] \cup [h_n(w'_1)(y_n), h_n(w'_1 u'_1)(y_n)]$ is contained in the $10\Delta_n$-neighborhood of $[y_n, h_n(w_1)(y_n)]$, since u_1 is admissible for w_1 and u'_1 is admissible for w'_1. Suppose we are in the first case. Then w_1 would be forbidden since there is a subword of w'_1, denoted by $\overline{w'_1}$, such that

$$|h_n(\overline{w'_1})(y_n) - h_n(w_1 u_1)(y_n)| \leq \frac{1}{5} |h_n(w_1)(y_n) - h_n(w_1 u_1)(y_n)|.$$

Such a subword exists because $L(S_n) \leq \frac{\Delta_n}{4}$ and $|h_n(w_1)(y_n) - h_n(w_1 u_1)(y_n)| \geq 100\Delta_n$. We got a contradiction since w_1 is not forbidden. Similarly, if we are in the second case, then w'_1 would be forbidden, which is a contradiction. We showed that $u_1 = u'_1$.

Next we show $w_1 = w'_1$. We first show that $h_n(w_1 w'^{-1}_1) \in F(h_n(u_1))$. Suppose not. Then, as in the previous paragraph, the small cancelation property of the separators implies that either w_1 or w_1' is forbidden, a contradiction. We are left with the case that $h_n(w_1 w_1'^{-1}) \in F(h_n(u_1))$. But this means that $h_n(w_1)$ and $h_n(w_1')$ are in the same (right) coset with respect to $F(h_n(u_1))$. Since both w_1 and w_1' are adequate, it implies $w_1 = w_1'$.

Since $u_1 = u_1'$, $w_1 = w_1'$, it follows that $W_1 = w_2 u_2 \cdots w_{q+1} u_{q+1}$ and $W_1' = w_2' u_2' \cdots w_{q+1}' u_{q+1}'$ are feasible elements of type q with $h_n(W_1) = h_n(W_1')$. By the induction hypothesis, we have $w_i = w_i'$ for all $i \geq 2$, and we are done. ∎

4.5. Proof of Proposition 3.2

We prove Proposition 3.2.

Proof. Recall that D is the WPD-function and we set $\mathcal{D} = D(100\delta)$. Recall that for every m, we have

$$\frac{1}{m} \log(|B_m(L, \eta(S))| \geq \log e(L, \eta(S)).$$

Given $\varepsilon > 0$, choose and fix a large enough m such that

$$\frac{1}{m+b}(\log |B_m(L, \eta(S))| - \log(D^4(D+1)) \geq \frac{1}{m} \log |B_m(L, \eta(S))| - \varepsilon.$$

The constant b is from Lemma 4.2. Such m exists since $\lim_{m\to\infty} \frac{\log |B_m(L,\eta(S))|}{m} = \log e(L, \eta(S))$, and b does not depend on n, m. (We will choose n right after this.)

Choose n large enough such that h_n is injective on $B_{2m}(L, \eta(S))$, which defines the forbidden elements in $B_m(L, \eta(S))$. Then for all q:

$$|B_{q(m+b)}(G, f_n(S))| \geq \left(\frac{|B_m(L, \eta(S))|}{\mathcal{D}^4(\mathcal{D}+1)} \right)^q$$

because the number of adequate elements in $B_m(L, \eta(S))$ is at least $\frac{|B_m(L,\eta(S))|}{\mathcal{D}^4(\mathcal{D}+1)}$ by the estimate (4.3), and Φ_n is injective on the set of feasible elements of type q by Lemma 4.8.

Then by the above three inequalities,

$$\begin{aligned}
\log(e(G, f_n(S)) &= \lim_{q\to\infty} \frac{1}{q(m+b)} \log |B_{q(m+b)}(G, f_n(S))| \\
&\geq \frac{1}{m+b} \log \left(\frac{|B_m(L, \eta(S))|}{\mathcal{D}^4(\mathcal{D}+1)} \right) \geq \frac{1}{m} \log(|B_m(L, \eta(S))| - \varepsilon \\
&\geq \log e(L, \eta(S)) - \varepsilon.
\end{aligned}$$

Since we have this for all large enough n, and $\varepsilon > 0$ is arbitrary, we have

$$\lim_{n\to\infty} \log(e(G, f_n(S)) \geq \log e(L, \eta(S)).$$

As we said the other direction is trivial, hence the equality holds. ∎

4.6. The version for a family of groups

We state a variation of Proposition 3.2 for a family of groups. We do not use this proposition in this paper, but it would be useful in the future.

We explain the setting. Let $\delta > 0$ be a constant, $M > 0$ an integer, and $D(\varepsilon)$ a function for WPD. Suppose we have a family of finitely generated groups $\{G_\alpha\}$ such that each G_α is not virtually cyclic and acts on a δ-hyperbolic space X_α. Assume that for any finite generating set S of any G_α, the set S^M contains a hyperbolic element on X_α that is D-WPD.

Now suppose that S_n is a finite generating set of G_n that is in the family for $n > 0$. Assume that the (infinite) sequence $e(G_n, S_n)$ is bounded from above. Then by Proposition 2.10, there is a constant $A > 0$ that depends only on M, δ and the function D such that $e(G_n, S_n) \geq A|S_n|^A$; therefore, the sequence $|S_n|$ is bounded from above.

As before, passing to a subsequence, we may assume that there exists $\ell > 0$ with $|S_n| = \ell$ for all n. This gives a sequence of surjections $f_n : (F, S) \to (G_n, f_n(S))$ with $f_n(S) = S_n$, where (F, S) is a pair of free group F and a free generating set S with $|S| = \ell$.

Since F is countable, passing to a further subsequence if necessary, we may assume that the sequence $\{f_n\}$ converges to a limit epimorphism $\eta : (F, S) \to (L, \eta)$.

Then we have the following.

Proposition 4.9 (cf. Proposition 3.2). *Assume that there exists a sequence of epimorphisms $\{h_n : L \to G_n\}$ such that for all n, the homomorphism f_n and the limit epimorphism: $\eta : F \to L$ satisfy $f_n = h_n \circ \eta$:*

$$\begin{array}{ccc} (F, S) & & \\ \downarrow{\scriptstyle \eta} & \searrow^{f_n} & \\ (L, \eta(S)) & \xrightarrow[h_n]{} & (G_n, f_n(S)) \end{array}$$

Then, we have

$$\lim_{n\to\infty} e(G_n, f_n(S)) = e(L, \eta(S)).$$

We omit a proof of this proposition since it is identical to the argument for Proposition 3.2. Instead, we make some comments comparing the two settings. First, regarding the assumption, in Proposition 3.2, the existence of h_n for a large enough n is a consequence of the assumption that G is equationally Noetherian. In the current setting, the existence of h_n is an assumption and we avoid an assumption related to equational Noetherianity of G_ns.

Second, by the existence of h_n, we immediately have $\lim_n e(G_n, f_n(S)) \leq e(L, \eta(S))$ in this setting, and $\lim_n e(G, f_n(S)) \leq e(L, \eta(S))$ in the previous setting for large enough n. The main issue was to show the other inequality $\lim_n e(G, f_n(S)) \geq e(L, \eta(S))$. It was done by constructing separators for each action of L on X, via $h_n : L \to G$. That argument applies without change to the current setting using $h_n : L \to G_n$, since the constants δ, M and the function $D(\varepsilon)$ are common for all (G_n, S_n) and X_n.

5. Examples

An obvious example for Theorem 1.1 is a non-elementary hyperbolic group, G. Let X be a Cayley graph of G, then it is δ-hyperbolic, and the action by G is (uniformly) proper, so that acylindrical, and non-elementary. The existence of the constant M is known [18]. As we said G is equationally Noetherian [26, 28]. Therefore, $\xi(S)$ is well ordered. This is proved in [13], and we adapted their argument to prove Theorem 3.1 in this paper. In this section, we discuss some other examples.

5.1. Relatively hyperbolic groups

We treat a relatively hyperbolic group. For example, see [5] for the definition and basic properties. Suppose G is hyperbolic relative to a collection of finitely generated subgroups $\{P_1, \dots, P_n\}$. Suppose G is not virtually cyclic and not equal to any P_i. Then it acts properly discontinuously on a proper δ-hyperbolic space X such that [5, Proposition 6.13]:

(1) There is a G-invariant collection of points, $\Pi \subset \partial X$, with Π/G finite. For each i, there is a point $p_i \in \Pi$ such that the stabilizer of p_i is P_i. The union of the G-orbits of the p_is is Π.

(2) For every $r > 0$, there is a G-invariant collection of horoballs $B(p)$ at each $p \in \Pi$ such that they are r-separated, that is, $d(B(p), B(q)) \geq r$ for every distinct $p, q \in \Pi$.

(3) The action of G on $X \backslash \big(\bigcup_{p\in\Pi} \operatorname{int} B(p)\big)$ is co-compact.

A subgroup $H < G$ is called *parabolic* if it is infinite, fixes a point in ∂X, and contains no hyperbolic elements. The fixed point is unique and called a *parabolic point*. In fact, Π is the set of parabolic points [5, Propositions 6.1 and 6.13].

For this action, we have the following.

Lemma 5.1 ([31, Proposition 5.1]). *Let G and X be as above. Then there exists M such that for any finite generating set S of G, the set S^M contains a hyperbolic element on X.*

Also we have the following.

Lemma 5.2. *Let G and X be as above. Then the action is uniformly WPD.*

Proof. Given $\varepsilon > 0$, take a G-invariant collection of horoballs that are $(10\varepsilon + 100\delta)$-separated in X. Let $\mathcal{B}$ denote the union of the interior of the horoballs in the collection. Then since the action of G on $X \backslash \mathcal{B}$ is co-compact, there exists a constant $D(\varepsilon)$ such that for any $y \in X \backslash \mathcal{B}$, the cardinality of the following set is at most $D(\varepsilon)$:

$$\{h \in G \mid |y - h(y)| \leq \varepsilon + 10\delta\}.$$

We argue that the action is uniformly WPD with respect to $D = D(\varepsilon)$. Let $g \in G$ be hyperbolic with an 10δ-axis γ. Let $x, y \in \gamma$ with $|x - y| \geq \lambda(g)$. It suffices to show that the following set contains at most $D(\varepsilon)$ elements:

$$\{h \in G \mid |x - h(x)| \leq \varepsilon, |y - h(y)| \leq \varepsilon\}.$$

We divide the case into four:

(1) $x \notin \mathcal{B}$. Then there are at most $D(\varepsilon)$ elements $h \in G$ such that $|x - h(x)| \leq \varepsilon$, and we are done.

(2) $y \notin \mathcal{B}$. This is same as (1).

(3) $x, y \in \mathcal{B}$ such that $x \in B(p)$ and $y \in B(q)$ with $p \neq q$. Then, there must be $z \in [x, y]$ with $z \notin N_{5\varepsilon+30\delta}(\mathcal{B})$ since the horoballs in $\mathcal{B}$ are $(10\varepsilon + 100\delta)$-separated. For this point z, we have $|z - h(z)| \leq \varepsilon + 10\delta$ for all h in the above. But there are at most $D(\varepsilon)$ such elements.

(4) $x, y \in B(p)$ for some p. Then $g(x) \in B(q)$ for some $q \neq p$, since g is hyperbolic and does not preserve any horoball. Since horoballs are $(10\varepsilon + 100\delta)$-separated, we have $|x - g(x)| \geq 10\varepsilon + 100\delta$. So, $\lambda(g) \geq 10\varepsilon + 50\delta$. Now, there are two possibilities depending on x and y: one is that $g(x) \in \gamma$ is between x and y. Then, as in (3), there must be $z \in [x, y]$ with $z \notin N_{5\varepsilon+30\delta}(\mathcal{B})$, say, $z \in [x, g(x)]$. Again, as in (3), for this point z, we have $|z - h(z)| \leq \varepsilon + 10\delta$, and we are done. The other possibility is that y is between x and $g(x)$ on γ. But in this case, since $|x - y| \geq \lambda(g)$, we have $|y - g(x)| \leq 50\delta$. Then the distance between $B(p)$ and $B(q)$ is at most 50δ since $y \in B(p)$ and $g(x) \in B(q)$. But it contradicts the separation of horoballs, so this case does not happen. ∎

We quote a theorem [15, Theorem D].

Theorem 5.3 (Equationally Noetherian [15]). *If G is hyperbolic relative to equationally Noetherian subgroups, then G is equationally Noetherian.*

We are ready to state a theorem.

Theorem 5.4 (Well-orderedness for relatively hyperbolic groups). *Let G be a group that is hyperbolic relative to a collection of subgroups $\{P_1, \dots, P_n\}$. Suppose G is not virtually cyclic, and not equal to P_i for any i. Suppose each P_i is finitely generated and equationally Noetherian. Then $\xi(G)$ is well ordered.*

Proof. By Theorem 5.3, G is equationally Noetherian. Take a hyperbolic space X with a G action as above. The action is non-elementary since G contains a hyperbolic isometry and is not virtually $\mathbb{Z}$. Then Theorem 3.1 applies by Lemmas 5.1 and 5.2. ∎

5.2. Rank 1 lattices

There are many examples of relatively hyperbolic groups, but we mention one standard family that Theorem 5.4 applies to.

Let G be a lattice in a simple Lie group of rank 1. It is always finitely generated and has exponential growth, and in fact uniform exponential growth [10]. If it is a uniform lattice, then it is hyperbolic, so that $\xi(G)$ is well ordered. We prove the following as an immediate application of Theorem 5.4.

Theorem 5.5 (Rank 1 lattices). *Let G be one of the following groups:*

(1) *A lattice in a simple Lie group of rank* 1.

(2) *The fundamental group of a complete Riemannian manifold M of finite volume such that there exist $a, b > 0$ with $-b^2 \leq K \leq -a^2 < 0$, where K denotes the sectional curvature.*

Then $\xi(G)$ is well ordered.

Proof. We only need to argue for non-uniform lattices since otherwise, G is a non-elementary hyperbolic group and the conclusion holds. Suppose that G is a non-uniform lattice. Then, it is known that G is relatively hyperbolic with respect to the parabolic subgroups, called *peripheral subgroups* $\{H_i\}$ that are associated with the cusps [5, 11]. Moreover, G is not virtually cyclic, and not equal to any H_i. Also, those H_i are finitely generated virtually nilpotent groups (see, e.g., [11]). It is known that finitely generated virtually nilpotent groups are equationally Noetherian [7], so that H_i are equationally Noetherian. With those facts, Theorem 5.4 applies to G and the conclusion holds. ∎

We remark that it is known that if G is linear over a field, then it is equationally Noetherian [2]. For lattices in a simple Lie group, one can apply this result as well to see that G is equationally Noetherian. (Consider the adjoint representation on its Lie algebra. It is faithful since the Lie group is simple.)

5.3. Mapping class groups

We discuss mapping class groups as a possible application. Let $MCG(\Sigma)$ be the mapping class group of a compact oriented surface $\Sigma = \Sigma_{g,p}$ of genus g, with p punctures, where *complexity* $c(\Sigma) = 3g + p$. It is known that it is either virtually abelian or has exponential growth, and then uniform exponential growth [19]. Let $\mathcal{C}(\Sigma)$ be the *curve graph* of Σ. The graph $\mathcal{C}(\Sigma)$ has vertex set representing the non-trivial homotopy classes of simple closed curves on Σ, and edges joining vertices representing the homotopy classes of disjoint curves. The group $MCG(\Sigma)$ naturally acts on it by isometries.

We assume $c(\Sigma) > 4$. Then $MCG(\Sigma)$ has exponential growth. We recall some facts:

(1) The graph $\mathcal{C}(\Sigma)$ is δ-hyperbolic, and any pseudo-Anosov element in $MCG(\Sigma)$ acts hyperbolically on $\mathcal{C}(\Sigma)$ [21].

(2) The action of $MCG(\Sigma)$ on $\mathcal{C}(\Sigma)$ is acylindrical [4]. It is non-elementary.

(3) There exists $T(\Sigma) > 0$ such that for any pseudo-Anosov element g, $\lambda(g) \geq T(\Sigma) > 0$ for the action on $\mathcal{C}(\Sigma)$ [21]. (This follows from the acylindricity as we pointed out.)

(4) There exists $M(\Sigma)$ such that for any finite $S \subset MCG$ such that $\langle S \rangle$ contains a pseudo-Anosov element, then S^M contains a pseudo-Anosov element [20].

In summary, the action of $MCG(\Sigma)$ on $\mathcal{C}(\Sigma)$ satisfies all the assumptions of Theorem 1.1 except we do not know if $MCG(\Sigma)$ is equationally Noetherian or not. We expect it to hold (e.g., see an announcement by Daniel Groves in [15]), but it does not exist in the literature yet. Once that is verified, it would imply that $\xi(MCG(\Sigma))$ is well ordered if $c(\Sigma) > 4$.

We remark that for $\Sigma_{1,1}$, $\Sigma_{1,0}$, $\Sigma_{0,4}$, the conclusion holds by [13] since $MCG(\Sigma)$ is hyperbolic (a well-known fact, e.g., [21]).

Question 5.6. Let $G = MCG(\Sigma)$. Is $\xi(G)$ well ordered? If so, the infimum is attained. It would be interesting to know its value and generating sets that attain the minimum for each Σ.

5.4. Three-manifold groups

We discuss 3-manifold groups. Let M be a closed, orientable 3-manifold. The manifold M is called *irreducible* if $\pi_1(M)$ does not admit a non-trivial splitting over the trivial group.

Let M be a closed, orientable, irreducible 3-manifold which is not a torus bundle over a circle. Then there is a finite collection of embedded disjoint essential tori T_i in M such that each connected component of $M \setminus \bigcup_i T_i$ is geometric, that is, either Seifert fibered, or admitting hyperbolic or Sol-geometry. Such a collection of smallest number of tori is called the *JSJ-decomposition* of M. A JSJ-decomposition in this sense exists by the solution of the geometrization conjecture of 3-manifolds. The collection of tori is maybe empty. Otherwise, we say M has non-trivial JSJ-decomposition. A non-trivial JSJ-decomposition gives a graph of groups decomposition of $\pi_1(M)$ along subgroups isomorphic to $\mathbb{Z}^2$, and $\pi_1(M)$ has exponential growth. Its Bass–Serre tree is called the *JSJ-tree*, T_M. An action of a group G on a tree T is called *k-acylindrical* if for every non-trivial element g, the subtree of fixed points by g is either empty or of diameter at most k. It is proved that the action of $\pi_1(M)$ on T_M is 4-acylindrical [30, Proposition 4.2]. It implies that the action is uniformly 4-WPD. It is known that if an action of G on a tree T is k-acylindrical for some k, then it is acylindrical [22, Lemma 5.2]; therefore, it is uniformly WPD (Lemma 2.3). Moreover, the acylindricity constants $R(\varepsilon)$, $N(\varepsilon)$ and the uniformly WPD constant $D(\varepsilon)$ depend only on k and ε. The moreover part easily follows from the proof of [22, Lemma 5.2].

Theorem 5.7. *Let M be a closed orientable 3-manifold, and $G = \pi_1(M)$. If M is one of the following, then G has exponential growth and $\xi(G)$ is well ordered:*

(1) *M is not irreducible such that G is not isomorphic to $\mathbb{Z}_2 * \mathbb{Z}_2$.*

(2) *M is irreducible such that it is not a torus bundle over a circle, and that it has a non-trivial JSJ-decomposition.*

(3) *M admits hyperbolic geometry.*

(4) *M is Seifert fibered such that the base 2-orbifold is hyperbolic.*

Proof. First, every 3-manifold group is equationally Noetherian [16]:

(1) In this case, G is a non-trivial free product $A * B$. Since it is not $\mathbb{Z}_2 * \mathbb{Z}_2$, it has exponential growth. Let G acts on the Bass–Serre tree T of this free product. Then for any finite generating set S, the set S^2 contains a hyperbolic element on T [27]. The action of G on T is 0-acylindrical, so that the action is uniformly D-WPD for some D. Since T is hyperbolic, Theorem 3.1 applies with $M = 2$.

(2) In this case, let T_M be the JSJ-tree of M. Then the action of G on T_M is 4-acylindrical, so that it is uniformly D-WPD for some D. Theorem 3.1 applies with $M = 2$.

(3) If M is hyperbolic, then G is a non-elementary, hyperbolic group. Then $\xi(G)$ is well ordered by [13].

(4) In this case, we have the following exact sequence:

$$1 \to \mathbb{Z} \to G \to H \to 1,$$

where H is the orbifold-fundamental group of the base 2-orbifold, and $\mathbb{Z}$ is the fundamental group of the fiber circle. We denote this subgroup C. By assumption, H is a non-elementary, hyperbolic group.

We claim that $\xi(G) = \xi(H)$. To see it, let S be a finite generating set of G. Let $\overline{S}$ be the image of S by the projection $G \to H$. We have $|\overline{S}^n| \leq |S^n|$. But it is known that the subgroup C is not distorted in G in the sense that there is a constant $K > 0$ such that for any $n > 0$, we have $|S^n \cap C| \leq Kn$ [23, Proposition 1.2 (2)]. It implies that for all n, we have $|S^n| \leq Kn|\overline{S}^n|$. It follows that $e(G, S) = e(H, \overline{S})$; therefore, $\xi(G) \subset \xi(H)$. On the other hand, if S is a finite generating set of H, then there is a finite generating set $\widetilde{S}$ of G which projects to S, by lifting each element of S to G, then adding a generator of C, which gives $\widetilde{S}$. Then as we saw, $e(G, \widetilde{S}) = e(H, S)$, which implies $\xi(H) \subset \xi(G)$. We proved the claim.

But by [13], $\xi(H)$ is well ordered, therefore so is $\xi(G)$. ∎

A torus bundle over a circle either admits Sol-geometry, or it is Seifert fibered (see, e.g., [30]). Therefore, the theorem covers all closed, orientable 3-manifolds with fundamental groups of exponential growth except that M has Sol-geometry. We leave it as a question.

Question 5.8. Let M a closed, orientable 3-manifold that has the geometry of three-dimensional solvable group, Sol. Then is $\xi(\pi_1(M))$ well ordered? Also, in view of Proposition 2.10, is it true that if $|S| \to \infty$ then $e(\pi_1(M), S) \to \infty$?

6. The set of growth of subgroups

We discuss the set of growth of subgroups in a finitely generated group G. Define

$$\Theta(G) = \{e(H, S) \mid S \subset G, |S| < \infty, H = \langle S \rangle, e(H, S) > 1\}.$$

The set $\Theta(G)$ is countable and contains $\xi(G)$ as a subset. If G is a hyperbolic group, it is known by [13, Section 5] that $\Theta(G)$ is well ordered.

6.1. Subgroups with hyperbolic elements

As usual, suppose G acts on a δ-hyperbolic space X. We introduce a subset in $\Theta(G)$ as follows:

$$\Theta_X(G) = \{e(H, S) \mid S \subset G, |S| < \infty, H = \langle S \rangle, e(H, S) > 1\},$$

where in addition we only consider S such that $\langle S \rangle$ *contains a hyperbolic element on* X. This set depends on the action on X.

Theorem 6.1 (cf. Theorem 3.1). *Suppose G acts on a δ-hyperbolic space X, and G is not virtually cyclic. Let $D(\varepsilon)$ be a function for WPD. Assume that there exists a constant M such that if $\langle S \rangle$ contains a hyperbolic element on X for a finite subset $S \subset G$, then S^M contains a hyperbolic element that is D-WPD. Assume that G is equationally Noetherian. Then, $\Theta_X(G)$ is a well-ordered set.*

Proof. The proof is nearly identical to Theorem 3.1.

Let $\{S_n\}$ be a sequence of finite generating sets of subgroups of exponential growth in G, $\{H_n\}$, such that each $H_n = \langle S_n \rangle$ contains a hyperbolic element on X and that $\{e(H_n, S_n)\}$ is a strictly decreasing sequence with $\lim_{n\to\infty} e(H_n, S_n) = d$, for some $d \geq 1$.

By our assumption, Proposition 2.10 applies to H_n. Therefore, $|S_n|$ is uniformly bounded from above. By passing to a subsequence we may assume that $|S_n|$ is constant, $|S_n| = \ell$, for the entire sequence.

Let $S_n = \{x_1^{(n)}, \ldots, x_\ell^{(n)}\}$. Let F be the free group of rank ℓ with a free generating set: $S = \{s_1, \ldots, s_\ell\}$. For each index n, we define a map: $f_n : F \to G$, by setting: $f_n(s_i) = x_i^{(n)}$. By construction: $e(H_n, S_n) = e(H_n, f_n(S))$.

Then, as before, the sequence $\{f_n : F \to G\}$ subconverges to a surjective homomorphism $\eta : F \to L$, where L is a limit group over G.

By assumption, G is equationally Noetherian. By the general principle (Lemma 1.11), there exists an epimorphism $h_n : L \to G$ such that by passing to a subsequence we may assume that all the homomorphisms $\{f_n\}$ factor through the limit epimorphism: $\eta : F \to L$, that is, $f_n = h_n \circ \eta$.

Then, we have the following.

Proposition 6.2 (cf. Proposition 3.2). *The following holds* .

$$\lim_{n\to\infty} e(H_n, f_n(S)) = e(L, \eta(S)).$$

Remark 6.3. This proposition is a special case of Proposition 4.9, when all δ-hyperbolic spaces X_n are X.

Proof. The proof is identical to Proposition 3.2. As in the beginning of the argument (Section 4.1), for each n, we pick one hyperbolic isometry $g \in S_n^M$ on X that is D-WPD to start with. This is possible since $H_n = \langle S_n \rangle$ contains a such hyperbolic element by the definition of $\Theta_X(G)$. The rest is same, and we omit it. ■

We continue the proof of the theorem. As in the proof of Theorem 3.1, Proposition 6.2 proves that there is no strictly decreasing sequence of rates of growth, $\{e(H_n, S_n)\}$, since a strictly decreasing sequence cannot approach its upper bound, a contradiction. Hence, $\Theta_X(G)$ is well ordered. The theorem is proved. ■

6.2. Relatively hyperbolic groups

We apply Theorem 6.1 relatively hyperbolic groups. Let $(G, \{P_i\})$ be a relatively hyperbolic group with P_i finitely generated. Define

$$\Theta_{\text{non-elem.}}(G) = \{e(H, S) \mid S \subset G, |S| < \infty, H = \langle S \rangle, e(H, S) > 1\},$$

where in addition we only consider H that is not conjugate into any P_i. This is a subset in $\Theta(G)$.

We first characterize the subgroups H that appear in the definition in terms of the action on a δ-hyperbolic space X that we described in Section 5.1. Fix such X and an action by G.

We recall a lemma. This is straightforward from the classification of subgroups that act on a hyperbolic space [14, Section 3.1].

Lemma 6.4. *Let $H < G$ be a subgroup. Then the following two are equivalent:*

(1) *H has an element g that is hyperbolic on X, and H is not virtually $\mathbb{Z}$.*

(2) *H is infinite, not virtually $\mathbb{Z}$ and not conjugate into any P_i.*

The lemma implies that

$$\Theta_{\text{non-elem.}}(G) = \Theta_X(G).$$

We prove the following.

Theorem 6.5. *Let G be a group that is hyperbolic relative to a collection of subgroups $\{P_1, \ldots, P_n\}$. Suppose G is not virtually cyclic, and not equal to P_i for any i. Suppose each P_i is finitely generated and equationally Noetherian. Then $\Theta_{\text{non-elem.}}(G)$ is well ordered.*

Proof. It suffices to argue that $\Theta_X(G)$ is well ordered. For that, we apply Theorem 6.1 to the action of G on X. We already checked the assumptions in the proof of Theorem 5.4, except that Lemma 5.1 holds for subgroups. Namely, there is a constant M such that for any finite set $S \subset G$ such that $\langle S \rangle$ contains a hyperbolic isometry on X, then S^M contains a hyperbolic isometry. But this is true and the argument is identical to the proof of [31, Proposition 5.1], and we omit it. ■

As an example of Theorem 6.5, we prove the following.

Theorem 6.6. *Let G be a group in Theorem 5.5. Then $\Theta(G)$ is a well-ordered set.*

Proof. As we said in the proof of Theorem 5.5, G is relatively hyperbolic with respect to the parabolic subgroups $\{H_i\}$, which are associated with the cusps, and H_i are virtually nilpotent. By applying Theorem 6.5 to $(G, \{H_i\})$, we have that $\Theta_{\text{non-elem.}}(G)$ is well ordered. But if a subgroup $H = \langle S \rangle$ is conjugate into one of H_i, then it is virtually nilpotent, so that H has polynomial growth. It follows that $\Theta_{\text{non-elem.}} = \Theta(G)$ holds for G, so that $\Theta(G)$ is well ordered. ■

6.3. Subgroups in mapping class groups

We discuss mapping class groups. A subgroup $H < MCG(\Sigma)$ is called *large* if it contains two independent pseudo-Anosov elements. Such H has exponential growth.

We define

$$\Theta_{\text{large}}(MCG(\Sigma)) = \{e(H, S) \mid S \subset MCG(\Sigma), |S| < \infty, \langle S \rangle = H, e(H, S) > 1\},$$

where in addition $H < MCG(\Sigma)$ is large. Note that $\xi(MCG(\Sigma)) \subset \Theta_{\text{large}}(MCG(\Sigma))$.

Theorem 6.7. *If $G = MCG(\Sigma)$ is equationally Noetherian, then $\Theta_{\text{large}}(MCG(\Sigma))$ is well ordered.*

Proof. We suppress Σ and denote MCG. Let X the curve graph of Σ. Then, as we said, Theorem 6.1 applies to the action of MCG on X. We conclude that $\Theta_X(MCG)$ is well ordered. Now, we claim $\Theta_X(MCG) = \Theta_{\text{large}}(MCG)$. Indeed, given $S \subset MCG$, if $H = \langle S \rangle$ contains a hyperbolic element on X, then it is a pseudo-Anosov element, and moreover, from $e(H, S) > 1$, H must be large. We showed $\Theta_X(MCG) \subset \Theta_{\text{large}}(MCG)$. On the other hand, for $S \subset MCG$ if $H = \langle S \rangle$ is large in MCG, then H contains hyperbolic isometries on X, so that $\Theta_{\text{large}}(MCG) \subset \Theta_X(MCG)$. ■

It is natural to ask the following question. To deal with a non-large subgroup, considering the action on the curve graph does not seem to be enough.

Question 6.8. Is $\Theta(MCG(\Sigma))$ well ordered?

7. Finiteness

7.1. Finiteness of equal growth generating sets

If G is a hyperbolic group, it is known by [13, Section 3] that for $\rho \in \xi(G)$, there are only finitely many generating sets S of G, up to $\mathrm{Aut}(G)$, such that $\rho = e(G, S)$. We discuss this issue.

Theorem 7.1 (Finiteness, cf. [13, Theorem 3.1]). *Suppose that a finitely generated group G acts on a δ-hyperbolic space X and G is not virtually cyclic. Let $D(\varepsilon)$ be a WPD-function. Assume that there exists a constant M such that if S is a finite generating set of G, then S^M contains a hyperbolic element that is D-WPD. Assume that G is equationally Noetherian.*

Then for any $\rho \in \xi(G)$, up to the action on $\mathrm{Aut}(G)$*, there are at most finitely many finite generating set S such that $e(G, S) = \rho$.*

Proof. We argue by contradiction. Suppose that there are infinitely many finite sets of generators $\{S_n\}$ that satisfy: $e(G, S_n) = \rho$, and no pair of generating sets S_n is equivalent under the action of the automorphism group $\mathrm{Aut}(G)$. As in the proof of Theorem 3.1, by Proposition 2.10, the cardinality of the generating sets $\{S_n\}$ is bounded, so we may pass to a subsequence that have a fixed cardinality ℓ. Hence, each generating set S_n corresponds to an epimorphism, $f_n : F \to G$, where S is a fixed free generating set of F, and $f_n(S) = S_n$.

By passing to a further subsequence, we may assume that the sequence of epimorphisms $\{f_n\}$ converges to a limit group L with $\eta : F \to L$ the associated quotient map. As in the proof of Theorem 3.1, since G is equationally Noetherian, by Lemma 1.11, for large n, $f_n = h_n \circ \eta$, where $h_n : L \to G$ is an epimorphism. In particular, $S_n = h_n(\eta(S))$. We pass to a further subsequence such that for every n, we have $f_n = h_n \circ \eta$. We keep using $\{f_n\}$ and $\{h_n\}$ to denote those subsequences. From now on, we discuss those subsequences.

Since for every index n, h_n is an epimorphism from L onto G that maps $\eta(S)$ to $f_n(S)$, we have $e(G, f_n(S)) \leq e(L, \eta(S))$. We prove the following.

Proposition 7.2 (cf. [13, Proposition 3.2]). *If* $\ker(h_{n_0})$ *is infinite for some n_0, then $e(G, f_{n_0}(S)) < e(L, \eta(S))$.*

We postpone the proof of the proposition until the next sections, and proceed. We prove a lemma.

Lemma 7.3. *The group L contains a finite normal subgroup $N = N_L$ that contains all finite normal subgroups in L, such that $|N| \leq 2D(100\delta)$.*

We recall one fact we use in the proof. If a finitely generated group G acts on a δ-hyperbolic space X such that G is not virtually cyclic and G contains a hyperbolic element on X that is D-WPD, then G contains a maximal finite normal subgroup $N < G$. Moreover, $|N| \leq 2D(100\delta)$. We sometimes denote N by N_G.

The existence of such N is known for an acylindrically hyperbolic group [8, Theorem 6.14], and the same proof applies to our setting, which we briefly recall. Indeed, N is the intersection of $E(g)$ for all hyperbolic elements $g \in G$ on X. It is obvious that N is normal. By assumption, there must be a hyperbolic and WPD element, g. Also, there is another element h such that g and h are independent. Then, by Proposition 2.6, $E(g) \cap E(h)$ is finite, so that N is finite. On the other hand, if N' is a finite normal subgroup in G, then for every hyperbolic element $g \in G$, there is $n > 0$ such that N' is contained in the centralizer of g^n, so that $N' < E(g)$. It implies that $N' < N$. We showed that N is maximal.

Lastly, to see $|N| \leq 2D(100\delta)$, consider the exact sequence $1 \to F(g) \to E(g) \to C \to 1$ for the hyperbolic and D-WPD element g. Recall that $|F(g)| \leq D(100\delta)$ if C is cyclic. From this we have $|N| \leq 2D(100\delta)$.

We prove the lemma.

Proof. Let $N < L$ be a finite normal subgroup. Since h_n is surjective, $h_n(N) < G$ is a finite normal subgroup; therefore, $h_n(N) < N_G$ for any h_n. Also, for sufficiently large n, the surjection $h_n : L \to G$ is injective on N. But since $|N_G| \leq 2D(100\delta)$, we have $|N| \leq 2D(100\delta)$.

If $N_1, N_2 < L$ are two finite normal subgroups, then $N_1 N_2$ is a finite normal subgroup. Combined with the fact in the previous paragraph, there must be the maximal finite normal subgroup N_L in L with $|N_L| \leq 2D(100\delta)$. ∎

We go back to the proof of the theorem. By Proposition 3.2, $\lim_{n\to\infty} e(G, S_n) = e(L, \eta(S))$. By our assumption, for every index n, $e(G, S_n) = \rho$. Hence, $e(L, \eta(S)) = \rho$, so that for every n, $e(G, S_n) = e(L, \eta(S))$.

It follows from Proposition 7.2 that for every n, $\ker(h_n)$ is finite. Since $\ker(h_n)$ is a normal subgroup in L, by Lemma 7.3, $\ker(h_n) < N_L$. Since N_L is a finite group, there are only finitely many possibilities for $\ker(h_n)$. It follows that there must be $N_0 < N_L$ such that $\ker(h_n) = N_0$ for infinitely many n.

The map h_n induces an isomorphism from $L/\ker(h_n)$ to G. Notice that this gives an isomorphism from $(L/\ker(h_n), \eta(S))$ to (G, S_n) since h_n gives a bijection between $\eta(S)$ and S_n. (Here, we may assume that each S_n consists of distinct elements, so that no two elements in $\eta(S)$ are identified by h_n.) But this implies that $(L/N_0, \eta(S))$ is isomorphic to (G, S_n) for infinitely many n by h_n, that is, those (G, S_n) are isomorphic to each other. This is a contradiction since all of them must be non-isomorphic. Theorem 7.1 is proved. ∎

7.2. Idea of the proof of Proposition 7.2

We prove Proposition 7.2. The argument is long and complicated, but the main idea is same as the proof of [13, Proposition 3.2], and we adapt it to our setting. Also, the proof is similar to the proof of Proposition 3.2, which also follows the counterpart in the paper [13]. The difference between this paper and [13] is that while they use the action of the limit group L on a limit object, called a limit tree, while in our paper we use the actions of L on X induced from the maps $h_n : L \to G$. But this approach is already taken in the proof of Proposition 3.2.

So, rather than giving a full formal proof, we first explain the strategy of the proof, then give all definitions and intermediate claims, which appear in the proof of [13, Proposition 3.2], then explain the part where we need to make technical modifications, most of which already appeared in Section 4. One advantage of not using the action of L on a limit object is that one does not need to deal with the degeneration of the action on the limit object. A trade-off is that we need to keep attention to the various constants related to the actions induced by h_n through the argument.

Strategy of the proof. We start with explaining the idea. The constant n_0 is given in the assumption, which gives the homomorphism $h_{n_0} : (L, \eta(S)) \to (G, f_{n_0}(S))$ with an infinite kernel. To show that $e(G, f_{n_0}(S)) < e(L, \eta(S))$, for each $g \in G$, we will produce not only infinitely many (which is obvious by the assumption), but "exponentially many" elements in the preimage of g by h_{n_0}. They are exponentially many in terms of the word length of g with respect to $f_{n_0}(S)$. See the estimate (7.2). Assigning the set of those elements to g is given by the map ϕ_n.

In the proof two positive integers m and n appear. They will be chosen and fixed around the end of the proof. The constant $m > 1$ is first chosen. It will be used to measure the gap between $e(G, f_{n_0}(S))$ and $e(L, \eta(S))$. The constant n depends on m, so that h_n is injective on $B_{2m}(L, \eta(S))$. Also, a positive integer q is used to make the word length of g, which is mq longer and take a limit at the end of the proof.

We now explain more concretely. Similar to $B_m(G, S_n)$, the ball of radius m in $\mathrm{Cayley}(G, S_n)$ centered at the identity, let $\mathrm{Sph}_m(G, S_n)$ denote the sphere of radius m. It is an elementary fact that if $e(G, S_n) > 1$, then

$$e(G, S_n) = \limsup_{m \to \infty} |\mathrm{Sph}_m(G, S_n)|^{\frac{1}{m}}.$$

Given $m > 0$, for a large enough $n > 0$ depending on m, we will define a "map" ϕ_n:

$$\phi_n : \mathrm{Sph}_{mq}(G, S_{n_0}) \to B_{q(m+2b)}(L, \eta(S))$$

for all $q > 0$, where b is a constant that does not depend on n, m, q. The map ϕ_n is similar to the map Φ_n in Section 4.2, but strictly speaking ϕ_n is not a map, but $\phi_n(g)$ is a finite set of elements in $B_{q(m+2b)}(L, \eta(S))$ for each g. But we abuse the notation and call them maps in the following account.

We make a remark on some confusing point. As we just said, ϕ_n is defined only for a sufficiently large n. But the n_0 in the assumption does not have to be large (it is given in the assumption, and we do not change it). We try to understand the set $h_{n_0}^{-1}(g) \subset L$ using the action on X induced by h_n. We need to take n large enough for a better understanding.

We will arrange the following two properties for every $g \in \mathrm{Sph}_{mq}(G, S_{n_0})$. Set $\mathcal{D} = D(200\delta)$.

(i)

$$\left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q \leq |\phi_n(g)|.$$

See the estimate (7.1).

(ii) $h_{n_0} \circ \phi_n(g) = g$. This implies that for $g \neq h$, we have $\phi_n(g) \cap \phi_n(h) = \emptyset$.

Once we have such a map ϕ_n, we argue as follows: Fix a (large) m. Since $\phi_n(g) \subset B_{q(m+2b)}(L, \eta(S))$, we have from (i) and (ii) that

$$|\mathrm{Sph}_{mq}(G, S_{n_0})| \left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q \leq |B_{q(m+2b)}(L, \eta(S))|.$$

Taking log, dividing by mq, and letting $q \to \infty$, we have as the limsup

$$\log e(G, S_{n_0}) + \frac{\log(m-1) - \log(\mathcal{D}^4(\mathcal{D}-1))}{m} \leq \frac{m+2b}{m} \log e(L, \eta(S)).$$

Since b does not depend on m, choosing m large enough, this inequality implies

$$\log e(G, S_{n_0}) < \log e(L, \eta(S)).$$

Roughly speaking, the construction of ϕ_n is as follows. As in the construction of Φ_n, we first construct separators. To define separators in L, we use a non-trivial element $r_n \in \ker h_{n_0}$. Separators will be products of conjugates of r_n, so that they are also in $\ker h_{n_0}$, which will imply property (ii). The separators depend on n.

For each n, the map $h_n : (L, \eta(S)) \to (G, S_n)$ gives a canonical bijection between $\eta(S)$ and $S_n = f_n(S_n)$. This gives a bijection between the words on $\eta(S)$ (not elements in L) and the words on S_n.

Let $m, q > 0$ be integers. We will fix m and let $q \to \infty$ later. Given an element $g \in \mathrm{Sph}_{mq}(G, S_{n_0})$, we choose a word $w(g)$ of length mq on S_{n_0} that represents g. We divide $w(g)$ into q subwords of length m. As we said, each subword of length m canonically gives a word of length m on $\eta(S)$ via the map h_{n_0}. We further subdivide each of the subwords of length m on $\eta(S)$ into two words of length k and $m-k$. We choose k to satisfy $1 \leq k \leq m-1$. In this way, for each choice of a q-tuple of such ks, we divided the word on $\eta(S)$ corresponding to $w(g)$ into $2q$ subwords. There are $(m-1)^q$ ways to subdivide it.

To each of such subdivision, we insert separators to the $(2q-1)$ break points and obtain an element in $B_{q(m+2b)}(L, \eta(S))$ since the word length of each separator is at most b. We obtain $(m-1)^q$ such elements. Since separators are in $\ker(h_{n_0})$, those elements (words) are mapped to g by h_{n_0}.

But we do not know if they are all distinct as elements in L, but we will show that there are at least $\left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q$ elements that are distinct. They are called *feasible elements*. This collection of feasible elements is denoted by $\phi_n(g)$. They are significantly many, so that $e(L, \eta(S))$ is strictly larger than $e(G, S_{n_0})$ as we computed in the above.

Lastly, to show that those feasible elements in L defined for each g are distinct, as in Section 4, we let them act on the space X via the map h_n for a large enough n. (We choose n such that h_n is injective on $B_{2m}(L, \eta(S))$.) We then argue that the images of the base point $y_n \in X$ by those elements are distinct. In the paper [13], they use a limit tree Y on which L acts to argue that feasible elements are distinct. Here, we use the action on X. This difference already appeared in Section 4.

7.3. Proof

We prove Proposition 7.2. It will occupy the whole subsection.

Proof. As in Section 4.1, set $\mathcal{D} = D(200\delta)$, and for each n let $y_n \in X$ be a point where $L(S_n^{2\mathcal{D}M})$ is achieved and put

$$\Delta_n = 100\delta + 4\mathcal{D}L(S_n^{2\mathcal{D}M}).$$

We consider germs with respect to Δ_n.

As we did in Lemma 4.2, in the next lemma we construct separators $u_i \in G$, which give $\hat{u}_1, \hat{u}_2, \hat{u}_3, \hat{u}_4 \in L$ by pulling them back by h_n. In addition to the properties in Lemma 4.2, they satisfy $h_{n_0}(\hat{u}_i) = 1$ in G. We state it as a lemma. The constant b is different from the constant b in Lemma 4.2, and it depends on $\ker(h_{n_0})$ but not on n.

Lemma 7.4 (cf. [13, Lemmas 3.3 and 3.9]). *Suppose* $\ker(h_{n_0})$ *is infinite. Then there exists a constant b with the following property: If n is sufficiently large, then there are elements $u_1, u_2, u_3, u_4 \in S_n^b$ that satisfy the conditions* (i)*,* (ii)*, and* (iii) *in Lemma 4.2, and in addition to that, all u_i satisfy*

(iv) $u_i \in h_n(\ker h_{n_0})$.

Moreover, those elements are such that there are elements $\hat{u}_1, \hat{u}_2, \hat{u}_3, \hat{u}_4 \in \eta(S)^b$ with $h_n(\hat{u}_i) = u_i$ and $u_i \in$ $\ker(h_{n_0})$ *for each i.*

A few remarks are in order. We will call both u_i and $\hat{u}_i$ *separators*. The moreover part will be immediate from the construction of separators, that is, we first construct $\hat{u}_i$ then map them by h_n. The separators depend on n, but we will not explicitly write n. What will be important is that the constant b does not depend on n.

Proof. In the proof there will be several constants $b_1, b_2, \ldots,$ for which we do not try to give explicit values. The important property of those constants is that they do not depend on n.

First, since $\ker(h_{n_0})$ is infinite, choose distinct elements $r_1, \ldots, r_{\mathcal{D}+1} \in L$ that are in the kernel. Let b_1 be the maximum of the word lengths of the r_i in terms of $\eta(S)$. If n is large enough, then the image of those $\mathcal{D}+1$ elements by h_n are all distinct. From now on, we only consider such n. In the following we fix each such n and argue. We have $\mathcal{D}+1$ distinct elements $\{h_n(r_i)\}$, and clearly the word lengths of those with respect to S_n are bounded by b_1.

Second, choose and fix an element $g_n \in S_n^M$ that is hyperbolic on X such that its 10δ-axis is at distance at most 10δ from the point y_n. By assumption, such an element exists. Also there is $s \in S_n$ such that g_n and sg_ns^{-1} are independent. Choose such s.

Third, we choose one element, r, from the r_is as follows: if there is r_i such that

(I) The element $h_n(r_i)$ is hyperbolic on X,
then choose one of such r_i and set $r = r_i$. Otherwise choose r_i with

(II) $h_n(r_i) \notin F(g_n)$
and set $r = r_i$. This is clearly possible since $|F(g_n)| \leq \mathcal{D}$.

Note that the element r depends on n. From now on we suppress n and write g_n as g.

Now we divide the case into two depending on (I) or (II). Suppose we are in case (I). We consider the power g^k with $k = 60\mathcal{D}$, then we have (see Lemma 2.8)

$$\langle g^k, sg^ks^{-1}\rangle = \langle g^k\rangle * \langle sg^ks^{-1}\rangle.$$

Note that we have

$$\lambda(g^k) \leq \Delta_n - 100\delta.$$

This is because since $k = 60\mathcal{D}$ and $g \in S_n^M$, we have

$$\lambda(g^k) = 30\lambda(g^{2\mathcal{D}}) \leq 30L(S_n^{2\mathcal{D}M}) \leq \Delta_n - 100\delta.$$

The last inequality is by $\mathcal{D} \geq 10$ and $\Delta_n = 100\delta + 4\mathcal{D}L(S_n^{2\mathcal{D}M})$.

Recall that the axes of g^k, sg^ks^{-1} are at at most $40\delta + L(S_n^{2M\mathcal{D}})$ from y_n.

In the proof of Lemma 4.2, we set $w = g^k, z = sg^ks^{-1}$ and produce u_i as words on w, z, but this time we take into account the germs of the element $h_n(r)$ and choose $z, w \in \langle g^k\rangle * \langle sg^ks^{-1}\rangle$ as follows.

Notice that six elements $g^k, g^{-k}, sg^ks^{-1}, sg^{-k}s^{-1}, g^ksg^ks^{-1}, g^ksg^{-k}s^{-1}$, define six germs at y_n that are mutually opposite since $\lambda(g^k) \leq \Delta_n - 100\delta$ as we noted. From the six, choose four distinct germs that are opposite to the germs for $h_n(r)$ and $h_n(r)^{-1}$. If those two germs are empty, then ignore this condition. Denote those four germs as $\gamma_1, \gamma_2, \gamma_3, \gamma_4$. Now choose $w, z \in \langle g^k, sg^ks^{-1}\rangle$ such that the germ for w, w^{-1}, z, z^{-1} is equivalent to $\gamma_1, \gamma_3, \gamma_2, \gamma_4$, respectively, such that the axes of w, z are at most $60\delta + L(S_n^{2\mathcal{D}M})$ from y_n.

We also arrange that the axes of w, z are not parallel to each other, and no element of G flips the axes of w or z. This is achieved by a similar technique to the one we used to prove Lemma 2.8, so we do not repeat. As in the proof of Lemma 4.2, there exists a constant b_2 that depends only on δ, M and $\mathcal{D} = D(200\delta)$ such that the word lengths of w, z with respect to S_n are bounded by b_2. The constant b_2 does not depend on n, nor the choice of r.

Now, choose $\widehat{z}, \widehat{w} \in L$ with $h_n(\widehat{z}) = z, h_n(\widehat{w}) = w$, whose word length with respect to $\eta(S)$ is also bounded by b_2. Using $\widehat{z}, \widehat{w}, r$ we define $\widehat{u}_i \in L$ as follows:

$$\begin{aligned}
\widehat{u}_1 &= \widehat{w} r \widehat{w}^{-1} \cdot \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^2 r \widehat{w}^{-2} \cdot \widehat{z} r \widehat{z}^{-1} \cdots \widehat{w}^{19} r \widehat{w}^{-19} \cdot \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{-20} r \widehat{w}^{20},\\
\widehat{u}_2 &= \widehat{w}^{21} r \widehat{w}^{-21} \cdot \widehat{z}^{-1} r \widehat{z} \cdot \widehat{w}^{22} r \widehat{w}^{-22} \cdot \widehat{z}^{-1} r \widehat{z} \cdots \widehat{w}^{40} r \widehat{w}^{-40} \cdot z^{-1} r \widehat{z},\\
\widehat{u}_3 &= \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{-41} r \widehat{w}^{41} \cdot \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{-42} r \widehat{w}^{42} \cdot \widehat{z} r \widehat{z}^{-1} \cdots \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{-60} r \widehat{w}^{60},\\
\widehat{u}_4 &= \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{61} r \widehat{w}^{-61} \cdot \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{62} r \widehat{w}^{-62} \cdots \widehat{z} r \widehat{z}^{-1} \cdot \widehat{w}^{80} r \widehat{w}^{-80} \cdot \widehat{z}^{-1} r \widehat{z}.
\end{aligned}$$

The word length of $\widehat{u}_i$ in terms of $\eta(S)$ is at most b_3, which does not depend on n. Indeed, we may set $b_3 = 21b_1 + 1420b_2$.

Clearly, $\widehat{u}_i$ are in $\ker h_{n_0}$ since they are products of conjugates of $r \in \ker h_{n_0}$. Finally, define $u_i = h_n(\widehat{u}_i) \in S_n^{b_3}$. Then, they satisfy property (iv).

We need to check those elements that satisfy the other properties, (i), (ii), and (iii) in Lemma 4.2. Regarding (i), the germ γ_1 is the germ of w, the germ γ_3 is the germ of w^{-1}, the germ γ_2 is the germ of z, and the germ γ_4 is the germ of z^{-1}. Then u_1, u_2, u_3, u_4 satisfy (i). Property (iii) is a consequence of that the axes of w and z are not parallel to each other. We skip details of the arguments since it is similar to Lemma 4.2. We point out that some of the arguments slightly differ depending on whether the germs for $h_n(r), h_n(r)^{-1}$ are defined or empty. In the empty case, we use the property that the axes of w, z are not flipped by any element of G.

In conclusion, those are desired elements, so in this case we take $b = b_3$ and we are done.

Suppose we are in case (II) when we chose r. In this case, we replace r with another element r' such that $h_n(r')$ is hyperbolic on X. We explain how we produce such an r'. First choose an element $\widehat{g} \in L$ with $h_n(\widehat{g}) = g$, where the word length of $\widehat{g}$ is at most M in terms of $\eta(S)$. Then we consider an element of the form

$$r' = r \widehat{g}^{Q} r \widehat{g}^{-Q}.$$

We will show that if $Q \geq 40\mathfrak{D}$, $h_n(r')$ is hyperbolic (see Lemma 7.6). Also, since $r \in \ker h_{n_0}$, we have $r' \in \ker h_{n_0}$. It is a well-known method to produce a hyperbolic element as a product of two non-hyperbolic elements, and we postpone an explanation on this, and proceed.

But, then the word length of r' in terms of $\eta(S)$ is at most b_5, where $b_5 = 2b_1 + 60\mathfrak{D}M$, which does not depend on n.

With the element r' we repeat the same argument as we did for r in case (I) and obtain desired u_i with a bound on the word length, uniformly over all n, which will finish the proof of Lemma 7.4.

We now explain some details on how to produce r' from r. For an element $k \in \mathrm{Isom}(X)$, define a set

$$\mathrm{Min}(k) = \{x \in X \mid |x - k(x)| \leq 100\delta\}.$$

This is a k-invariant set. We state a few standard facts on this set.

(M1) If k is not hyperbolic then $\mathrm{Min}(k)$ is not empty, since if it is empty, then $L(k) > 100\delta$, which implies k is hyperbolic, a contradiction.

(M2) If k is not hyperbolic, then for any point $y \in X$, we have

$$|y - k(y)| \geq 2(d(y, \mathrm{Min}(k)) - 400\delta).$$

We prove (M2). If $d(y, \mathrm{Min}(k)) \leq 400\delta$, then nothing to show, so let us assume $d(y, \mathrm{Min}(k)) > 400\delta$. Let $x \in \mathrm{Min}(k)$ be a point with $d(x, y) = d(\mathrm{Min}(k), y)$.

Let $z \in [x, y]$ be the point with $|x - z| = 350\delta$. We claim that $z \notin N_{10\delta}([k(x), k(y)])$ and $k(z) \notin N_{10\delta}([x, y])$. To prove the first claim by contradiction, suppose not, that is, $z \in N_{10\delta}([k(x), k(y)])$. Let $v, w \in [x, y]$ be with $|x - v| = 150\delta$ and $|x - w| = 200\delta$. Then we have $k([v, w]) \subset N_{10\delta}([x, y])$ since $|x - k(x)| \leq 100\delta$ and $z \in N_{10\delta}([k(x), k(y)])$. Then we have $|v - k(v)| \leq 50\delta$ since k is not hyperbolic. (Otherwise, v is "pushed" along the geodesic $[x, y]$ by at least 40δ by k, which implies that k is hyperbolic, impossible.) But it implies that $v \in \mathrm{Min}(k)$, which is a contradiction. We showed $z \notin N_{10\delta}([k(x), k(y)])$.

By the same argument, we can show $k(z) \notin N_{10\delta}([x, y])$.

Having those two claims, we have

$$\begin{aligned} |y - k(y)| &\geq |y - z| + |z - k(z)| + |k(z) - k(y)| - 100\delta \\ &\geq 2(|y - x| - 350\delta) - 100\delta = 2(|y - x| - 400\delta). \end{aligned}$$

We showed (M2).

We go back to the explanation. For a hyperbolic isometry g and its 10δ-axis $\mathrm{Ax}(g)$, we consider the nearest points projection in X to $\mathrm{Ax}(g)$. We denote the projection by π_g. For every point $x \in X$, although $\pi_g(x)$ is not a point, the diameter of $\pi_g(x)$ is bounded by 100δ. The following lemma is well known (see Figure 5).

Lemma 7.5 (Bounded projection). *If $g \in G$ is hyperbolic and D-WPD, and $k \in G \setminus E(g)$ and k is not hyperbolic, then the image of* $\mathrm{Min}(k)$ *in* $\mathrm{Ax}(g)$ *by the projection π_g is bounded by $2D(100\delta)L(g) + 200\delta$ in diameter.*

We prove the lemma for readers' convenience.

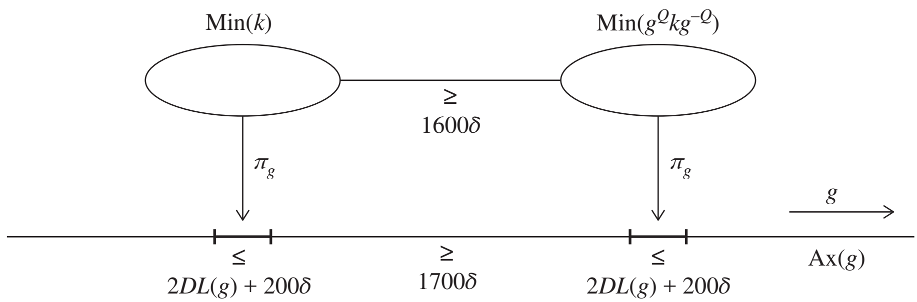


Figure 5. The Min sets and the projection to an axis.

Proof. Let $x, y \in \mathrm{Min}(k)$ and suppose $p \in \pi_g(x), q \in \pi_g(y)$. Assume that $|p-q|$ is larger than 200δ, since otherwise there is nothing to show. Then $[x,p] \cup [p,q] \cup [q,y]$ is a uniform quasi-geodesic and every point on it is moved by k by at most 200δ. Moreover, each point x on $[p,q]$ with $|x-p|, |x-q| \geq 50\delta$ is moved by k by at most 20δ since k is not hyperbolic.

Now, consider the points $p', q' \in [p,q]$ with $|p-p'| = |q-q'| = 50\delta$. Then since g is D-WPD, we have $|p'-q'| \leq 2D(100\delta)L(g) + 100\delta$ by Lemma 2.5 (2). This is because, otherwise, $k \in E(g)$, impossible. It follows that $|p-q| \leq 2D(100\delta)L(g) + 200\delta$. Lemma 7.5 is proved. ∎

The following lemma is also standard, and this is what we need for our purpose (see Figure 5).

Lemma 7.6 (Producing hyperbolic element). *For g, k as in Lemma 7.5, the element $kg^Q kg^{-Q}$ is hyperbolic if $Q \geq 40D(100\delta)$.*

We also give a brief proof.

Proof. Set $\mathfrak{D} = D(100\delta)$. (Or one can set $\mathfrak{D} = D(200\delta)$ as usual.) It does not matter since $D(100\delta) \leq D(200\delta)$.) In general, $\mathrm{Ax}(g)$ is not exactly g-invariant, but if $L(g) \geq 10\delta$ by definition. We also know that $L(g) \geq 50\delta/\mathfrak{D}$ by Lemma 2.2 (1) for our g. So, if necessary, by replacing g by $g^{\mathfrak{D}}$, we may assume that $\mathrm{Ax}(g)$ is g-invariant. For simplicity, in the following argument, we assume that $\mathrm{Ax}(g)$ is g-invariant.

Consider the set $\mathrm{Min}(g^Q kg^{-Q})$, which is equal to the set $g^Q(\mathrm{Min}(k))$. We consider the projection of those two sets by π_g. Then since $\mathrm{Ax}(g)$ is g-invariant, the projection π_g is g-equivariant. It implies that $\pi_g(\mathrm{Min}(g^Q kg^{-Q})) = g^Q(\pi_g(\mathrm{Min}(k)))$.

Then by Lemma 7.5, the distance between $\pi_g(\mathrm{Min}(g^Q kg^{-Q}))$ and $\pi_g(\mathrm{Min}(k)))$ is at least

$$QL(g) - (2\mathfrak{D}L(g) + 200\delta) \geq 38\mathfrak{D}L(g) - 200\delta \geq 1700\delta$$

since $L(g) \geq 50\delta/\mathfrak{D}$. It follows that the distance between $\mathrm{Min}(g^Q kg^{-Q})$ and $\mathrm{Min}(k)$ is at least 1600δ. It follows that the product of k and $g^Q kg^{-Q}$ is hyperbolic (this is a well-known fact in δ-hyperbolic geometry, i.e., if the distance between $\mathrm{Min}(a)$ and $\mathrm{Min}(b)$ is at least 1000δ, then ab is hyperbolic since both $\mathrm{Min}(a), \mathrm{Min}(b)$ satisfy property (M2)). Lemma 7.6 is proved. ∎

This finishes the explanation for the part to produce r' from r, and case (ii) is done. We proved Lemma 7.4. ∎

We go back to the proof of Proposition 7.2. With Lemma 7.4, the rest is very similar to [13]. Fix n that is large enough to apply Lemma 7.4. We explain how we define the map ϕ_n.

Let $g \in \mathrm{Sph}_m(G, S_{n_0})$. Choose a shortest representative word $w(g)$ of length m on S_{n_0} for g. By the bijection h_{n_0} between S_{n_0} and $\eta(S)$, $w(g)$ canonically gives a word w of length m on $\eta(S)$.

From this word w, we construct a collection of elements in L. Given a positive integer k with $1 \le k \le m-1$, we divide the word w into a prefix of length k, and a suffix of length $m-k$. The prefix corresponds to an element in L that we denote w_p^k, and the suffix corresponds to an element in L that we denote w_s^k.

Now, from the four separators we constructed in Lemma 7.4 we choose a separator $\hat{u}$ for h_n such that $\hat{u}$ is admissible for w_p^k and u^{-1} is admissible for $(w_s^k)^{-1}$, after we map them to G by h_n. We are writing $\hat{u}$ instead of u to indicate that the separator is in L.

To the pair w_p^k, w_s^k, we associate the following element in L:

$$w_p^k \hat{u} w_s^k \in B_{m+b}(L, \eta(S)).$$

Note that $h_{n_0}(w_p^k \hat{u} w_s^k) = h_{n_0}(w)$ for all k since $\hat{u} \in \ker(h_{n_0})$.

In this way, we obtain $m-1$ "words" in L from w, but possibly, some of them represent the same elements in L. To address this issue, we define a subcollection of words, called *forbidden* words (in somewhat similar way to what we did in Section 4).

Given m, take n large enough such that h_n is injective on $B_{2m}(L, \eta(S))$ from now on.

Definition 7.7 (Forbidden words, cf. [13, Definition 3.4]). Let w be a word of length m on $\eta(S)$ in the above explanation. We say that a word $w_p^k \hat{u} w_s^k$, from the collection that is built from w, is *forbidden* if there exists f, $1 \le f \le m$ such that

$$d_X(h_n(w_p^k \hat{u})(y_n), h_n(w_p^f)(y_n)) \le \frac{1}{5} d_X(y_n, h_n(\hat{u})(y_n)).$$

We give a bound on the number of forbidden words.

Lemma 7.8 (cf. [13, Lemma 3.5]). *For m and each word w as in Definition 7.7, there are at most $\frac{1}{\mathcal{D}+1} m$ forbidden words of the form: $w_p^k \hat{u} w_s^k$ for $k = 1, \dots, m-1$.*

In the proof of this lemma, we use the assumption that h_n is injective on $B_{2m}(L, \eta(S))$ as we did in the proof of Lemma 4.6.

Proof. The strategy of the proof is same as the proof of Lemma 4.6. In there, we ran the argument in $B_m(L, \eta(S))$, but here, we do it in the set $Z(w) = \{w_p^k \hat{u} w_s^k \mid k = 1, \dots, m-1\}$. Namely, if $w_p^k \hat{u} w_s^k \in Z(w)$ is forbidden for some k, then there are two other elements in $Z(w)$ that are candidates for non-forbidden elements. Then if at least one of them is forbidden, then there are two other elements in $Z(w)$ that are candidates for non-forbidden elements, and so on. We omit details. ■

Then we have the following.

Lemma 7.9 (cf. [13, Lemma 3.6]). *For m and the word w as above, the non-forbidden words: $w_p^k \hat{v}_{i,j} w_s^k$, for all k, $1 \le k \le m-1$, are distinct elements in L.*

Proof. The proof is nearly identical to the proof of [13, Lemma 3.6], and we omit it. ■

We continue with the proof of the proposition. As in Section 4, we define adequate elements in $Z(w)$. We choose a maximal subset in the set of non-forbidden elements in $Z(w)$ such that for any two distinct elements z_1, z_2 in the subset, $h_n(z_1), h_n(z_2)$ are not in the same coset with respect to $F(h_n(u_i))$ for any separator u_i. We call those elements *adequate elements*. This notion depends on n. Then, as before,

$$\frac{|\{\text{non-forbidden elements}\}|}{\mathcal{D}^4} \leq |\{\text{adequate elements}\}|.$$

This is because, as we explained, there are only four separators, $\hat{u}$, and $|F(h_n(\hat{u}))| \leq \mathcal{D}$. In conclusion, since $|Z(w)| = m - 1$,

$$\frac{m-1}{\mathcal{D}^4(\mathcal{D}+1)} \leq |\{\text{adequate elements in } Z(w)\}|. \tag{7.1}$$

Using adequate elements, we construct a collection of *feasible* words in L.

Definition 7.10 (Feasible words in L, cf. [13, Definition 3.7]). Let m, q be positive integers. Let w be a word of length mq on $\eta(S)$ that is associated with an element $g \in \mathrm{Sph}_{mq}(G, S_{n_0})$ as in the above discussion. We present w as a concatenation of q subwords of length m: $w = w(1) \cdots w(q)$.

Then for any choice of integers: $k_1, \ldots, k_q$ with $1 \leq k_t \leq m-1$, and $t = 1, \ldots, q$, for which all the elements, $w(t)_p^{k_t} v^t w(t)_s^{k_t}$, are adequate (here we drop the "hat" from v^t deliberately although they are in L to avoid confusion since we want to use it right in the below), we associate a *feasible* word (of type q) on $\eta(S)$ (in L):

$$w(1)_p^{k_1} v^1 w(1)_s^{k_1}\, \hat{v}^1\, w(2)_p^{k_2} v^2 w(2)_s^{k_2}\, \hat{v}^2\, \cdots w(q)_p^{k_q} v^q w(q)_s^{k_q},$$

where for each t, $1 \leq t \leq q-1$, $\hat{v}^t$ is one of the separators from Lemma 4.2 such that $\hat{v}^t$ is admissible for $w(t)_s^{k_t}$ and $(\hat{v}^t)^{-1}$ is admissible for $(w(t+1)_p^{k_{t+1}})^{-1}$.

Finally, we define the map ϕ_n. Suppose positive integers m, q are given. Then for $g \in \mathrm{Sph}_{mq}(G, S_{n_0})$, choose one shortest representative $w(g)$ of length mq on S_{n_0}, which defines a word $\tilde{w}(g)$ of length mq on $\eta(S)$ as Definition 7.10. From $\tilde{w}(g)$ we produce feasible words on $\eta(S)$, which define *feasible* elements in L. Note that those elements are in $B_{q(m+2b)}(L, \eta(S))$ and mapped to g by h_{n_0}. We denote this collection as $\phi_n(g)$.

We have the following.

Lemma 7.11 (cf. [13, Lemma 3.8]). *For any positive integers m, q, the feasible elements in the collection $\phi_n(g)$ we obtain for each $g \in \mathrm{Sph}_{mq}(G, S_{n_0})$ are all distinct in L.*

Moreover, all the feasible elements obtained from all the elements g in $\mathrm{Sph}_{mq}(G, S_{n_0})$ are all distinct in L.

In this lemma, n must be large enough in the sense that Lemma 7.4 applies and also, for the given m, the map h_n is injective on $B_{2m}(L, \eta(S))$ (cf. Lemma 4.5). We will choose such n in the proof.

Proof. The proof of the first sentence is similar to the proof of Lemma 4.8. So we omit it (cf. the proof of [13, Lemma 3.8].) Then the moreover part immediately follows. ■

We note that for each g, we have

$$\left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q \leq \phi_n(g). \tag{7.2}$$

This is from the lower bound (7.1) on the number of adequate elements and the lemma.

We finish the proof of the proposition 7.2. We review the setting. The constant n_0 is given to start with. We want to show $e(G, f_{n_0}(S)) < e(L, \eta(S))$. Then the constant b is given by Lemma 7.4, which does not depend on n. Choose n large enough so that we can apply Lemma 7.4.

Then choose m such that

$$\log(m-1) > 2b\log(e(L,\eta(S)) + \log(\mathcal{D}^4(\mathcal{D}-1)).$$

This implies that

$$\log(m-1) > 2b\log(e(G,S_{n_0})) + \log(\mathcal{D}^4(\mathcal{D}-1)).$$

Then choose $n > 0$ larger if necessary such that h_n is injective on $B_{2m}(L,\eta(S))$. We need this to apply Lemma 7.9.

Now, for all $q > 0$, by combining the lower bound (7.2) and (the moreover part of) Lemma 7.11, we have

$$|\mathrm{Sph}_{mq}(G,S_{n_0})|\left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q \leq |B_{q(m+2b)}(L,\eta(S))|.$$

From this,

$$\begin{aligned}
\log e(L,\eta(S)) &\geq \lim_{q\to\infty} \frac{\log\left(|\mathrm{Sph}_{mq}(G,S_{n_0})|\left(\frac{m-1}{\mathcal{D}^4(\mathcal{D}-1)}\right)^q\right)}{q(m+2b)} \\
&= \lim_{q\to\infty} \frac{\log(|\mathrm{Sph}_{mq}(G,S_{n_0})|)}{q(m+2b)} + \frac{q(\log(m-1)-\log(\mathcal{D}^4(\mathcal{D}-1)))}{q(m+2b)} \\
&= \log(e(G,S_{n_0}))\frac{m}{m+2b} + \frac{\log(m-1)-\log(\mathcal{D}^4(\mathcal{D}-1))}{m+2b} \\
&> \log(e(G,S_{n_0})).
\end{aligned}$$

The last inequality is by the way we chose m. Hence, $e(L,\eta(S)) > e(G,S_{n_0})$. We proved Proposition 7.2. ■

7.4. Family version

We state a family version of Proposition 7.2. We summarize the setting and the assumption. They are same as Proposition 4.9. Let δ, M be constants and $D(\varepsilon)$ a function for

WPD. Suppose X_n is δ-hyperbolic, a group G_n acts on X_n and S_n is a finite generating set of G_n such that S_n^M contains a hyperbolic element on X_n that is D-WPD.

Suppose that $|S_n| = \ell$ for all n, and let F be a free group of rank ℓ with a free generating set S. Let $f_n : (F, S) \to (G_n, S_n)$ be a surjection with a bijection $f_n(S) = S_n$. Assume that the sequence $\{f_n\}$ converges to $\eta : (F, S) \to (L, \eta(S))$.

Also, assume that for all n, there is a surjection $h_n : (L, \eta(S)) \to (G_n, S_n)$ with $f_n = h_n \circ \eta$.

Then we have the following generalizing Proposition 7.2, which is a special case in the sense that G_n are common and X_n are common. The proof is identical and we omit it.

Proposition 7.12 (cf. Proposition 7.2). *If* $\ker(h_{n_0})$ *is infinite for some* n_0*, then we have*

$$e(G_{n_0}, f_{n_0}(S)) < e(L, \eta(S)).$$

7.5. Finiteness for $\Theta_X(G)$

A finiteness result similar to Theorem 7.1 holds for subgroups. It is known for hyperbolic groups [13, Theorem 5.3].

Let $S_1, S_2 \subset G$ be two finite subsets. Let $H_i = \langle S_i \rangle < G$ be the subgroup generated by S_i. We say (H_1, S_1) and (H_2, S_2) are *isomorphic* if there is a bijection between S_1, S_2 that induces an isomorphism between H_1, H_2.

Theorem 7.13 (Finiteness in the subgroups case). *Assume the same condition on G as in Theorem 7.1. Moreover, we assume that if S is a finite set of G such that $\langle S \rangle$ contains a hyperbolic isometry on X, then S^M contains a hyperbolic isometry that is D-WPD. Let $\rho \in \Theta_X(G)$, Then there are at most finitely many (H, S), up to isomorphism, such that $S \subset G$ is finite, $H = \langle S \rangle$, H contains a hyperbolic element on X, and $\rho = e(H, S)$.*

The proof is nearly identical to the proof of Theorem 7.1, and we only need to modify the setting from the entire group G to subgroups.

Proof. To argue by contradiction, let $\rho \in \Theta_X(G)$ and suppose that there are infinitely many distinct, up to isomorphism, (H_n, S_n) with $e(H_n, S_n) = \rho$ such that H_n contains a hyperbolic isometry on X. Note that by assumption, S_n^M contains a hyperbolic isometry that is D-WPD.

As before, by Proposition 2.10, passing to a subsequence, one may assume that there is ℓ with $|S_n| = \ell$ for all n. Then we obtain $f_n : F \to G$ with $f_n(S) = S_n$, where F is the free group on S with $|S| = \ell$. Then, passing to a subsequence again, f_n converges to a limit group $(L, \eta(S))$ with $\eta : F \to L$. Since G is equationally Noetherian, by Lemma 1.11, passing to a further subsequence, we may assume that there are $h_n : L \to H_n < G$ with $h_n \circ \eta = f_n$ for all n.

First, by Proposition 6.2, we have that $e(H_n, S_n) = e(L, \eta(L))$ for all n.

On the other hand, we prove a version of Proposition 7.2 for subgroups: if $\ker(h_{n_0})$ is infinite for some n_0, then $e(H_{n_0}, S_{n_0}) < e(L, \eta(S))$. The proof is same and we only outline it. As before set $\mathcal{D} = D(100\delta)$. First, since H_n contains a hyperbolic isometry on X that is D-WPD, each H_n contains the maximal finite normal subgroup, which we denote by N_{H_n} with $|N_{H_n}| \leq 2\mathcal{D}$ (by Lemma 7.3).

The key step is to prove a lemma similar to Lemma 7.4. The argument is same. We use that S_n^M has a hyperbolic and D-WPD element for all n. Then, as before, for all sufficiently large n, there exists an element $r(n) \in \ker(h_{n_0})$ such that $h_n(r(n))$ is hyperbolic on X, and that the word length of $r(n)$ in terms of $\eta(S)$ is bounded uniformly on n. Then the rest is same as proving the lemma. Once we have the lemma, the rest is same to show the proposition. (We point out that this is a special case of Proposition 7.12, where X_n are common. But we did not describe the details of the argument for that.)

Combining those two, we conclude that $\ker(h_n)$ is finite for all n.

Finally, there are only finitely many possibilities for $\ker(h_n)$ since it is contained in N_L and $|N_L| \leq 2\mathcal{D}$. It implies that the desired finiteness for (H_n, S_n) holds as before. ■

7.6. Examples

We give some examples of Theorems 7.1 and 7.13. We start with relatively hyperbolic groups.

Theorem 7.14 (Finiteness for relatively hyperbolic groups). *Let G be a group that is hyperbolic relative to a collection of subgroups $\{P_1, \ldots, P_n\}$. Suppose G is not virtually cyclic, and not equal to P_i for any i. Suppose each P_i is finitely generated and equationally Noetherian. Then for each $\rho \in \xi(G)$ there are at most finitely many finite generating sets S_n of G, up to* Aut(G)*, such that $e(G, S_n) = \rho$.*

Moreover, for each $\rho \in \Theta_{\text{non-elem.}}(G)$, there are at most finitely many (H_n, S_n), up to isomorphism, such that $e(H_n, S_n) = \rho$, where $S_n \subset G$ is finite and $H_n = \langle S_n \rangle$ is not conjugate into any P_i.

Proof. Let X be a hyperbolic space on which G acts as we explained in Section 5.1. We also verified that all the assumption of Theorem 7.1 for G and the action of G on X. Recall that the action of G on X is uniformly WPD by Lemma 5.2. It implies the first part of the theorem.

For the moreover part, we apply Theorem 7.13. As Lemma 6.4 shows, $\Theta_{\text{non-elem.}}(G) = \Theta_X(G)$, which implies the conclusion. ■

Theorem 7.14 immediately implies the following as Theorem 6.5 implies Theorem 6.6.

Theorem 7.15 (Finiteness for lattices). *Let G be a group in Theorem 5.5. Then for each $\rho \in \xi(G)$ there are at most finitely many finite generating sets S_n, up to* Aut(G)*, such that $e(G, S_n) = \rho$.*

Moreover, for each $\rho \in \Theta(G)$, there are at most finitely many (H_n, S_n), up to isomorphism of H_n, such that $e(H_n, S_n) = \rho$, where $S_n \subset G$ is finite and $H_n = \langle S_n \rangle$.

Lastly, we record the following (potential) example.

Theorem 7.16 (Finiteness for MCG). *Let $MCG = MCG(\Sigma)$ be the mapping class group of a compact orientable surface Σ. Assume that it is equationally Noetherian.*

Then for each $\rho \in \xi(MCG)$ there are at most finitely many finite generating sets S_n, up to $\mathrm{Aut}(MCG)$*, such that $e(MCG, S_n) = \rho$.*

Moreover, for each $\rho \in \Theta_{\mathrm{large}}(MCG)$, there are at most finitely many (H_n, S_n), up to isomorphism, such that $e(H_n, S_n) = \rho$, where $S_n \subset MCG$ is finite and $H_n = \langle S_n \rangle$ is a large subgroup.

Proof. As we explained in Section 5.3, the action of $MCG(\Sigma)$ on the curve graph $X = \mathcal{C}(\Sigma)$ satisfies the assumption of Theorem 7.1. It is uniformly WPD.

For the moreover part, the conclusion holds for $\Theta_X(MCG)$ by Theorem 7.13. But, as we said in the proof of Theorem 6.7, we have $\Theta_X(MCG) = \Theta_{\mathrm{large}}(MCG)$, so that the conclusion holds. ∎

Acknowledgments. This work is a continuation of [13]. The author would like to thank Zlil Sela for generously offering numerous comments and suggestions throughout the work. He is grateful to Mladen Bestvina, Emmanuel Breuillard, Thomas Delzant, Daniel Groves, and Wenyuan Yang for useful comments. He feels extremely grateful to the referee for reading several versions of the paper thoroughly and making many suggestions, by which the paper significantly improved not only in terms of the presentation, but also mathematically, in particular concerning Proposition 7.2.

Funding. The author is supported in part by Grant-in-Aid for Scientific Research (No. 15H05739).

References

[1] G. N. Arzhantseva and I. G. Lysenok, A lower bound on the growth of word hyperbolic groups. *J. London Math. Soc. (2)* **73** (2006), no. 1, 109–125 Zbl 1093.20024 MR 2197373

[2] G. Baumslag, A. Myasnikov, and V. Remeslennikov, Algebraic geometry over groups I. Algebraic sets and ideal theory. *J. Algebra* **219** (1999), no. 1, 16–79 Zbl 0938.20020 MR 1707663

[3] M. Bestvina and K. Fujiwara, Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.* **6** (2002), 69–89 Zbl 1021.57001 MR 1914565

[4] B. H. Bowditch, Tight geodesics in the curve complex. *Invent. Math.* **171** (2008), no. 2, 281–300 Zbl 1185.57011 MR 2367021

[5] B. H. Bowditch, Relatively hyperbolic groups. *Internat. J. Algebra Comput.* **22** (2012), no. 3, article no. 1250016, pp. 66 Zbl 1259.20052 MR 2922380

[6] E. Breuillard and K. Fujiwara, On the joint spectral radius for isometries of non-positively curved spaces and uniform growth. *Ann. Inst. Fourier (Grenoble)* **71** (2021), no. 1, 317–391 Zbl 07362105 MR 4275871

[7] R. M. Bryant, The verbal topology of a group. *J. Algebra* **48** (1977), no. 2, 340–346 Zbl 0408.20022 MR 0453878

[8] F. Dahmani, V. Guirardel, and D. Osin, Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces. *Mem. Amer. Math. Soc.* **245** (2017), no. 1156, v+152 Zbl 1396.20041 MR 3589159

[9] T. Delzant, Kähler groups, $\mathbb{R}$-trees, and holomorphic families of Riemann surfaces. *Geom. Funct. Anal.* **26** (2016), no. 1, 160–187 Zbl 1350.53093 MR 3494488

[10] A. Eskin, S. Mozes, and H. Oh, On uniform exponential growth for linear groups. *Invent. Math.* **160** (2005), no. 1, 1–30 Zbl 1137.20024 MR 2129706

[11] B. Farb, Relatively hyperbolic groups. *Geom. Funct. Anal.* **8** (1998), no. 5, 810–840 Zbl 0985.20027 MR 1650094

[12] K. Fujiwara, Subgroups generated by two pseudo-Anosov elements in a mapping class group. I. Uniform exponential growth. In *Groups of diffeomorphisms*. Adv. Stud. Pure Math. 52, pp. 283–296, Math. Soc. Japan, Tokyo, 2008 Zbl 1170.57017 MR 2509713

[13] K. Fujiwara and Z. Sela, The rates of growth in a hyperbolic group. *Invent. Math.* **233** (2023), no. 3, 1427–1470 Zbl 1521.20097 MR 4623546

[14] M. Gromov, Hyperbolic groups. In *Essays in group theory*. Math. Sci. Res. Inst. Publ. 8, pp. 75–263, Springer, New York, 1987 Zbl 0634.20015 MR 0919829

[15] D. Groves and M. Hull, Homomorphisms to acylindrically hyperbolic groups I: Equationally noetherian groups and families. *Trans. Amer. Math. Soc.* **372** (2019), no. 10, 7141–7190 Zbl 1515.20220 MR 4024550

[16] D. Groves, M. Hull, and H. Liang, Homomorphisms to 3-manifold groups. [v1] 2021, [v3] 2024, arXiv:2103.06095v3

[17] V. S. Guba, Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems. *Math. Notes Acad. Sci. USSR* **40** (1986), 688–690 Zbl 0621.20009

[18] M. Koubi, Croissance uniforme dans les groupes hyperboliques. *Ann. Inst. Fourier (Grenoble)* **48** (1998), no. 5, 1441–1453 Zbl 0914.20033 MR 1662255

[19] J. Mangahas, Uniform uniform exponential growth of subgroups of the mapping class group. *Geom. Funct. Anal.* **19** (2010), no. 5, 1468–1480 Zbl 1207.57005 MR 2585580

[20] J. Mangahas, A recipe for short-word pseudo-Anosovs. *Amer. J. Math.* **135** (2013), no. 4, 1087–1116 Zbl 1277.57019 MR 3086070

[21] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves I: Hyperbolicity. *Invent. Math.* **138** (1999), no. 1, 103–149 Zbl 0941.32012 MR 1714338

[22] A. Minasyan and D. Osin, Acylindrical hyperbolicity of groups acting on trees. *Math. Ann.* **362** (2015), no. 3-4, 1055–1105 Zbl 1360.20038 MR 3368093

[23] H. T. Nguyen and H. Sun, Subgroup distortion of 3-manifold groups. *Trans. Amer. Math. Soc.* **373** (2020), no. 9, 6683–6711 Zbl 1483.20079 MR 4155188

[24] D. Osin, Acylindrically hyperbolic groups. *Trans. Amer. Math. Soc.* **368** (2016), no. 2, 851–888 Zbl 1380.20048 MR 3430352

[25] A. Sambusetti, Growth tightness of free and amalgamated products. *Ann. Sci. Éc. Norm. Supér. (4)* **35** (2002), no. 4, 477–488 Zbl 1018.20036 MR 1981169

[26] Z. Sela, Diophantine geometry over groups I: Makanin-Razborov diagrams. *Publ. Math. Inst. Hautes Études Sci.* (2001), no. 93, 31–105 Zbl 1018.20034 MR 1863735

[27] J.-P. Serre, *Trees* (Translated from the French by J. Stillwell). Springer, Berlin, New York, 1980 MR 0607504

[28] R. Weidmann and C. Reinfeldt, Makanin-Razborov diagrams for hyperbolic groups. *Ann. Math. Blaise Pascal* **26** (2019), no. 2, 119–208 Zbl 1454.20088 MR 4140867

[29] J. S. Wilson, On exponential growth and uniformly exponential growth for groups. *Invent. Math.* **155** (2004), no. 2, 287–303 Zbl 1065.20054 MR 2031429

[30] H. Wilton and P. Zalesskii, Profinite detection of 3-manifold decompositions. *Compos. Math.* **155** (2019), no. 2, 246–259 Zbl 1436.57020 MR 3896566

[31] X. Xie, Growth of relatively hyperbolic groups. *Proc. Amer. Math. Soc.* **135** (2007), no. 3, 695–704 Zbl 1121.20033 MR 2262865

Received 4 August 2021.

Koji Fujiwara
Department of Mathematics, Kyoto University, Kitashirakawa Oiwakecho, Kyoto 606-8502, Japan; kfujiwara@math.kyoto-u.ac.jp