RANDOM WALKS, WPD ACTIONS, AND THE CREMONA GROUP

JOSEPH MAHER, GIULIO TIOZZO

Abstract. We study random walks on groups of isometries of non-proper δ-hyperbolic spaces under the assumption that at least one element in the group satisfies Bestvina-Fujiwara's *WPD condition*. We show that in this case typical elements are WPD, and the Poisson boundary coincides with the Gromov boundary. Moreover, we show that the random walk satisfies a form of *asymptotic acylindricality*, and we use this to show that the normal closure of random elements yields almost surely infinitely many different normal subgroups. Moreover, the probability that the normal closure is free tends to 1 if and only if the maximal normal subgroup coincides with the center of the group. We apply such techniques to the Cremona group, thus obtaining that the dynamical degree of random Cremona transformations grows exponentially fast, producing many different normal subgroups, and identifying the Poisson boundary. We also give a new identification of the Poisson boundary of $\mathrm{Out}(F_n)$. Our methods give bounds on the rates of convergence for these results.

1. Introduction

A great deal of information on the geometric and algebraic properties of a group G can be derived by its isometric actions on a metric space (X, d). Following Gromov, a metric space is δ*-hyperbolic* if geodesic triangles are δ-thin. Recall that a metric space is *proper* is closed balls are compact. If the group G is not hyperbolic, then it cannot admit a proper, cocompact action on a hyperbolic metric space, but there are many interesting actions on *nonproper* hyperbolic metric spaces.

Notable examples include *relatively hyperbolic groups* which act on the coned-off Cayley graph, *right-angled Artin groups*, acting on the extension graph; the *mapping class group* of a surface, which acts on the *curve complex*; the group $\mathrm{Out}(F_n)$ of outer automorphisms of the free group, and the *Cremona group* of birational transformations of the complex projective plane.

We are going to be interested in properties of random elements of G, defined by constructing a *random walk* on G. Namely, let μ be a probability measure on G, and let (g_n) be a sequence of independent, identically distributed elements of G, with distribution μ. We are going to study the random product

$$w_n := g_1 \dots g_n$$

of elements of G. The group G need not be countable, but we will only consider probability distributions μ with countable support.

Recall that a δ-hyperbolic space X is equipped with the *Gromov boundary* ∂X given by asymptote classes of quasigeodesic rays. Under mild conditions on μ, we

Date: November 13, 2018.

proved in [MT16] that almost every sample path $(w_n x)$ converges to a point on the boundary ∂X, and that the random walk has positive drift.

Since the spaces on which G acts are not proper, some weak notion of properness is still needed in order to be able to extract information on the group from the action, and several candidate notions have been proposed in the last two decades.

First of all, following [Sel97], [Bow08], [Osi16], the action of a group G on X is *acylindrical* if for any two points x, y in X which are sufficiently far apart, the set of group elements which coarsely fixes both x and y has bounded cardinality. More precisely, given a constant $K \geqslant 0$, we define the *joint coarse stabilizer* of x and y as

$$\mathrm{Stab}_K(x,y) := \{g \in G \ : \ d(x,gx) \leqslant K \text{ and } d(y,gy) \leqslant K\}.$$

Then the action of G on X is acylindrical if for any $K \geqslant 0$, there are constants $R(K)$ and $N(K)$ such that for all points x and y in X with $d(x,y) \geqslant R(K)$, we have the following bound (where $\#|A|$ is the cardinality of A):

$$\#|\mathrm{Stab}_K(x,y)| \leqslant N(K). \tag{1}$$

This condition is quite useful, and it is verified in certain important cases (e.g. the action of the mapping class group on the curve complex [Bow08], or the action of a RAAG on its extension graph [KK14]). Under the assumption of acylindricality, we proved in [MT16] that the Gromov boundary of X is the Poisson boundary of the random walk.

However, acylindricality is too strong a condition in several other cases, such as the action of $\mathrm{Out}(F_n)$ on its related complexes, and the Cremona group. For this reason, in this paper we will consider group actions which satisfy the *weak proper discontinuity (WPD)* property, a weaker notion introduced by Bestvina and Fujiwara [BF02] in the context of mapping class groups. Intuitively, an element is WPD if it acts properly on its axis. In formulas, an element $g \in G$ is *WPD* if for any $x \in X$ and any $K \geqslant 0$ there exists $N > 0$ such that

$$\#|\mathrm{Stab}_K(x, g^N x)| < +\infty. \tag{2}$$

In other words, the finiteness condition is not required of all pairs of points in the space, but only of points along the axis of a given loxodromic element.

Let μ be a probability measure on the group G. We say that μ is *countable* if the support of μ is countable, and we denote as Γ_μ the semigroup generated by the support of μ. In this paper we show that as long as the semigroup Γ_μ contains *at least one* WPD element, then generic elements have all the properness properties one could wish for. In particular, one can identify the Poisson boundary, and study the normal closure of random elements. As an application, we will use this condition to derive results on the Cremona group.

1.1. **Genericity of WPD elements.** Maher [Mah11] and Rivin [Riv08] considered random walks on the mapping class group acting on the curve complex, and showed that pseudo-Anosov mapping classes are typical for random walks. More generally, in [MT16], we showed that for a group G acting non-elementarily on a Gromov hyperbolic space X, loxodromic elements are typical for the random walk: i.e., the probability that the random product of n elements is loxodromic tends to one as n tends to infinity. In this paper, we prove that as long as there is one WPD element in the support of the measure generating the random walk, then WPD elements are generic.

We say that a measure μ is *non-elementary* if Γ_μ contains at least two independent loxodromic elements, and is *bounded* if for some $x \in X$ the set $(gx)_{g \in \operatorname{supp} \mu}$ is bounded in X. Finally, μ is *WPD* if Γ_μ contains an element h which is WPD in G.

We will show that generic elements are WPD with an explicit bound on the rate of convergence: we say that a sequence of numbers (p_n) tends to 1 with *square root exponential decay* if there are constants $B > 0$ and $c < 1$ such that $p_n \geqslant 1 - Bc^{\sqrt{n}}$.

Theorem 1.1. *(Genericity of WPD elements.) Let G be a group acting on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, bounded, WPD probability measure on G. Then*

$$\mathbb{P}(w_n \text{ is WPD}) \to 1$$

as $n \to \infty$, with square root exponential decay.

In fact, we obtain that most random elements have bounded coarse stabilizer, where the bound does not depend on the point chosen. We call this property *asymptotic acylindricality*. We prove the following estimate on the joint coarse stabilizer.

Theorem 1.2. *(Asymptotic acylindricality.) Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable, non-elementary, bounded, WPD probability measure on G, and let $x \in X$. Then for any $K \geqslant 0$ there is an $N > 0$ such that*

$$\mathbb{P}(\#|\mathrm{Stab}_K(x, w_n x)| \leqslant N) \to 1$$

with square root exponential decay, where the implicit constants depend on K.

1.2. **Normal closure.** Another application of our methods is the study of normal groups obtained by taking the normal closure of a random element w_n.

In order to state the theorem, we need some assumption. We call a measure μ *reversible* if the semigroup Γ_μ generated by the support of μ is indeed a group. This condition is satisfied e.g. when the support of μ is closed under taking inverses. Given a subgroup $H < G$, we define its *injectivity radius* as

$$\mathrm{inj}(H) := \inf_{\substack{g \in H \setminus \{1\} \\ x \in X}} d(x, gx).$$

We prove that the injectivity radius of the normal closure of a random element is almost surely unbounded, and taking the normal closure of random elements yields many different normal subgroups.

Theorem 1.3. *(Abundance of normal subgroups.) Let G be a group acting on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, reversible, bounded, WPD probability measure on G. Then there exists k such that, if we consider the normal closure $N_n(\omega) := \langle\langle w_n^k \rangle\rangle$ we have:*

1. *for any $R > 0$ the probability that $\mathrm{inj}(N_n) \geqslant R$ tends to 1 as $n \to \infty$;*
2. *for almost every sample path ω, the sequence*

$$\{N_1(\omega), N_2(\omega), \dots, N_n(\omega), \dots\}$$

contains infinitely many different normal subgroups of G.

Our techniques also allow us to determine the value of k in the previous result. Moreover, one can determine the group structure of the normal closure of a random element, in particular whether it is free.

To be precise, let us denote as $E_\mu := \{g \in \Gamma_\mu \;:\; gx = x \text{ for all } x \in \partial X\}$ the pointwise stabilizer of ∂X. Note that if $G = \Gamma_\mu$, then $E_\mu = E(G)$ is the maximal finite normal subgroup of G (i.e., the largest finite subgroup of G which is normal: that such a subgroup exists is a consequence of the WPD property).

Since E_μ is normal in Γ_μ, conjugacy yields a homomorphism

$$\Gamma_\mu \to \text{Aut } E_\mu.$$

Let us denote as H_μ the image of Γ_μ in Aut E_μ. The size of H_μ will determine the structure of the normal closure of a random element, in particular whether it is free.

Theorem 1.4. *(Structure of the normal closure.) Let G be a group acting on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, reversible, bounded, WPD probability measure on G. Let us denote as H_μ the image of Γ_μ in* Aut E_μ. *Then:*

(1) the probability that the normal closure $\langle\langle w_n \rangle\rangle$ of w_n in G is free satisfies

$$\mathbb{P}(\langle\langle w_n \rangle\rangle \text{ is free}) \to \frac{1}{\# H_\mu}$$

as $n \to \infty$.

(2) Moreover, if $k = \# H_\mu$, then

$$\mathbb{P}(\langle\langle w_n^k \rangle\rangle \text{ is free}) \to 1$$

as $n \to \infty$, and indeed there exist constant $B > 0, c < 1$ such that

$$\mathbb{P}(\langle\langle w_n^k \rangle\rangle \text{ is free}) \geqslant 1 - Bc^{\sqrt{n}}$$

for any n.

Note that $k = \# H_\mu$ is also precisely the k in Theorem 1.3.

Moreover, as a corollary of Theorem 1.4, the probability that the normal closure of a random element is free detects the following algebraic property of the group:

Corollary 1.5. *If $\Gamma_\mu = G$, then*

$$\mathbb{P}\left(\langle\langle w_n \rangle\rangle \text{ is free}\right) \to 1 \qquad \text{as } n \to \infty$$

if and only if the maximal finite normal subgroup $E(G)$ equals the center $Z(G)$.

In particular, we will show later that this is the case for mapping class groups.

1.3. **The Poisson boundary.**

The well-known *Poisson representation formula* expresses a duality between bounded harmonic functions on the unit disk and bounded functions on its boundary circle. Indeed, bounded harmonic functions admit radial limit values almost surely, while integrating a boundary function against the Poisson kernel gives a harmonic function on the interior of the disk.

This picture is intimately connected with the geometry of $SL_2(\mathbb{R})$; then in the 1960's Furstenberg and others extended this duality to more general groups. In particular, let G be a countable group of isometries of a Riemannian manifold X, and let us consider a probability measure μ on G. One defines μ-harmonic functions as functions on G which satisfy the mean value property with respect to averaging using μ; in formulas $f : G \to \mathbb{R}$ is μ-harmonic if

$$f(g) = \sum_{h \in G} f(gh)\ \mu(h) \qquad \forall g \in G.$$

Following Furstenberg [Fur63], a measure space (M, ν) on which G acts is then a boundary if there is a duality between bounded, μ-harmonic functions on G and L^∞ functions on M.

A related way to interpret this duality is by looking at random walks on G. In many situations, (e.g. when X is hyperbolic) the space X is equipped naturally with a topological boundary ∂X, and almost every sample path $(w_n x)$ converges to some point on the boundary of X. Hence, one can define the *hitting measure* of the random walk as the measure ν on ∂X given on a subset $A \subseteq \partial X$ by

$$\nu(A) := \mathbb{P}\left(\lim_{n\to\infty} w_n x \in A\right).$$

A fundamental question in the field is then whether the pair $(\partial X, \nu)$ equals indeed the Poisson boundary of the random walk (G, μ), i.e. if all harmonic functions on G can be obtained by integrating a bounded, measurable function on ∂X.

In the proper case, the classical criteria in order to identify the Poisson boundary can be applied and one gets that the Gromov boundary $(\partial X, \nu)$ with the hitting measure is a model for the Poisson boundary. In the non-proper case, the classical entropy criterion is not expected to work, as there may be infinitely many group elements contained in a ball of fixed diameter.

We prove, however, that as long as Γ_μ contains a WPD element, the Poisson boundary indeed coincides with the Gromov boundary.

Theorem 1.6. *(Poisson boundary for WPD actions.) Let G be a countable group which acts by isometries on a δ-hyperbolic metric space (X, d), and let μ be a non-elementary probability measure on G with finite logarithmic moment and finite entropy. Suppose that there exists at least one WPD element h in the semigroup generated by the support of μ. Then the Gromov boundary of X with the hitting measure is a model for the Poisson boundary of the random walk (G, μ).*

The result extends our earlier result in [MT16] for acylindrical actions.

1.4. **The Cremona group.** The *Cremona group* is the group $G = \mathrm{Bir}\ \mathbb{P}^2(\mathbb{C})$ of birational transformations of the projective plane.

Let $f : \mathbb{P}^2(\mathbb{C}) \to \mathbb{P}^2(\mathbb{C})$ be a birational map. Then f is given in homogeneous coordinates by

$$f([x : y : z]) := [P : Q : R]$$

where P, Q, R are polynomials of degree d without common factors. We call d the *degree* of f, and we denote it as $\deg f$.

Now, one notes that $\deg(f^{n+m}) \leqslant \deg(f^n) \cdot \deg(f^m)$, but the equality need not hold: the most famous example is the *Cremona involution*

$$g([x : y : z]) := [yz : xz : xy]$$

which has degree 2, but g^2 is the identity; the Cremona group is in fact generated by degree 1 transformations and the Cremona involution. Hence, following [Fri95], [RS97] we define the *dynamical degree* of f as

$$\lambda(f) := \lim_{n\to\infty} (\deg\ f^n)^{1/n}.$$

The dynamical degree is always an algebraic integer [DF01], and it is related to the topological entropy by $h_{top}(f) \leqslant \log \lambda(f)$. In fact, equality is conjectured [Fri95].

We are interested in the properties of random Cremona transformations. Let us fix a probability measure μ on the Cremona group, with countable support. Then let us draw a sequence (g_n) of elements independently with distribution μ, and consider the random product

$$f_n := g_1 g_2 \dots g_n.$$

It is known that the Cremona group acts by isometries on an infinite dimensional hyperbolic space which is contained in the *Picard-Manin space* (see Section 3). Using such an action, we will determine asymptotic properties for random walks on the Cremona group. A measure μ on the Cremona group has *finite first moment* if $\int \deg f \, d\mu(f) < +\infty$, and is *bounded* if there exists $D < +\infty$ such that $\deg f \leqslant D$ for any $f \in \operatorname{supp}(\mu)$.

First of all, we prove that the degree and dynamical degree of a random Cremona transformation grow exponentially fast.

Theorem 1.7. *Let μ be a countable non-elementary probability measure on the Cremona group with finite first moment. Then there exists $L > 0$ such that for almost every random product $f_n = g_1 \dots g_n$ of elements of the Cremona group we have the limit*

$$\lim_{n\to\infty} \frac{1}{n} \log \deg(f_n) = L.$$

Moreover, if μ is bounded then for almost every sample path we have

$$\lim_{n\to\infty} \frac{1}{n} \log \lambda(f_n) = L.$$

Proof. The Cremona group acts by isometries on the (half) hyperboloid $\mathbb{H}_{\mathbb{P}^2}$ inside the Picard-Manin space. The space $\mathbb{H}_{\mathbb{P}^2}$ is a Gromov hyperbolic metric space, hence we can apply our techniques, and in particular for any Cremona transformation f one has $\deg(f) = \cosh d(x, fx)$ if $x = [H]$. The first claim follows from the fact that the drift of the random walk is positive (Theorem 2.5 (2)). The second claim follows from the fact that translation length of random elements grows linearly (Theorem 4.1) and that the translation length $\tau(f)$ and the dynamical degree are related by the equation $\log \lambda(f) = \tau(f)$. □

In [CL13], Cantat and Lamy showed that there exist infinitely many distinct normal subgroups of G, thus proving a celebrated conjecture of Mumford. Using our techniques, we will prove a random walk version of their theorem, namely that for almost every sample path the normal closures of random elements produce infinitely many different normal subgroups of the Cremona group.

For this group, the *injectivity radius* of a subgroup H can also be defined as

$$\operatorname{inj}(H) := \inf_{f \in H \setminus \{1\}} \deg f.$$

Let us state the consequence of Theorems 1.3 and 1.4 when applied to the Cremona group.

Theorem 1.8. *Let μ be a countable non-elementary, reversible, bounded, WPD probability measure on the Cremona group. Then there exists k such that, if we consider the normal closure $N_n(\omega) := \langle\langle f_n^k \rangle\rangle$ we have:*

(1) the probability that N_n is free tends to 1 as $n \to \infty$;
(2) for any $R > 0$ the probability that $\operatorname{inj}(N_n) \geqslant R$ texnds to 1 as $n \to \infty$;
(3) for almost every sample path ω, the sequence

$$\{N_1(\omega), N_2(\omega), \dots, N_n(\omega), \dots\}$$

contains infinitely many different normal subgroups of Bir $\mathbb{P}^2(\mathbb{C})$.

Note that the action of the Cremona group on the infinite-dimensional hyperbolic space is not acylindrical, but WPD elements actually exist: in particular, by Shepherd-Barron [SB13], a loxodromic map is WPD if and only if it is not conjugate to a monomial map (see also [Ure18]). Moreover, by ([Lon16], Proposition 4), for each $n \geqslant 2$, the transformation given in affine coordinates by $(x, y) \mapsto (y, y^n - x)$ is WPD. We wonder if one can take $k = 1$ in the above result for the Cremona group.

Finally, we have the following result concerning the Poisson boundary, which follows immediately from Theorem 1.6. Let us denote as $\mathbb{H}_{\mathbb{P}^2}$ the hyperboloid inside the Picard-Manin space of $\mathbb{P}^2(\mathbb{C})$ (see Section 3 for details).

Theorem 1.9. *Let μ be a countable, non-elementary, WPD probability measure on the Cremona group with finite entropy and finite logarithmic moment. Then the Gromov boundary of the hyperboloid $\mathbb{H}_{\mathbb{P}^2}$ with the hitting measure is a model for the Poisson boundary of (G, μ).*

A related notion to WPD is the notion of *tight* element from [CL13]. In fact, in order to produce new normal subgroups, Cantat and Lamy take the normal closure of tight elements. Let us note that in the Cremona group, centralizers of loxodromic elements are virtually cyclic; as a consequence, if an element is tight then it is also WPD.

Note that for simplicity we have dealt with the Cremona group over $\mathbb{C}$, but Theorems 1.7, 1.8, and 1.9 are still true (and with the same proofs) for the Cremona group over any algebraically closed field k.

1.5. **Tame automorphism groups.** Other groups arising in algebraic geometry admit an action on a non-proper δ-hyperbolic space with WPD elements.

First of all, the group $\mathrm{Aut}(\mathbb{C}^2)$ of polynomial automorphisms of $\mathbb{C}^2$ (see [FL10] and references therein, as well as [MO15]) can be written as an amalgamated product of two of its subgroups, hence it acts on the corresponding Bass-Serre tree, which is a Gromov-hyperbolic space; in fact, for this action every loxodromic element is WPD, but the action is not acylindrical.

Remarkably, Lamy and Przytycki recently extended this work to three variables. They considered the *tame automorphism group* $\mathrm{Tame}(\mathbb{C}^3)$, which is the group generated by affine and elementary automorphisms of $\mathbb{C}^3$ (see [LP16] for a precise definition), and showed that this group also acts on a Gromov-hyperbolic complex and there are WPD elements, so the methods of the present paper apply.

Let us finally remark that much less is known about the structure of the Cremona group in three variables, and these methods do not easily apply since there exist two linearly independent divisor classes with positive self-intersection, hence the Picard-Manin space contains a two-dimensional flat, so the analog of the hyperboloid is no longer δ-hyperbolic.

1.6. **Outer automorphisms of the free group.** Another application of our setup is to the group $Out(F_n)$ of outer automorphisms of a finitely generated free group F_n of rank $n \geqslant 2$.

There are several hyperbolic graphs on which $Out(F_n)$ acts: the main two are the *free factor complex* and the *free splitting complex*. In particular, the free factor complex $\mathcal{FF}(F_n)$ is hyperbolic by work of Bestvina and Feighn [BF14]. Moreover, an element is loxodromic on $\mathcal{FF}(F_n)$ if and only if it is *fully irreducible*, and all fully irreducible elements satisfy the WPD property. However, it is not known whether the action of $Out(F_n)$ on the free factor complex is acylindrical.

On the other hand, the free splitting complex is also hyperbolic, but the action on the free splitting complex $\mathcal{FS}(F_n)$ is known not to be acylindrical, by work of Handel and Mosher [HM13]. Moreover, an element is loxodromic if and only if it admits a filling lamination pair. This is a weaker condition than being fully irreducible, and the stabilizer of the quasi-axis of a loxodromic element need not be virtually cyclic. Thus, for this action it is not true that every loxodromic element satisfies the WPD property. However, by Theorem 1.1 even for this action WPD elements are generic for the random walk.

We have the following identification for the Poisson boundary of $Out(F_n)$.

Theorem 1.10. *Let μ be a measure on $Out(F_n)$ such that the semigroup generated by the support of μ contains at least two independent fully irreducible elements. Moreover, suppose that μ has finite entropy and finite logarithmic moment for the simplicial metric on the free factor complex. Then the Gromov boundary of the free factor complex is a model for the Poisson boundary of (G, μ).*

Proof. By [BF14], the action of fully irreducible elements on the free factor complex is WPD. Hence, the claim follows by Theorem 1.6. □

Note that the identification of the Poisson boundary for $Out(F_n)$ has been obtained by Horbez [Hor16] using the action of $Out(F_n)$ on the outer space CV_n. In our theorem above, the moment condition required is a bit weaker, as we only need the logarithmic moment condition to hold with respect to the metric on $\mathcal{FF}(F_n)$ instead of the metric on CV_n (recall that there is a coarsely defined Lipschitz map $CV_n \to \mathcal{FF}(F_n)$).

1.7. **Mapping class groups.** Let $S_{g,n}$ be a topological surface with genus g and n punctures, and let $Mod(S_{g,n})$ be its mapping class group, i.e. the group of homeomorphisms of $S_{g,n}$, up to isotopy. The mapping class group acts on a locally infinite, δ-hyperbolic graph, known as the *curve complex* [MM99]. Loxodromic elements for this action are the pseudo-Anosov mapping classes, and as they are all WPD elements, all results in our paper apply.

As an application of Theorem 1.4, we prove that the normal closure of random mapping classes is a free group, answering a question of Margalit [Mar18, Problem 10.11].

Theorem 1.11. *Let $G = Mod(S_{g,n})$ be the mapping class group of a surface of finite type, and suppose that G is infinite. Let μ be a probability measure on G with bounded support in the curve complex and such that $\Gamma_\mu = G$, and let w_n be the n^{th} step of the random walk generated by μ. Then the probability that the normal closure $\langle\langle w_n \rangle\rangle$ is free tends to 1 as $n \to \infty$.*

The result follows from Theorem 1.4 and the fact that, by the Nielsen realization theorem, the maximal normal subgroup of $Mod(S_{g,n})$ always equals its center (which is trivial unless the mapping class group contains a central hyperelliptic involution). See Section 11.2 for details. Note that in fact the action is acylindrical [Bow08], hence some applications such as the Poisson boundary already follow from [MT16].

1.8. **Matching estimates and rates.** In order to obtain our results, we need to show that a random element has finite joint coarse stabilizer, and to do so we recur to what we call *matching estimates*.

Following [CM15], we say that two geodesics γ and γ' in X have a *match* if there is a subsegment of γ close to a G-translate of a subsegment of γ' (see Definition 7.4).

Let $x \in X$ be a basepoint and (w_n) be a sample path. The two key estimates we will prove and use are the following.

(1) *Matching estimate* (Proposition 7.6): given a loxodromic element g, we show that the probability that the geodesic $[x, w_n x]$ has a match with a translate of the axis of g is at least $1 - Bc^{\sqrt{n}}$.
(2) *Non-matching estimate* (Proposition 8.2): given a geodesic segment η in X of length s, the probability that there is a match between $[x, w_n x]$ and a G-translate of η is at most $Bc^{\sqrt{s}}$.

The rate that we obtain comes from the following estimate, which is given in detail in Section 7. If T is an ergodic transformation of a probability space $(\Omega, \mathbb{P})$, then for any measurable set A of positive measure, the measure of $A_n = A \cup T^{-1}A \cup \cdots T^{-n}A$ tends to one as n tends to infinity. However, for arbitrary measurable sets, this rate of convergence can be arbitrarily slow. The ergodic transformation we shall consider is the shift map acting ergodically on the bi-infinite step space of the random walk, which is just the product probability space $(G, \mu)^{\mathbb{Z}}$. In [MT16] we showed that the image of a random walk on G in X converges to the Gromov boundary ∂X almost surely, and this gives a map $\partial_+ \colon (G, \mu)^{\mathbb{Z}} \to \partial X$, which we call the forward boundary map. Similarly, the image of a random walk in reverse time also converges to the boundary, and we shall call this the backward boundary map. We shall consider sets A whose images under the forward and backward boundary maps contain an open set in the boundary, and we shall show that for these sets, $\mathbb{P}(A_n) \to 1$ with square root exponential decay.

1.9. **Asymmetric elements.** Another important tool in our proofs is the notion of asymmetric element, which was introduced in [MS18]. We call a loxodromic element $g \in G$ *asymmetric* if any element which coarsely stabilizes a segment of the axis of g actually coarsely stabilizes the set $\{g^i x\}_{i \in \mathbb{Z}}$ (see Definition 9.1 for the precise statement). In [MS18] it is proven that if the action of G is acylindrical, then asymmetric elements are generic. In this paper, we generalize this result to WPD actions, and use it to prove the other results.

Let G_{WPD} be the set of WPD elements in G. For a loxodromic $g \in G$, let us denote as $\Lambda(g) := \{\lambda_g^+, \lambda_g^-\}$ the two fixed points of g on ∂X. We denote as $E_G(g)$ the stabilizer of $\Lambda(g)$ as a set, and as $E_G^+(g)$ the pointwise stabilizer of $\Lambda(g)$. Moreover, for a subgroup $H < G$ we denote as

$$E_G(H) := \bigcap_{H \cap G_{WPD}} E_G(h)$$

the intersection of all $E_G(h)$ as h lies in $H \cap G_{WPD}$ (a priori, this set may be smaller than the set of WPD elements for the action of H on X). Note that $E_G(G)$ is the maximal finite normal subgroup of G.

We have the following characterization of $E_G(w_n)$ for generic elements w_n. Let $E_\mu := E_G^+(\Gamma_\mu)$.

Theorem 1.12. *Given $\delta \geqslant 0$ there are constants K and L with the following properties. Let G be a group acting by isometries on a $\delta-$hyperbolic space X, and let μ be a countable, non-elementary, reversible, bounded, WPD probability distribution on G. Then there are constants $B > 0$ and $c < 1$ such that the probability that w_n is loxodromic, $(1, L, K)$-asymmetric, and WPD with*

$$E_G(w_n) = E_G^+(w_n) = \langle w_n \rangle \ltimes E_\mu$$

is at least $1 - Bc^{\sqrt{n}}$.

Note that the action of E_μ on $E_G(w_n)$ is precisely responsible for the value of k in Theorems 1.3 and 1.4. Indeed, one obtains that the cyclic group $\langle w_n \rangle$ is normal in $E_G(w_n)$ if and only if the image of w_n in Aut E_μ is trivial. Now, the random walk on Γ_μ pushes forward to a random walk on the finite group Aut E_μ, and this random walk equidistributes on the image of Γ_μ inside Aut E_μ, which we denote as H_μ. This explains the asymptotic probability of $\frac{1}{\# H_\mu}$ in Theorem 1.4.

1.10. **Further questions.** We conclude with a few questions for further exploration.

1. Can one drop "reversible" as an hypothesis in Theorem 1.4?
2. Can one take $k = 1$ in Theorems 1.3 and 1.4 when G is the Cremona group?
3. Do our results still hold for measures μ with finite exponential moment, rather than bounded measures?
4. Is the actual rate of convergence in Theorems 1.1, 1.2, and 1.4 exponential (i.e. of order c^n for some $c < 1$) instead of just of order $c^{\sqrt{n}}$?
5. Does the radius of injectivity inj(N_n) typically goes to infinity as $n \to \infty$, and at what rate?

We believe that the answers to all these questions should be positive, but we do not attempt to solve them here.

1.11. **Acknowledgements.** We would like to thank Mladen Bestvina for pointing out that the Poisson boundary result from [MT16] holds in the WPD case. We also thank Carolyn Abbott, Jeffrey Diller, Igor Dolgachev, Mattias Jonsson, Stephane Lamy and Piotr Przytycki for useful discussions and comments. The first named author acknowledges support from the Simons Foundation and PSC-CUNY. The second named author is partially supported by NSERC and the Alfred P. Sloan Foundation.

2. Background material

Let X be a δ-hyperbolic metric space, and let G be a group of isometries of X. Let μ be a probability measure on G. This defines a *random walk* by choosing for each n an element g_n of G with distribution μ independently of the previous ones, and considering the product

$$w_n := g_1 \ldots g_n.$$

The sequence $(w_n)_{n \geqslant 0}$ is called a *sample path* of the random walk, and we are interested in the asymptotic behaviour of typical sample paths.

2.1. **Isometries of hyperbolic spaces.** Recall that isometries of a δ-hyperbolic space (even if it is not proper) can be classified into three types; in particular, $g \in \mathrm{Isom}(X)$ is:

- *elliptic* if g has bounded orbits;
- *parabolic* if it has unbounded orbits, but zero translation length;
- *loxodromic* (or *hyperbolic*) if it has positive translation length.

Here, the *translation length* of $g \in \mathrm{Isom}(X)$ is the quantity

$$\tau(g) := \lim_{n\to\infty} \frac{d(g^n x, x)}{n}$$

where the limit always exists and is independent of the choice of x. Moreover, a loxodromic element has two fixed points on the Gromov boundary of X, one attracting and one repelling.

A semigroup inside $\mathrm{Isom}(X)$ is *non-elementary* if it contains two loxodromic elements which have disjoint fixed point sets on ∂X.

We will use the following elementary properties of δ-hyperbolic spaces, which we state without proof. A *quasi-axis* for a loxodromic isometry g of X is a quasigeodesic which is invariant under g.

Proposition 2.1. *Given a constant $\delta \geqslant 0$, there is a constant K_1 such that every loxodromic isometry of a δ-hyperbolic space has a $(1, K_1)$-quasi-axis.*

By abuse of notation, we will refer to a choice of $(1, K_1)$-quasi-axis as an axis for g. We will use the following fellow-travelling properties of quasigeodesics in Gromov hyperbolic spaces, whose proofs we omit.

If two parameterized geodesics have endpoints which are close together, then they are (parameterized) fellow travelers.

Lemma 2.2. *There is a constant K', which only depends on δ, such that for any two geodesics γ and γ' with unit speed parameterizations, for any constant $K \geqslant 0$ and $a \leqslant b \leqslant c$, if $d(\gamma(a), \gamma'(a)) \leqslant K$ and $d(\gamma(c), \gamma'(c)) \leqslant K$, then $d(\gamma(b), \gamma'(b)) \leqslant K + K'$.*

Given a finite geodesic $\gamma = [x, y]$, let $\gamma_K^- = \gamma \setminus (B_K(x) \cup B_K(y))$. Then we have:

Proposition 2.3. *Given $\delta \geqslant 0$ and $K_1 \geqslant 0$, there is a constant L such that for any $K \geqslant 0$ and for any two $(1, K_1)$-quasigeodesics γ and η in a δ-hyperbolic space, with endpoints distance at most K apart, any point on γ_K^- lies within distance at most L from a point on η.*

Let us recall that given $x, y \in X$ and $R \geqslant 0$, we define the *shadow* $S_x(y, R)$ as

$$S_x(y, R) := \{z \in X \ : \ (z \cdot y)_x \geqslant d(x, y) - R\}.$$

The number $r = d(x, y) - R$ is called the *distance parameter* of the shadow.

Proposition 2.4. *Given constants $\delta \geqslant 0$ and K_1, there are constants D and L with the following properties. Let $A = S_x(y, R)$ and $B = S_x(z, R)$ be disjoint shadows in a δ-hyperbolic space X, and let γ_1 and γ_2 be $(1, K_1)$-quasigeodesics each with one endpoint in A and the other endpoint in B. Then $\gamma_1 \setminus (S_x(y, R+D) \cup S_x(x, R+D))$ is contained in an L-neighbourhood of $\gamma_2 \setminus (S_x(y, R + D) \cup S_x(x, R + D))$.*

2.2. **Random walks on weakly hyperbolic groups.** In [MT16], we established many properties of typical sample paths for random walks on general groups of isometries of δ-hyperbolic spaces. Namely:

Theorem 2.5 ([MT16])**.** *Let μ be a countable, non-elementary measure on a group of isometries of a δ-hyperbolic metric space X, and let $x \in X$. Then*

(1) almost every sample path $(w_n x)$ converges to some point ξ in the Gromov boundary of X;

(2) if μ has finite first moment in X, there exists $L > 0$ such that for almost all sample paths we have

$$\lim_{n\to\infty} \frac{d(w_n x, x)}{n} = L;$$

(3) moreover, if μ is bounded, there exists $L > 0, B \geqslant 0$ and $c < 1$ such that the translation length grows linearly with exponential decay:

$$\mathbb{P}(\tau(w_n) \geqslant nL) \geqslant 1 - Bc^n.$$

Note that in [MT16] the previous result is proven under the assumption that X is *separable*, i.e. it contains a countable dense set. However, since the measure μ is countable one can drop the separability assumption, as remarked in [GST, Remark 4]. In fact, the only point where separability is used is to prove convergence to the boundary, and one can prove it for general metric spaces from the separable case and the following fact, whose proof we write here in detail.

Lemma 2.6 ([GST, Remark 4])**.** *Let Γ be a countable group of isometries of a δ-hyperbolic metric space X. Then there exists a separable metric space X' (in fact, a simplicial graph with countably many vertices) and a Γ-equivariant quasi-isometric embedding $i\colon X' \to X$. As a consequence, i extends continuously to a Γ-equivariant inclusion $\partial X' \to \partial X$ between the Gromov boundaries.*

Proof. Consider the orbit Γx of some point $x \in X$, and pick for each pair of points $p_1, p_2 \in \Gamma x$ a geodesic γ between p_1 and p_2. On each such geodesic, pick a maximal collection of points such that any two points are at least 10δ apart. Let Y be the union of the Γ-orbits of all these additional points together with Γx.

Now, for each pair of points of Y which have distance at most 100δ in X, pick a geodesic between them, and let us denote as $\mathcal{E}$ the union of all Γ-orbits of such geodesics. This way we obtained a Γ-invariant collection Y of points of X, and a Γ-invariant collection $\mathcal{E}$ of geodesic segments connecting them, where each segment has length $\leqslant 100\delta$. Let us define the graph X' to have Y as vertex set, and to have one edge for each geodesic in $\mathcal{E}$. The group Γ acts on the graph X', which has countably many vertices and edges, and so is separable. Furthermore, there is a map $i\colon X' \to X$ which sends each vertex in X' to the corresponding point in X, and each edge (v, w) to the corresponding geodesic segment. By construction, this map is Γ-equivariant.

We now show that i is a quasi-isometric embedding, and in particular

$$10\delta d_{X'}(v, w) \leqslant d_X(v, w) \leqslant 100\delta d_{X'}(v, w).$$

For the right-hand inequality, observe that if two vertices v and w are connected by a path of edge length $d_{X'}(v, w)$, then as the image of each edge in X has length at most 100δ, this shows that $d_X(v, w) \leqslant 100\delta d_{X'}(v, w)$.

For the left-hand inequality, let v and w be any two vertices in X'. Then there are two geodesics $[ax, bx]$ and $[cx, dx]$, connecting orbit points of x, such that $v \in [ax, bx]$ and $w \in [cx, dx]$. By thin triangles, any geodesic $[v, w]$ in X is contained in a 2δ-neighbourhood of the union of the geodesics $[ax, bx], [cx, dx], [ax, cx]$ and $[bx, dx]$. In particular, each point $p \in [v, w]$ lies within distance 12δ of a point $p' \in Y$. Let $\{p_i\}$ be a maximal collection of points on $[v, w]$ all at least 10δ apart, with an order inherited from an orientation on $[v, w]$. Let $\{p'_i\}$ be a corresponding collection of vertices p'_i in Y, with $d_X(p_i, p'_i) \leqslant 12\delta$ for all i. Each adjacent pair of points p'_i and p'_{i+1} is distance at most 44δ apart, and so is connected by an edge in X'. The collection p_i contains at most $d_X(v, w)/10\delta$ points, and so the distance in X' between v and w is at most $d_X(v, w)/10\delta$.

Finally, we observe that X' is Gromov hyperbolic, as $d_{X'}$ is quasi-isometric to the restriction of the Gromov hyperbolic metric d_X on the image of the vertices of X' in X. A quasi-isometric embedding then induces an inclusion map on the Gromov boundaries. $\square$

By the theorem in the separable case, given $x' \in X'$ almost every sample path $(w_n x')$ converges to a point $\xi' \in \partial X'$, hence if $x = i(x')$ then almost every sample path $(w_n x)$ converges to $i(\xi') \in \partial X$, hence the random walk on X converges almost surely to the boundary.

Another ingredient in the proof of the previous theorem is the following lemma about the measure of shadows [MT16, Proposition 5.4], which we will use in the later sections.

Proposition 2.7. *Let G be a non-elementary, countable group acting by isometries on a Gromov hyperbolic space X, and let μ be a non-elementary probability distribution on G. Then there is a number R_0 such that if $g, h \in G$ are group elements such that h and $h^{-1}g$ lie in the semigroup generated by the support of μ, then*

$$\nu(\overline{S_{hx}(gx, R_0)}) > 0,$$

where $\overline{A}$ denotes the closure in $X \cup \partial X$.

We will also use the well-known fact that in a Gromov hyperbolic space the complement of a shadow is approximately a shadow, as in the following proposition. We omit the proof, but we draw the appropriate approximate tree in Figure 1 below.

Proposition 2.8. *Given non-negative constants δ, K and L, there are constants C and D, such that in any δ-hyperbolic space X we have:*

(1) for any pair of points x, y in X and any $R \geqslant 0$ we have

$$X \setminus S_x(y, R) \subseteq S_y(x, d(x, y) - R + C)$$

(2) for any $R \geqslant 0$, and any bi-infinite (K, L)-quasigeodesic γ, parameterized such that $\gamma(0)$ is the closest point on γ to the basepoint x, then for any shadow set $V = S_x(\gamma(t), R)$ which does not contain x, with $t \geqslant 0$, and for any point $y \in U = S_x(\gamma(t + D), R)$, we have the inclusion

$$X \setminus V \subseteq S_y(x, d(x, \gamma(t)) - R + C).$$

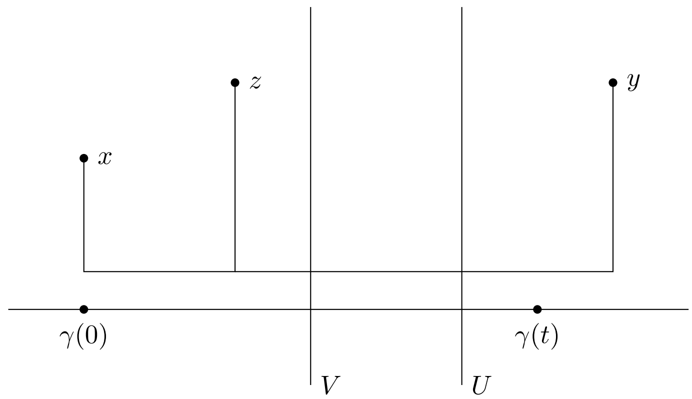

FIGURE 1. The complement of a shadow is contained in a shadow.

We will also use the following exponential decay estimates. For $Y \subset X$ let $H^+(Y)$ denote the probability that the random walk ever hits Y, i.e. that there is at least one index $n \in \mathbb{N}$ such that $w_n x \in Y$.

Lemma 2.9 (Exponential decay of shadows, [Mah12, Lemma 2.10])**.** *Let G be a group which acts by isometries on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, bounded probability measure on G. Then there exist constants $B > 0$ and $c < 1$ such that for any shadow $S_x(y, R)$ with distance parameter $r = d(x, y) - R$, we have the estimates*

$$\nu(S_x(y, R)) \leqslant Bc^r, \tag{3}$$

and

$$H^+(S_x(y, R)) \leqslant Bc^r. \tag{4}$$

In particular, for all n:

$$\mathbb{P}(w_n x \in S_x(y, R)) \leqslant Bc^r. \tag{5}$$

Indeed, equation (3) is [Mah10b, Lemma 5.4], and equation (4) follows from (3) as in [MT16, Equation (5.3)]. Equation (5) is an immediate consequence of (4).

Finally, we will also use the following positive drift, or linear progress, result.

Proposition 2.10 (Exponential decay of linear progress, [Mah12])**.** *Let μ be a countable, non-elementary measure on G which has bounded support in X. Then there exist constants $B > 0$, $L > 0$ and $0 < c < 1$ such that for all n:*

$$\mathbb{P}(d(x, w_n x) \leqslant Ln) \leqslant Bc^n.$$

2.3. **The Poisson boundary.** Given a countable group G and a probability measure μ on G, one defines the space of bounded μ-harmonic functions as

$$H^\infty(G, \mu) := \left\{ f : G \to \mathbb{R} \text{ bounded } \quad : \quad f(g) = \sum_{h \in G} f(gh)\mu(h) \ \forall g \in G \right\}.$$

Suppose now that G acts by homeomorphisms on a measure space (M, ν). Then the measure ν is μ-stationary if

$$\nu = \sum_{h \in G} \mu(h) \ h_\star \nu.$$

A G-space M with a μ-stationary measure ν is called a *μ-boundary* if for almost every sample path (w_n) the measure $w_n\nu$ converges to a δ-measure. Given a μ-boundary, one has the *Poisson transform* $\Phi : L^\infty(M,\nu) \to H^\infty(G,\mu)$ defined as

$$\Phi(f)(g) := \int_M f(gx)\ d\nu(x).$$

Definition 2.11. The space (M,ν) is the *Furstenberg-Poisson boundary* of (G,μ) if the Poisson transform Φ is a bijection between $L^\infty(M,\nu)$ and $H^\infty(G,\mu)$.

It turns out that the Furstenberg-Poisson boundary is well-defined up to G-equivariant measurable isomorphisms. Moreover, it is the maximal μ-boundary in following sense: if (B_{FP},ν_{FP}) is the Furstenberg-Poisson boundary and (B,ν) is another μ-boundary, then there exists a G-equivariant measurable map $(B_{FP},\nu_{FP}) \to (B,\nu)$. Finally, such a boundary can be defined as the measurable quotient of the sample space of the random walk (G,μ) by identifying two sample paths if they eventually coincide (to be precise, one should cast this definition in the context of measurable partitions, as defined by Rokhlin [Roh52]).

2.4. **The strip criterion.** In order to obtain the Poisson boundary for WPD actions, we will use Kaimanovich's *strip criterion.* This basically says that if bi-infinite paths for the random walks can be approximated by subsets of G, called *strips*, then one can conclude that the relative entropies of the conditional random walks vanish, hence the proposed geometric boundary is indeed the Poisson boundary.

Given a measure μ on G, its *reflected measure* if $\check{\mu}(g) := \mu(g^{-1})$. Moreover, we denote as $\check{\nu}$ the hitting measure for the random walk associated to the reflected measure $\check{\mu}$. We say that the measure μ has *finite entropy* if

$$H(\mu) := -\sum_{g\in G} \mu(g)\log\mu(g) < \infty.$$

Finally, it has *finite logarithmic moment* if $\int_G \log^+ d(x,gx)\ d\mu(g) < \infty$. Let us denote as

$$B_G(g) := \{h \in G\ :\ d(x,hx) \leqslant d(x,gx)\}.$$

We shall use the following *strip criterion* by Kaimanovich.

Theorem 2.12 ([Kai00])**.** *Let μ be a probability measure with finite entropy on G, and let $(\partial X,\nu)$ and $(\partial X,\check{\nu})$ be μ- and $\check{\mu}$-boundaries, respectively. If there exists a measurable G-equivariant map S assigning to almost every pair of points $(\alpha,\beta) \in \partial X \times \partial X$ a non-empty "strip" $S(\alpha,\beta) \subset G$, such that for all g*

$$\frac{1}{n}\log|S(\alpha,\beta)g \cap B_G(w_n)| \to 0 \qquad \text{as } n\to\infty,$$

for $\nu\times\check{\nu}$-almost every $(\alpha,\beta)\in\partial X\times\partial X$, then $(\partial X,\nu)$ and $(\partial X,\check{\nu})$ are the Poisson boundaries of the random walks (G,μ) and $(G,\check{\mu})$, respectively.

3. Background on the Cremona group

We will start by recalling some fundamental facts about the Cremona group, and especially its action on the Picard-Manin space. For more details, see [CL13], [DF01], [Fav08] and references therein.

3.1. **The Picard-Manin space.** If X is a smooth, projective, rational surface the group

$$N^1(X) := H^2(X, \mathbb{Z}) \cap H^{1,1}(X, \mathbb{R})$$

is called the *Néron-Severi group*. Its elements are Cartier divisors on X modulo numerical equivalence. The intersection form defines an integral quadratic form on $N^1(X)$. We denote $N^1(X)_\mathbb{R} := N^1(X) \otimes \mathbb{R}$.

If $f : X \to Y$ is a birational morphism, then the pullback map $f^\star : N^1(Y) \to N^1(X)$ is injective and preserves the intersection form, so $N^1(Y)_\mathbb{R}$ can be thought of as a subspace of $N^1(X)_\mathbb{R}$.

A *model* for $\mathbb{P}^2(\mathbb{C})$ is a smooth projective surface X with a birational morphism $X \to \mathbb{P}^2(\mathbb{C})$. We say that a model $\pi' : X' \to \mathbb{P}^2(\mathbb{C})$ dominates the model $\pi : X \to \mathbb{P}^2(\mathbb{C})$ if the induced birational map $\pi^{-1} \circ \pi' : X' \dashrightarrow X$ is a morphism. By considering the set $\mathcal{B}_X$ of all models which dominate X, one defines the space of *finite Picard-Manin classes* as the injective limit

$$\mathcal{Z}(X) := \lim_{X' \in B_X} N^1(X')_\mathbb{R}.$$

In order to find a basis for $\mathcal{Z}(X)$, one defines an equivalence relation on the set of pairs (p, Y) where Y is a model of X and p a point in Y, as follows. One declares $(p, Y) \sim (p', Y')$ if the induced birational map $Y \dashrightarrow Y'$ maps p to p' and is an isomorphism in a neighbourhood of p. We denote the quotient space as $\mathcal{V}_X$. Finally, the *Picard-Manin space* of X is the L^2-completion

$$\mathcal{Z}(X) := \left\{ [D] + \sum_{p \in \mathcal{V}_X} a_p [E_p] \ : \ [D] \in N^1(X)_\mathbb{R}, a_p \in \mathbb{R}, \sum_{p \in \mathcal{V}_X} a_p^2 < +\infty \right\}.$$

In this paper, we will only focus on the case $X = \mathbb{P}^2(\mathbb{C})$. Then the Néron-Severi group of $\mathbb{P}^2(\mathbb{C})$ is generated by the class $[H]$ of a line, with self-intersection $+1$. Thus, the Picard-Manin space is

$$\overline{\mathcal{Z}}(\mathbb{P}^2) := \left\{ a_0[H] + \sum_{p \in \mathcal{V}_{\mathbb{P}^2(\mathbb{C})}} a_p [E_p], \quad \sum_p a_p^2 < +\infty \right\}.$$

It is well-known that if one blows up a point in the plane, then the corresponding exceptional divisor has self-intersection -1, and intersection zero with divisors on the original surface.

Thus, the classes $[E_p]$ have self-intersection -1, are mutually orthogonal, and are orthogonal to $N^1(X)$. Hence, the space $\overline{\mathcal{Z}}(\mathbb{P}^2)$ is naturally equipped with a quadratic form of signature $(1, \infty)$, thus making it a Minkowski space of uncountably infinite dimension. Thus, just as classical hyperbolic space can be realized as one sheet of a hyperboloid inside a Minkowski space, inside the Picard-Manin space one defines

$$\mathbb{H}_{\mathbb{P}^2} := \{[D] \in \overline{\mathcal{Z}}(\mathbb{P}^2) \ : \ [D]^2 = 1, [H] \cdot [D] > 0\}$$

which is one sheet of a two-sheeted hyperboloid. The restriction of the quadratic intersection form to $\mathbb{H}_{\mathbb{P}^2}$ defines a Riemannian metric of constant curvature -1, thus making $\mathbb{H}_{\mathbb{P}^2}$ into an infinite-dimensional hyperbolic space. More precisely, the induced distance dist satisfies the formula

$$\cosh \operatorname{dist}([D_1], [D_2]) = [D_1] \cdot [D_2].$$

Each birational map f acts on $\overline{\mathcal{Z}}$ by orthogonal transformations. To define the action, recall that for any rational map $f : \mathbb{P}^2(\mathbb{C}) \dashrightarrow \mathbb{P}^2(\mathbb{C})$ there exist a surface X and morphisms $\pi, \sigma : X \to \mathbb{P}^2(\mathbb{C})$ such that $f = \sigma \circ \pi^{-1}$. Then we define $f^\star = (\pi^\star)^{-1} \circ \sigma^\star$, and $f_\star = (f^{-1})^\star$. Moreover, $f_\star$ preserves the intersection form, hence it acts as an isometry of $\mathbb{H}_{\mathbb{P}^2}$: in other words, the map $f \mapsto f_\star$ is a group homomorphism

$$\mathrm{Bir}\ \mathbb{P}^2(\mathbb{C}) \to \mathrm{Isom}(\mathbb{H}_{\mathbb{P}^2})$$

hence one can apply to the Cremona group the theory of random walks on groups acting on non-proper δ-hyperbolic spaces.

The space $\mathbb{H}_{\mathbb{P}^2}$ is not separable; however, any countable subgroup of the Cremona group preserves a closed, totally geodesic, separable, subset of $\mathbb{H}_{\mathbb{P}^2}$ (see also [DP12], Remark 1).

Definition 3.1. The *dynamical degree* of a birational transformation $f : X \dashrightarrow X$ is defined as

$$\lambda(f) := \lim_{n\to\infty} \|(f^n)^\star\|^{1/n}$$

where $\|\cdot\|$ is any operator norm on the space of endomorphisms of $H^\star(X, \mathbb{R})$.

Note that $\lambda(f) = \lambda(gfg^{-1})$ is invariant by conjugacy. Moreover, if f is represented by three homogeneous polynomials of degree d without common factors, then the action of $f^\star$ on the class $[H]$ of a line is $f^\star([H]) = d[H]$, hence

$$\lambda(f) = \lim_{n\to\infty} \deg(f^n)^{1/n}.$$

Moreover, the degree is related to the displacement in the hyperbolic space $\mathbb{H}_{\mathbb{P}^2}$: in fact, (see [Fav08], page 17)

$$\deg(f) = f^\star[H] \cdot [H] = [H] \cdot f_\star[H] = \cosh d(x, fx)$$

if $x = [H] \in \mathbb{H}_{\mathbb{P}^2}$. As a consequence, the dynamical degree $\lambda(f)$ of a transformation f is related to its translation length $\tau(f)$ by the equation ([CL13], Remark 4.5):

$$\tau(f) = \lim_{n\to\infty} \frac{\mathrm{dist}(x, f^n x)}{n} = \lim_{n\to\infty} \frac{\cosh^{-1} \deg(f^n)}{n} = \log \lambda(f).$$

Hence, a Cremona transformation f is loxodromic if and only if $\lambda(f) > 1$.

4. Growth of translation length

Let us now start by proving that for bounded probability measures translation length grows linearly along almost every sample path.

Theorem 4.1. *Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable non-elementary measure on G whose support is bounded in X. Then for almost every sample path we have*

$$\lim_{n\to\infty} \frac{\tau(w_n)}{n} = L$$

where $L > 0$ is the drift of the random walk.

This will follow using the following result, that the size of the Gromov products $(w_n x \cdot w_n^{-1} x)_x$ grows sublinearly.

Proposition 4.2. *Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable non-elementary measure on G whose support is bounded in X. Then for almost every sample path we have*

$$\lim_{n\to\infty} \frac{(w_n x \cdot w_n^{-1} x)_x}{n} = 0.$$

We now prove Theorem 4.1, assuming Proposition 4.2.

Proof of Theorem 4.1. Since the support is bounded in X, by Theorem 2.5 there exists $L > 0$ such that almost surely

$$\lim_{n\to\infty} \frac{d(x, w_n x)}{n} = L.$$

Moreover, by Proposition 4.2

$$\lim_{n\to\infty} \frac{(w_n x \cdot w_n^{-1} x)_x}{n} = 0.$$

The claim then follows by using the well-known formula (see [MT16], Appendix A)

$$\tau(g) = d(x, gx) - 2(gx \cdot g^{-1}x)_x + O(\delta).$$

□

Finally, we prove Proposition 4.2.

Proof of Proposition 4.2. In fact, we will show the following exponential decay statement: there exist $B > 0$ and $0 < c < 1$ such that for any $\epsilon > 0$ we have

$$\mathbb{P}((w_n x \cdot w_n^{-1} x)_x \geqslant \epsilon n) \leqslant B c^{\epsilon n}. \tag{6}$$

The statement of the Proposition then follows by Borel-Cantelli. The proof of (6) will be achieved in the following steps. First of all, for the sake of simplicity let us replace n by $2n$. Moreover, let us define the random variable $u_n := w_n^{-1} w_{2n}$, so that $w_{2n} = w_n u_n$. Note that by the Markov property w_n and u_n are independent, and the distribution of u_n^{-1} equals the distribution of the n^{th} step of the reflected random walk.

(1) For any $n \geqslant 0$ and any $\epsilon > 0$ we have

$$\mathbb{P}((w_n x \cdot u_n^{-1} x)_x \geqslant \epsilon n) \leqslant B c^{\epsilon n}$$

for some constants $B > 0$, $c < 1$.

Proof. By definition, the condition $(w_n x \cdot u_n^{-1} x)_x \geqslant \epsilon n$ is equivalent to $w_n x \in S_x(u_n^{-1} x, d(x, u_n^{-1} x) - \epsilon n)$. Hence

$$\mathbb{P}((w_n x \cdot u_n^{-1} x)_x \geqslant \epsilon n) = \sum_g \mathbb{P}\left(w_n x \in S_x(u_n^{-1} x, d(x, u_n^{-1} x) - \epsilon n) \mid u_n = g\right) \mathbb{P}(u_n = g)$$

and by independence

$$\mathbb{P}((w_n x \cdot u_n^{-1} x)_x \geqslant \epsilon) = \sum_g \mathbb{P}(w_n x \in S_x(g^{-1} x, d(x, g^{-1} x) - \epsilon n)) \mathbb{P}(u_n = g)$$

hence by exponential decay of shadows, Lemma 2.9,

$$\mathbb{P}((w_n x \cdot u_n^{-1} x)_x \geqslant \epsilon) \leqslant B c^{\epsilon n},$$

as required. □

(2) For any $n \geqslant 0$ we have

$$\mathbb{P}((w_{2n}x \cdot w_n x)_x \leqslant (L-\epsilon)n) \leqslant Bc^{\epsilon n}.$$

Proof. By definition of the Gromov product, if

$$(w_{2n}x \cdot w_n x)_x \leqslant (L-\epsilon)n$$

then

$$w_{2n}x \notin S_x(w_n x, d(x, w_n x) - (L-\epsilon)n).$$

Since $w_{2n} = w_n u_n$ and w_n acts by isometries, this implies that

$$u_n x \notin S_{w_n^{-1}x}(x, d(x, w_n x) - (L-\epsilon)n).$$

As the complement of a shadow is contained in a shadow (Proposition 2.8),

$$u_n x \in S_x(w_n^{-1}x, (L-\epsilon)n + C), \tag{7}$$

where C depends only on δ. Now, if $d(x, w_n x) \geqslant Ln$ the distance parameter of the above shadow is $r = d(x, w_n^{-1}x) - (L-\epsilon)n - C \geqslant \epsilon n - C$, hence by decay of shadows (Lemma 2.9) and decay of linear progress (Lemma 2.10), the probability that (7) holds is at most $Bc^{\epsilon n - C}$. Hence the claim holds after replacing B with a slightly larger B. □

(3) For any $n \geqslant 0$ we have

$$\mathbb{P}((w_{2n}^{-1}x \cdot u_n^{-1}x)_x \leqslant (L-\epsilon)n) \leqslant Bc^{\epsilon n}.$$

The proof is exactly as the previous one so we will not write it in detail.

(4) Finally, we will use the following lemma in hyperbolic geometry (for a proof, see [TT16]):

Lemma 4.3. *Let X be a δ-hyperbolic space, $x \in X$ a base point, and fix $A \geqslant 0$. Then for any four points $a, b, c, d \in X$ which satisfy $(a \cdot b)_x \geqslant A$, $(c \cdot d)_x \geqslant A$, and $(a \cdot c)_x \leqslant A - 3\delta$, one has $(b \cdot d)_x \leqslant (a \cdot c)_x - 2\delta \leqslant A - 5\delta$.*

By applying this lemma to $A = \epsilon n$, $a = w_n x$, $b = w_{2n}x$, $c = u_n^{-1}x$, and $d = w_{2n}^{-1}x$ we complete the proof of equation (6).

□

5. WPD actions

5.1. **The WPD condition.** Let G be a group acting by isometries on a metric space X. Recall that the action of G on X is *proper* if the map $G \times X \to X \times X$ given by $(g, x) \mapsto (x, gx)$ is proper, i.e. the preimages of compact sets are compact. A related notion is that the action is *properly discontinuous* if for every $x \in X$ there exists an open neighbourhood U of x such that $gU \cap U \neq \emptyset$ holds for at most finitely many elements g. If the space X is not proper, it is very restrictive to ask for the action to be proper (for instance, point stabilizers for a proper action must be finite). However, Bestvina-Fujiwara [BF02] defined the notion of *weak proper discontinuity*, or *WPD*; essentially, a loxodromic isometry g is a WPD element if its action is proper in the direction of its axis.

Definition 5.1. Let G be a group acting on a hyperbolic space X, and h a loxodromic element of G. One says that h satisfies the *weak proper discontinuity condition* (or h is a WPD element) if for every $K > 0$ and every $x \in X$, there exists $M \in \mathbb{N}$ such that

$$\#|\{g \in G \ : \ d(x, gx) < K, d(h^M x, gh^M x) < K\}| < \infty.$$

If we define the *joint coarse stabilizer* of two points $x, y \in X$ as

$$\mathrm{Stab}_K(x, y) := \{g \in G \ : \ d(x, gx) \leqslant K \text{ and } d(y, gy) \leqslant K\}$$

then the WPD condition says that for any K and any x there exists an integer M such that $\mathrm{Stab}_K(x, h^M x)$ is a finite set. A trivial consequence of the definition of WPD is the following.

Lemma 5.2. *Let G be a group acting on a Gromov hyperbolic space X, and let h be a WPD element in G. Then there are functions $M_W : \mathbb{R}_{\geqslant 0} \to \mathbb{N}$ and $N_W : \mathbb{R}_{\geqslant 0} \to \mathbb{N}$ such that for any $x \in X$, any $K \geqslant 0$, and for any $f \in G$ one has*

$$\#|Stab_K(fx, fh^{M_W(K)}x)| \leqslant N_W(K).$$

Proof. By definition, note that

$$\mathrm{Stab}_K(fx, fy) = f\mathrm{Stab}_K(x, y)f^{-1}$$

hence the cardinality

$$\#|\mathrm{Stab}_K(fx, fh^M x)| = \#|f(\mathrm{Stab}_K(x, h^M x))f^{-1}| = \#|\mathrm{Stab}_K(x, h^M x)|$$

is finite and independent of f, proving the claim. □

6. The Poisson boundary

Let us now use the WPD property to prove that the Poisson boundary coincides with the Gromov boundary, proving Theorem 1.6 in the Introduction. The idea is to define appropriately the strips for Kaimanovich's criterion using "elements of bounded geometry" as below, and using the WPD condition to show that the number of elements in such strips grows at most linearly.

Choose K large enough, let us fix a base point $x \in X$, and take $v = h^M$ to be a sufficiently high power of h so that the WPD condition holds with constant $22K$.

6.1. Elements of bounded geometry.

For any pair $(\alpha, \beta) \in \partial X \times \partial X$, with $\alpha \neq \beta$, define the set of *bounded geometry elements* as

$$\mathcal{O}(\alpha, \beta) := \{g \in G \ : \ \alpha \in \overline{S_{gvx}(gx, K)} \text{ and } \beta \in \overline{S_{gx}(gvx, K)}\}.$$

An example of a bounded geometry element is illustrated below in Figure 2. Note that for any $g \in G$ we have $\mathcal{O}(g\alpha, g\beta) = g\mathcal{O}(\alpha, \beta)$. Moreover, we define the ball in the group with respect to the metric on X as

$$B(y, r) := \{g \in G \ : \ d(y, gx) \leqslant r\}$$

where $y \in X$ and $r \geqslant 0$.

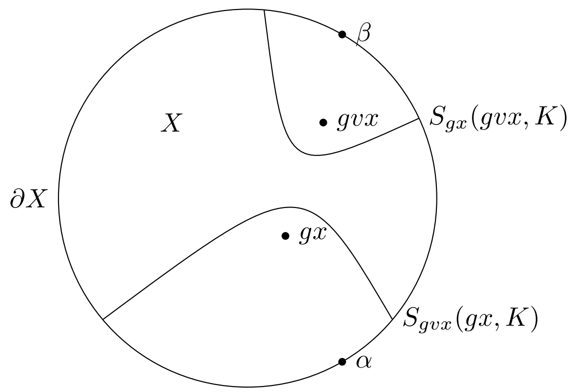

FIGURE 2. A bounded geometry element g in $\mathcal{O}(\alpha, \beta)$.

The most crucial property of bounded geometry elements is that their number in a ball grows linearly with the radius of the ball.

Proposition 6.1. *There exists a constant C such that for any radius $r > 0$ and any pair of distinct boundary points $\alpha, \beta \in \partial X$ one has*

$$\#|B(x, r) \cap \mathcal{O}(\alpha, \beta)| \leqslant Cr.$$

This fact follows from the next lemma, which uses the WPD property in a crucial way.

Lemma 6.2. *For any $K \geqslant 0$ there exists a constant N such that*

$$\#|B(z, 4K) \cap \mathcal{O}(\alpha, \beta)| \leqslant N$$

for any $z \in X$ and any pair of distinct boundary points α, β.

Proof. Let us consider two elements g, h which belong to $\mathcal{O}(\alpha, \beta) \cap B(z, 4K)$. Then if we let $f = hg^{-1}$, then

$$d(gx, fgx) \leqslant 8K. \tag{8}$$

Let γ be a quasigeodesic which joins α and β, and denote $S_1 := \overline{S_{gvx}(gx, K)}$, $S_2 := \overline{S_{gx}(gvx, K)}$. By construction, α belongs to both S_1 and fS_1 hence both α and $f\alpha$ belong to fS_1; similarly, β and $f\beta$ belong to fS_2. Hence, the two quasigeodesics γ and $f\gamma$ have endpoints in fS_1 and fS_2, hence they must fellow travel in their middle: more precisely, they must pass within distance $2K$ from both fgx and $y := fgvx$. Hence, if we call q a closest point to $f\gamma$ to fgx, we have $d(fgx, q) \leqslant 2K$. Moreover, if we call p a closest point on γ to y, and p' a closest point on $f\gamma$ to y, we have

$$d(p, p') \leqslant d(p, y) + d(y, p') \leqslant 4K$$

Combining this with eq. (8) we get

$$|d(gx, p) - d(fgx, p')| \leqslant 12K$$

Moreover, since f is an isometry we have $d(fgx, fp) = d(gx, p)$, hence

$$|d(fgx, fp) - d(fgx, p')| \leqslant 12K \tag{9}$$

Now, the points q, p' and fp both lie on the quasigeodesic $f\gamma$; let us assume that fp lies in between q and p', and draw a geodesic segment γ' between q and p', and let p'' be a closest point projection of fp to γ' (the case where p' lies between q and fp is completely analogous). By fellow traveling, we have $d(fp, p'') \leqslant L$. Then, since p', p'' and q lie on a geodesic, we have

$$d(p', p'') = |d(q, p') - d(q, p'')| \leqslant$$

and by using eq. (9)

$$\leqslant |d(fgx, p') - d(fgx, fp)| + d(fgx, q) + d(fgx, q) + d(fp, p'') \leqslant 12K + 2K + 2K + L$$

hence

$$d(fp, p') \leqslant 16K + 2L$$

and finally

$$d(y, fy) \leqslant d(y, p') + d(p', fp) + d(fp, fy) \leqslant 20K + 2L$$

Thus, if we choose K large enough so that $L \leqslant K$ we have $d(gvx, fgvx) = d(fgvx, f^2gvx) \leqslant 22K$ hence

$$f \in \mathrm{Stab}_{22K}(gx, gvx)$$

so by Lemma 5.2 there are only N possible choices of f, as claimed. $\square$

Proof of Proposition 6.1. Let γ be a quasi-geodesic in X which joins α and β. By definition, if g belongs to $\mathcal{O}(\alpha, \beta)$, then gx lies within distance $\leqslant 2K$ of γ. Then one can pick points $(z_n)_{n \in \mathbb{Z})}$ along γ such that any point of γ is within distance $\leqslant 2K$ of some z_n. Then, any ball of radius r contains at most cr of such z_n, where c depends only on K and the quasigeodesic constant of γ. The claim then follows from Lemma 6.2. $\square$

We now turn to the proof of Theorem 1.6. By Theorem 2.5, we know that since both μ and its reflected measure $\check{\mu}$ are non-elementary, both the forward random walk and the backward random walk converge almost surely to points on the boundary of X. Thus, one defines the two boundary maps $\partial_\pm : (G^{\mathbb{Z}}, \mu^{\mathbb{Z}}) \to \partial X$ as follows. Let $\omega = (g_n)_{n \in \mathbb{Z}}$ be a bi-infinite sequence of increments, and define

$$\partial_+(\omega) := \lim_{n \to \infty} g_1 \dots g_n x, \qquad \partial_-(\omega) := \lim_{n \to \infty} g_0^{-1} g_{-1}^{-1} \dots g_{-n}^{-1} x$$

the two endpoints of, respectively, the forward random walk and the backward random walk. Then define

$$\mathcal{O}(\omega) := \mathcal{O}(\partial_+(\omega), \partial_-(\omega))$$

the set of bounded geometry elements along the "geodesic" which joins $\partial_+(\omega)$ and $\partial_-(\omega)$. Note that, if $T : G^{\mathbb{Z}} \to G^{\mathbb{Z}}$ is the shift in the space of increments, we have

$$\mathcal{O}(T^n \omega) = \mathcal{O}(w_n^{-1} \partial_+, w_n^{-1} \partial_-) = w_n^{-1} \mathcal{O}(\omega).$$

Now we will show that for almost every bi-infinite sample path ω the set $\mathcal{O}(\omega)$ is non-empty and has at most linear growth. In fact, by definition of bounded geometry

$$\mathbb{P}(1 \in \mathcal{O}(\omega)) = p = \nu(\overline{S}) \check{\nu}(\overline{S'}) > 0$$

where $S = S_{vx}(x, K)$ and $S' = S_x(vx, K)$, and their measures are positive by Proposition 2.7. Moreover, since the shift map T preserves the measure in the space of increments, we also have for any n

$$\mathbb{P}(w_n \in \mathcal{O}(\omega)) = \mathbb{P}(1 \in \mathcal{O}(T^n \omega)) = p > 0.$$

Thus, by the ergodic theorem, the number of times w_n belongs to $\mathcal{O}(\omega)$ grows almost surely linearly with n: namely, for a.e. ω

$$\lim_{n\to\infty} \frac{\#|\{1 \leqslant i \leqslant n \ : \ w_i \in \mathcal{O}(\omega)\}|}{n} = p > 0.$$

Hence the set $\mathcal{O}(\omega)$ is almost surely non-empty (in fact, it contains infinitely many elements). On the other hand, by Proposition 6.1 the set $\mathcal{O}(\omega)$ has at most linear growth, i.e. there exists $C > 0$ such that

$$\#|\mathcal{O}(\omega) \cap B_G(z, r)| \leqslant Cr \qquad \forall r > 0. \tag{10}$$

The Poisson boundary result now follows from the strip criterion (Theorem 2.12). Let $P(G)$ denote the set of subsets of G. Then, we define the strip map $S : \partial X \times \partial X \to P(G)$ as $S(\alpha, \beta) := \mathcal{O}(\alpha, \beta)$; hence, by equation (10)

$$\#|S(\alpha, \beta)g \cap B_G(w_n)| \leqslant Cd(w_n x, x).$$

Then, since μ has finite logarithmic moment, one has almost surely

$$\lim_{n\to\infty} \frac{1}{n} \log d(w_n x, x) \to 0$$

which verifies the criterion of Theorem 2.12, establishing that the Gromov boundary of X is a model for the Poisson boundary of the random walk.

7. Asymptotic acylindricality

We say that a group G acting by isometries on a Gromov hyperbolic space is *acylindrical* if for all $K \geqslant 0$, there are constants $R \geqslant 0$ and $N \geqslant 0$, such that for all points x and y in X, with $d(x, y) \geqslant R$, one has the bound

$$\#|\mathrm{Stab}_K(x) \cap \mathrm{Stab}_K(y)| \leqslant N.$$

Let μ be a probability measure on a group G acting by isometries on a Gromov hyperbolic space.

Definition 7.1. We say that the random walk generated by μ is *asymptotically acylindrical* if there is a function $N_{ac} \colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ such that for all $K \geqslant 0$, the probability that

$$\#|\mathrm{Stab}_K(x) \cap \mathrm{Stab}_K(w_n x)| \leqslant N_{ac}(K)$$

tends to one as n tends to infinity.

In the following, we will actually need an explicit rate of convergence to one in the previous definition. Recall that we say that the random walk generated by μ is asymptotically acylindrical with *square root exponential decay* if for any $K \geqslant 0$ there exist constants $N > 0$, $B > 0$, $c < 1$ such that

$$\mathbb{P}(\#|\mathrm{Stab}_K(x) \cap \mathrm{Stab}_K(w_n x)| \leqslant N) \geqslant 1 - Bc^{\sqrt{n}}.$$

We now show that if Γ_μ contains a WPD element, then the random walk determined by μ is asymptotically acylindrical with square root exponential decay, which is Theorem 1.2 in the Introduction.

Theorem 7.2. *Let G be a group acting by isometries on a Gromov hyperbolic space X, let $x \in X$, and let μ be countable, non-elementary, bounded, WPD probability distribution on G. Then for any $K \geqslant 0$, there are constants $N > 0$, $B > 0$ and $c < 1$ such that*

$$\mathbb{P}\left(\#|\mathrm{Stab}_K(x, w_n x)| \leqslant N\right) \geqslant 1 - Bc^{\sqrt{n}}$$

and

$$\mathbb{P}\left(\#|\mathrm{Stab}_K(p, w_n p)| \leqslant N\right) \geqslant 1 - Bc^{\sqrt{n}}$$

where p is a closest point projection of x onto an axis α_{w_n} of w_n.

Let (Y, ν) be a probability space, and let $T : Y \to Y$ be a measure-preserving, ergodic map. Then for any set A of positive measure $\nu(A) > 0$, the ν-measure of the union of the iterates $A \cup T^{-1}A \cup \cdots T^{-n}A$ tends to one as n tends to infinity. The proof of the Theorem 7.2 depends crucially on the following effective estimate for the rate at which this measure tends to one, Proposition 7.3. Such a result does not hold in general for arbitrary measurable sets, but we show it holds in our context in which $Y = \partial X \times \partial X$, and A contains an open set in Y containing limit points of $\Gamma_\mu \times \Gamma_{\check{\mu}}$.

We shall write $(G, \mu)^{\mathbb{Z}}$ for the step space, and g_n for projection onto the n-th factor. We shall write w_n for the location of the random walk at time n, i.e. $w_0 = 1$ and $w_{n+1} = w_n g_{n+1}$. Let T be the shift map on $(G, \mu)^{\mathbb{Z}}$, i.e. $g_n(T(\omega)) = g_{n+1}(\omega)$, and let $\partial \colon (G, \mu)^{\mathbb{Z}} \to \overline{X} \times \overline{X}$ be the boundary map $\partial = \partial_+ \times \partial_-$. We shall write $\Lambda_\mu := \overline{\Gamma_\mu x} \cap \partial X$ for the limit set of Γ_μ in the Gromov boundary ∂X, which does not depend on the choice of $x \in X$.

The shift map $T \colon (G, \mu)^{\mathbb{Z}} \to (G, \mu)^{\mathbb{Z}}$ is ergodic, and so if a set $A \subset G^{\mathbb{Z}}$ has positive measure, then the measure of the union of images $A \cup T^{-1}A \cup \cdots T^{-n}A$ tends to one as n tends to infinity. We now obtain a rate of convergence, for sets A whose images under the boundary map contain an open neighbourhood of a pair of distinct points in the forward and backward limit sets of the random walk.

Proposition 7.3. *Let G be a group acting by isometries on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Let V_+ and V_- be a pair of sets in $X \cup \partial X$, such that there are a pair of distinct points $\lambda^+ \in \Lambda_\mu$ and $\lambda^- \in \Lambda_{\check{\mu}}$, for which $V_+ \times V_-$ contains an open neighbourhood of (λ^+, λ^-) in $\partial X \times \partial X$. Set $A = \{\omega \in (G, \mu)^{\mathbb{Z}} : \partial(\omega) \in V_+ \times V_-\}$. Then there are constants $B \geqslant 0$ and $c < 1$ such that*

$$\mu^{\mathbb{Z}}(A \cup T^{-1}A \cup \cdots \cup T^{-n}A) \geqslant 1 - Bc^{\sqrt{n}}.$$

Proof. Let γ be a $(1, K_1)$-quasigeodesic from λ^- to λ^+, and choose a unit speed parameterization of γ such that $\gamma(0)$ is a closest point on γ to the basepoint x, and furthermore $\lim_{t\to\infty} \gamma(t) = \lambda^+$.

Without loss of generality, we may replace V_+ and V_- by smaller sets, which are disjoint, and which are disjoint from the basepoint x, and such that $V_+ \times V_-$ still contains an open neighbourhood of (λ^+, λ^-) in $\partial X \times \partial X$. Furthermore, as shadow sets form a neighbourhood basis for the topology on ∂X, we may assume that V_+ and V_- are shadow sets $V_+ = S_x(\gamma(t), R_0)$ and $V_- = S_x(\gamma(-t), R_0)$, for some t sufficiently large, where R_0 is the constant from Proposition 2.7. Let D be the constant from Proposition 2.8, and let $U_+ = S_x(\gamma(t + D), R_0)$, and $U_- = S_x(\gamma(-t - D), R_0)$. This is illustrated below in Figure 3.

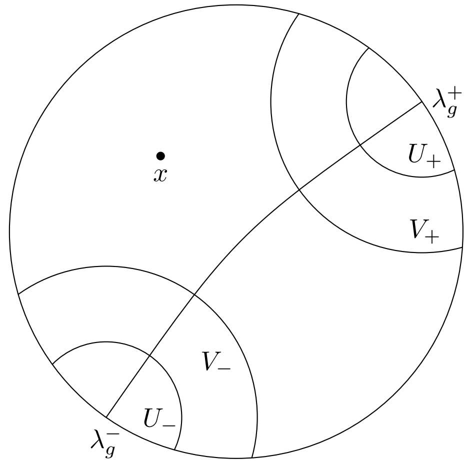

FIGURE 3. Nested shadows.

Let $A = \partial^{-1}(V_+ \times V_-)$. It suffices to show that

$$\mu^{\mathbb{Z}}(A \cup T^{-1}A \cup \cdots \cup T^{-n}A) \to 1,$$

or equivalently,

$$\mu^{\mathbb{Z}}(G^{\mathbb{Z}} \setminus (A \cup T^{-1}A \cup \cdots \cup T^{-n}A)) \to 0,$$

at the appropriate rate. If we write $A^c := G^{\mathbb{Z}} \setminus A$, we may rewrite this as

$$\mu^{\mathbb{Z}}(A^c \cap T^{-1}A^c \cap \cdots \cap T^{-n}A^c) \to 0.$$

For notational convenience, we will write A_n for $A^c \cap T^{-1}A^c \cap \cdots \cap T^{-n}A^c$.

It will be convenient to approximate A by sets of the following form. Let U_k be the set of sample paths for which $w_k x \in U_+$ and $w_{-k}x \in U_-$, i.e.

$$U_k = \{\omega \in G^{\mathbb{Z}} \ : \ (w_k(\omega)x, w_{-k}(\omega)x) \in U_+ \times U_-\},$$

and let V_k be the subset of U_k consisting of those sample paths for which $w_n(\omega) \in V_+$ for all $n \geqslant k$, and $w_{-n}(\omega) \in V_-$ for all $n \geqslant k$, i.e.

$$V_k = \left\{ \omega \in G^{\mathbb{Z}} \ \middle| \ \begin{array}{l} (w_n(\omega)x, w_{-n}(\omega)x) \in U_+ \times U_- \text{ for } n = k, \\ (w_n(\omega)x, w_{-n}(\omega)x) \in V_+ \times V_- \text{ for } n > k \end{array} \right\}.$$

In particular, $\partial(V_k) \subseteq V_+ \times V_-$, and so $V_k \subseteq A$. Therefore $A^c \subseteq V_k^c$, which immediately implies

$$A_n \subseteq V_k^c \cap T^{-1}V_k^c \cap \cdots \cap T^{-n}V_k^c = \bigcap_{i=0}^{n} T^{-i}V_k^c.$$

As we are taking intersections, we may choose a subcollection of the sets on the right. We shall choose numbers ℓ and k which grow at rate approximately $\sqrt{n}$. To be precise, choose k and ℓ to be the largest integers such that $2k+1 \leqslant \sqrt{n}$ and $\ell \leqslant \sqrt{n}$. Then $(2k+1)\ell \leqslant n$, so

$$A_n \subseteq \bigcap_{i=0}^{\ell-1} T^{-(2k+1)i}V_k^c.$$

As $V_k \subseteq U_k$, this implies that $V_k^c = U_k^c \cup (U_k \setminus V_k)$, which gives

$$A_n \subseteq \bigcap_{i=0}^{\ell-1} T^{-(2k+1)i}(U_k^c \cup (U_k \setminus V_k)).$$

By distributing out unions and intersections, this implies the inclusion

$$A_n \subseteq \left(\bigcap_{i=0}^{\ell-1} T^{-(2k+1)i} U_k^c \right) \cup \left(\bigcup_{i=0}^{\ell-1} T^{-(2k+1)i}(U_k \setminus V_k) \right).$$

The set U_k^c only depends on $g_{-k}, g_{-k+1}, \ldots g_k$, so the events $T^{-(2k+1)i}U_k^c$ are independent. Therefore

$$\mu^{\mathbb{Z}}(A_n) \leqslant \mu^{\mathbb{Z}}(U_k^c)^\ell + \ell \mu^{\mathbb{Z}}(V_k^c \setminus U_k^c).$$

By the definition of U_k, $\mu^{\mathbb{Z}}(U_k) = \mu_k(U_+)\check{\mu}_k(U_-)$. The convolution measures weakly converge to the hitting measures ([MT16], end of Section 4), hence by the portmanteau theorem $\liminf_k \mu_k(U_+) \geqslant \nu(U_+)$ and $\liminf_k \check{\mu}_k(U_-) \geqslant \check{\nu}(U_-)$. By [MT16, Proposition 5.4], $\nu(U_+) > 0$ and $\check{\nu}(U_-) > 0$, so there is a $c_1 < 1$ such that $\mu^{\mathbb{Z}}(U_k^c) < c_1 < 1$ for all k sufficiently large.

We now find an upper bound for $\mu^{\mathbb{Z}}(U_k \setminus V_k)$. Positive drift with exponential decay [MT16, Theorem 1.2] implies that there are constants $L > 0, B_2 \geqslant 0$ and $c_2 < 1$ such that

$$\mathbb{P}(d(x, w_k x) \leqslant Lk) \leqslant B_2 c_2^k,$$

and similarly, for the reflected measure

$$\mathbb{P}(d(x, w_{-k} x) \leqslant Lk) \leqslant B_2 c_2^k.$$

Therefore, we may assume that both $d(x, w_k x) \geqslant Lk$ and $d(x, w_{-k}x) \geqslant Lk$, for all but a set of sample paths in U_k of $\mu^{\mathbb{Z}}$-measure at most $2B_2c_2^k$. Then, by our choice of shadow sets, we may apply Proposition 2.8, which implies that $X \setminus V_+ \subseteq S_{w_k x}(x, R_0 + C)$ and $X \setminus V_- \subseteq S_{w_{-k}x}(x, R_0 + C)$, where C depends only on $x, \gamma(\pm t)$ and δ, as the quasigeodesic constants for a quasi-axis depend only on δ. By Lemma 2.9, the probability of ever hitting a shadow decays exponentially in the distance from the shadow, so there are constants $B_3 \geqslant 0$ and $c_3 < 1$ such that the probability that $w_k x$ lies in U_+ and $w_k x$ lies in $X \setminus V_+$ for some $n > k$ is at most $B_3 c_3^{Lk - R_0 - C}$, and similarly, the probability that $w_{-k}x$ lies in U_- and $w_{-n}x$ lies in $X \setminus V_-$ for some $n > k$ is at most $B_3 c_3^{Lk-R_0-C}$. Therefore, there are constants $B_4 \geqslant 0$ and $c_4 < 1$ such that

$$\mu^{\mathbb{Z}}(U_k \setminus V_k) \leqslant B_4 c_4^k.$$

Therefore

$$\mu^{\mathbb{Z}}(A_n) \leqslant c_1^\ell + \ell B_4 c_4^k,$$

and as we have chosen $2k + 1 \leqslant \sqrt{n}$ and $\ell \leqslant \sqrt{n}$, there are constants B_5 and c_5 such that

$$\mu^{\mathbb{Z}}(A^c \cap T^{-1}A^c \cap \cdots \cap T^{-n}A^c) \leqslant B_5 c_5^{\sqrt{n}},$$

as required. □

We will use the following definition from [CM15].

Definition 7.4. Let G be a group acting on a Gromov hyperbolic space X. We say that two geodesics γ and γ' in X have an $(L, K)-$*match* if there exist geodesic subsegments $\alpha \subseteq \gamma$ and $\alpha' \subseteq \gamma'$ of length $\geqslant L$ and some $g \in G$ such that $g\alpha$ and α' have Hausdorff distance $\leqslant K$.

If g is a loxodromic element, we denote by λ_g^+ and λ_g^-, respectively, the attracting and repelling fixed points of g on ∂X. We shall write γ_n for a geodesic in X from x to $w_n x$, and if g is a loxodromic isometry of X, we shall write α_g for an axis for g.

We will use the following result due to Dahmani and Horbez [DH15, Proposition 1.5]: they do not explicitly state the rate, but it follows immediately from the proof.

Proposition 7.5. *Let G be a group which acts on a δ-hyperbolic space X, and let $K \geqslant 0$. Let μ be a countable, non-elementary, bounded probability distribution on G, and let $\ell > 0$ be the drift of the random walk generated by μ. If w_n is loxodromic, let p denote a closest point projection of x to the axis of w_n. Then there exist constants $B > 0$, $c < 1$ such that for any $\epsilon > 0$ we have*

$$\mathbb{P}\left(\gamma_n \text{ has a } ((\ell - \epsilon)n, K)\text{-match with } [p, w_n p]\right) \geqslant 1 - Bc^{\epsilon n}.$$

We now show that for any loxodromic element g in Γ_μ, the probability that $[x, w_n x]$ has an (L, K)-match with a translate of the axis α_g of g tends to one as n tends to infinity, and give an explicit bound on the rate of convergence.

Proposition 7.6. *Given $\delta \geqslant 0$ there is a constant K_0 with the following properties. Let G be a group acting by isometries on a Gromov hyperbolic space, and let μ be a countable, non-elementary, bounded probability distribution on G.*

(1) Let g be a loxodromic element in Γ_μ, and let α_g be an axis for g. Given constants $K \geqslant K_0$ and L, there are constants $B_1 > 0$ and $c_1 < 1$ such that the probability that $\gamma_n = [x, w_n x]$ has an (L, K)-match with a G-translate of α_g is at least $1 - B_1 c_1^{\sqrt{n}}$.

(2) Furthermore, if w_n is loxodromic, then let α_{w_n} be an axis for w_n, and p be a closest point on α_{w_n} to the basepoint x. Then for any $K \geqslant K_0$ and any $L \geqslant 0$, there are constants $B_2 > 0$ and $c_2 < 1$ such that the probability that w_n is loxodromic and $[p, w_n p]$ has an (L, K)-match with a G-translate of α_g is at least $1 - B_2 c_2^{\sqrt{n}}$.

Proof. Let δ be a constant of hyperbolicity for X, and let K_1 be the quasigeodesic constant from Proposition 2.1. Given these values of δ and K_1, let K_0 be the fellow-travelling constant L from Proposition 2.4, and let D be the constant from Proposition 2.4.

There are constants B_1 and $c_1 < 1$ such that the probability that w_n is loxodromic is at least $1 - B_1 c_1^n$. Furthermore, the translation length grows linearly in n, and so there are constants $L_2 > 0, B_2$ and $c_2 < 1$ such that the probability that $\tau(w_n) \geqslant L_2 n$ is at least $1 - B_2 c_2^n$.

Let g be a loxodromic element in Γ_μ, and let us fix $K \geqslant K_0$. As shadow sets form a neighbourhood basis for the topology on ∂X, given two distinct points λ_g^+ and λ_g^- we may choose disjoint shadow sets $V_+ = S_x(y^+, R)$ and $V_- = S_x(y^-, R)$ such that λ_g^+ lies in the interior of V_+ and λ_g^- lies in the interior of V_-. Furthermore, we may choose the shadow sets to be distance at least $L + 2D$ apart. By Proposition 2.4, for any pair of points $\xi_1 \in V_-$ and $\xi_2 \in V_+$, any geodesic segment $[\xi_1, \xi_2]$ has an (L, K)-match with the axis α_g of g.

Let C be the largest distance from the basepoint x to any geodesic with endpoints in V_+ and V_-, i.e. $C := \sup_{(\xi_1,\xi_2) \in A} d(x, [\xi_1, \xi_2]) < +\infty$.

Set $A = V_+ \times V_-$, then if $(\partial_+(\omega), \partial_-(\omega)) \in A$ then the bi-infinite geodesic $\gamma(\omega) := [\partial_+(\omega), \partial_-(\omega)]$ has an (L, K)-match with the axis α_g of g. Furthermore, this match is contained in a $(C + L)$-neighbourhood of the basepoint x.

By definition of the shift map and the fact that the action is by isometries, $\omega \in T^{-n}A$ implies that $\gamma(\omega)$ matches the axis $w_n \alpha_g = \alpha_{w_n g w_n^{-1}}$ in a $(C+L)$-neighborhood of $w_n x$. Then by Proposition 7.3, there are constants B_1 and $c_1 < 1$ such that the probability that $\gamma(\omega)$ has an (L, K)-match with a translate of the axis of g, which contains the nearest point projection of $w_i x$ for some $0 \leqslant i \leqslant n$ is at least $1 - B_1 c_1^{\sqrt{n}}$.

Finally, let us denote as p_i the closest point projection of $w_i x$ to the geodesic $\gamma(\omega)$, and $\ell > 0$ be the drift of the random walk. Then by exponential decay of linear progress and of sublinear tracking ([MT16, Proposition 5.7]), given $\epsilon > 0$ there are constants B_2 and $c_2 < 1$ such that

$$\mathbb{P}(\exists\ i \in \{0, \ldots, n\}\ : |d(p_0, p_i) - \ell i| \geqslant \epsilon i) \leqslant B_2 c_2^n.$$

Therefore, with high probability p_i lies between $\gamma(\epsilon n)$ and $\gamma((\ell - \epsilon)n)$, so by thin triangles, the geodesic $\gamma_n = [x, w_n x]$ also matches a segment of length at least L of a translate of the axis of g.

Now, let p denote the closest point projection of x onto the axis of w_n. By Proposition 7.5 with high probability the segment $[x, w_n x]$ matches a subsegment of $[p, w_n p]$ of length at least $(\ell - \epsilon)n$. Then by the previous argument the segment $[p, w_n p]$ matches a translate of the axis of g. □

We may now complete the proof of Theorem 7.2.

Proof of Theorem 7.2. Let δ be a constant of hyperbolicity for X, and let K_1 be the quasigeodesic constant from Proposition 2.1. Given these values of δ and K_1, let K_0 be the fellow-traveling constant L from Proposition 2.4.

If $K' \geqslant K$ then $\mathrm{Stab}_K(x, y) \subseteq \mathrm{Stab}_{K'}(x, y)$, so without loss of generality, we may assume that $K \geqslant K_0$, where K_0 is the constant from the previous paragraph.

Choose $L \geqslant N_W(K + 2\delta)$, where N_W is the WPD function for g from Lemma 5.2.

If w_n is hyperbolic with axis α_{w_n}, let p be a nearest point on α_{w_n} to the basepoint x. By Proposition 7.6, there are constants B and $c < 1$ such that the probability that w_n is hyperbolic and both $[x, w_n x]$ and $[p, w_n p]$ have (L, K)-matches with a translate of α_g is at least $1 - Bc^{\sqrt{n}}$.

Thus, if h is an element which K-coarsely stabilizes both x and $w_n x$, then by hyperbolicity it also $(K + 2\delta)$-coarsely stabilizes a subsegment of length L of the axis of a conjugate of g. However, by definition of WPD there are only at most $N_W(K + 2\delta)$ elements h which do this, yielding the claim with $N = N_W(K + 2\delta)$. The exact same argument works for $[p, w_n p]$. □

Lemma 7.7. *Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable, non-elementary, bounded, WPD probability distribution on G, and let h be a WPD element in G which lies in Γ_μ. Then for any $\epsilon > 0$, any $K \geqslant K_0$, and any $L > 0$ there are constants $B > 0$ and $c < 1$ such that the probability that every segment $[w_i x, w_{i+\epsilon n} x]$ for $0 \leqslant i \leqslant n(1 - \epsilon)$ has a (L, K)-match with a translate of the axis of h is at least $1 - Bc^{\sqrt{n}}$.*

Proof. By Proposition 7.6, for each i the probability that $[w_i x, w_{i+\epsilon n} x]$ does not have a (L, K)-match with a translate of the axis of h is at most $B_1 c_1^{\sqrt{\epsilon n}}$ for some $c_1 < 1$, and there are at most $n(1-\epsilon)$ possible values of i, hence the total probability is $\leqslant B_1(1 - \epsilon) n c_1^{\sqrt{\epsilon n}}$. The result then follows for suitable choices of B and c. □

8. Non-matching estimates

So far, we have established generic properties of our random walks by proving *matching estimates*, i.e. by showing that with high probability there is a subsegment of the sample path that fellow travels some given element. However, in order to establish our results on the normal closure, we need to prove that the probability of such a matching to occur too often is not so high: we call this a *non-matching estimate*. Note that, while matching happens for random walks on any group of isometries of a hyperbolic space, to prove non-matching one uses crucially the WPD property (and in fact, non-matching may not hold in the non-WPD case, for example, for a dense subgroup of $SL(2,\mathbb{R})$ acting on $\mathbb{H}^2$).

We now define notation for the nearest point projection of a location $w_m x$ of the random walk to a geodesic γ_n from x to $w_n x$.

Definition 8.1. Given integers $0 \leqslant m \leqslant n$, let γ_n be a geodesic from x to $w_n x$, and let $\gamma_n(t_m)$ be a nearest point on γ_n to $w_m x$.

The main non-matching estimate is the following proposition, which says that the probability that γ_n contains in its neighbourhood a translate of a given geodesic segment η starting at $\gamma_n(t_m)$ is bounded above by an exponential function of $\sqrt{|\eta|}$. We will prove it by using the asymptotic acylindricality property established in the previous section.

Proposition 8.2. *Given a constant $\delta \geqslant 0$ there is a constant $K_0 \geqslant 0$ with the following properties. Let G be a group which acts by isometries on the δ-hyperbolic space X, and let μ be a countable, bounded probability distribution on G, such that the random walk generated by μ is asymptotically acylindrical with square root exponential decay.*

Then for any constant $K \geqslant K_0$ there are constants $B > 0$ and $c < 1$, such that for any geodesic segment η and any integer $m \geqslant 0$, the probability that a G-translate of η is contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + |\eta|)]$ is at most $Bc^{\sqrt{|\eta|}}$.

Before embarking on the details, we give a brief overview of the contents of this section. Fix a geodesic segment η of length $2s$. We wish to estimate the probability that some translate of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m+2s)]$. Let $U \subset (G,\mu)^{\mathbb{Z}}$ be the event that some translate of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m+2s)]$, and let V be the event that some translate of the first half of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m+s)]$. The event V is a subset of U, and so by conditional probability $\mathbb{P}(U) \leqslant \mathbb{P}(U \mid V)$. Let U_g be the event that a specific translate $g\eta$ is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m+2s)]$, and let V_g be the event that the first half of $g\eta$ is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m+s)]$. The event U is the union of the events U_g, and the event V is the union of the events V_g. It follows from exponential decay of shadows that $\mathbb{P}(U_g \mid V_g)$ decays exponentially in s. In order to use this fact to estimate $\mathbb{P}(U \mid V)$ we need the following extra information: it follows from asymptotic acylindricality that with high probability any point of V is contained in a bounded number of sets V_g, and this is enough for the exponential decay in s of $\mathbb{P}(U_g \mid V_g)$ to imply exponential decay in s of $\mathbb{P}(U \mid V)$.

We now give the details of the results discussed above. We will need information about the distribution of the nearest point projections of the locations $w_m x_0$ of the random walk to the geodesic γ_n, and we start with the following estimate on Gromov products, which follows directly from exponential decay of shadows.

Proposition 8.3. *Let G be a group acting by isometries on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants B and $c < 1$ such that for all $0 \leqslant i \leqslant n$ and for any $r \geqslant 0$,*

$$\mathbb{P}((x \cdot w_n x)_{w_i x} \geqslant r) \leqslant Bc^r.$$

Proof. If $(x \cdot w_n x)_{w_i x} \geqslant r$, then x lies in a shadow $S_{w_i}(w_n x, R)$, with $d(w_i x, w_n x) - R \geqslant r + O(\delta)$. The random variables w_i and $w_i^{-1} w_n$ are independent, so by exponential decay of shadows, this occurs with probability at most $Bc^{r+O(\delta)}$. □

Linear progress for the locations of the sample path $w_m x_0$ in X, and exponential decay for the distribution of the Gromov products $(x_0 \cdot w_n x_0)_{w_m x_0}$ imply that the points $\gamma_n(t_m)$ are reasonably evenly distributed along $\gamma_n = [x_0, w_n x_0]$. We now make this precise. As μ has bounded support in X, there is a constant D such that any point in γ_n lies within distance at most D from a nearest point projection $\gamma_n(t_i)$ of one of the locations of the random walk $w_i x$, for $0 \leqslant i \leqslant n$, and furthermore, we may choose D to be an upper bound for the diameter of the support of μ in X. For any constant $s \geqslant 0$, let P_s be the collection of indices $0 \leqslant i \leqslant n$ such that $t_i \in [s, s + D]$. This collection is non-empty if $s \leqslant |\gamma_n|$. We emphasize that P_s only contains indices between 0 and n, there may be other locations of the bi-infinite random walk which have nearest point projections to γ_n contained in $[\gamma(s), \gamma(s + D)]$, and we consider this separately in Proposition 8.5 below.

Proposition 8.4. *Let G be a group which acts by isometries on the hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants $0 < L_1 \leqslant L_2$, $B \geqslant 0$ and $c < 1$ such that for any $s > 0$ and any $n \geqslant 0$,*

$$\mathbb{P}(P_s \subseteq [L_1 s, L_2 s]) \geqslant 1 - Bc^s.$$

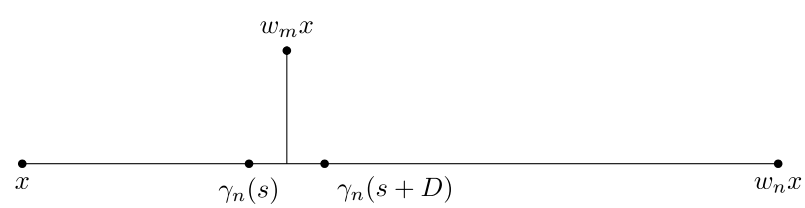

FIGURE 4. The set P_s defined in Proposition 8.4. The index m belongs to P_s as its projection to $[x, w_n x]$ lies within distance s and $s + D$ from the basepoint.

Proof. If $s > d(x, w_n x)$, then $P_s = \varnothing$, and the statement follows immediately, so we may assume that $\gamma_n(s)$ determines a point in γ_n.

By linear progress with exponential decay (Proposition 2.10), there are constants $L > 0, B_1 \geqslant 0$ and $c_1 < 1$ such that for any $m \geqslant 0$

$$\mathbb{P}(d(x, w_m x) \leqslant Lm) \leqslant B_1 c_1^m.$$

Therefore, by summing the geometric series we get

$$\mathbb{P}(d(x, w_m x) \leqslant Lm \text{ for any } m \geqslant N) \leqslant \frac{B_1}{1-c_1} c_1^N.$$

In particular, there are constants B_2 and $c_2 < 1$ such that

$$\mathbb{P}(d(x, w_m x) \geqslant Lm \text{ for all } m \geqslant 2s/L) \geqslant 1 - B_2 c_2^s. \tag{11}$$

If (11) holds, and if $m \geqslant 2s/L$, then $d(x, w_m x) \geqslant Lm \geqslant 2s$, so by thin triangles and the definition of the Gromov product, if the nearest point projection $\gamma(t_m)$ of $w_m x$ lies in $[\gamma_n(s), \gamma_n(s+D)]$, then

$$(x \cdot w_n x)_{w_m x} \geqslant d(x, w_m x) - s - D - O(\delta). \tag{12}$$

By exponential decay for Gromov products (Proposition 8.3), there are constants B_3 and c_3 such that $\mathbb{P}((x \cdot w_n x)_{w_m x} \geqslant r) \leqslant B_3 c_3^r$. In particular,

$$\mathbb{P}((x \cdot w_n x)_{w_m x} \geqslant Lm - s - D - O(\delta)) \leqslant B_3 c_3^{Lm-s-D-O(\delta)}.$$

This implies that there are constants B_4 and $c_4 < 1$ such that for any n

$$\mathbb{P}((x \cdot w_n x)_{w_m x} \geqslant Lm - s - D - O(\delta) \text{ for any } m \geqslant 2s/L) \leqslant B_4 c_4^s. \tag{13}$$

Except for a set of probability at most $B_2 c_2^s + B_4 c_4^s$, we may assume that (11) holds, and (13) does not hold. Equation (12) then implies that $\gamma(t_m)$ does not lie in $[\gamma_n(s), \gamma_n(s+D)]$ for all $m \geqslant 2s/L$. This gives the required upper bound, with $L_2 = 2/L$, and suitable choices of B and c. As μ has bounded support in X, the lower bound may be chosen to be $L_1 = 1/D$. □

We now obtain estimates for the nearest point projections of the remaining locations of the random walk $w_m x$ to a geodesic $\gamma_n = [x, w_n x]$, i.e. for those indices $m \leqslant 0$ and $m \geqslant n$.

Proposition 8.5. *Let G be a group which acts by isometries on the hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants B and c such that for all $s \geqslant 0$ the probability that all of the nearest point projections of $\{w_m x : m \leqslant 0\}$ to $\gamma_n = [x, w_n x]$ are contained within distance s of the initial point x, and all of the nearest point projections of $\{w_m x : m \geqslant n\}$ to γ_n are contained within distance s of the terminal point $w_n x$, is at least $1 - Bc^s$.*

Proof. By the Markov property, the backward random walk $(w_{-n} x)_{n \in \mathbb{N}}$ is independent of γ_n. Similarly, the forward random walk starting at $w_n x$ is also independent of γ_n. More precisely, applying the isometry w_n^{-1}, the random walk $w_n^{-1}(w_m x)_{m \geqslant n}$ starting at x, is independent of $w_n^{-1}\gamma_n$. Therefore, it suffices to show that for any geodesic ray γ starting at x, a random walk has nearest point projection to an initial segment of γ with high probability.

Let γ be a geodesic ray starting at x, with unit speed parameterization, and consider the forward locations of the random walk $(w_n x)_{n \in \mathbb{N}}$. Let $\gamma(t_n)$ be the nearest point projection of a location $w_n x$ to γ. If $t_n \geqslant s$, then $w_n x$ lies in the shadow $S_x(\gamma(s), R)$, for some R which only depends on δ. By (4) the probability that $(w_n)_{n \in \mathbb{Z}}$ ever hits $S_x(\gamma(s), R)$ is at most Bc^s. Therefore the probability that this does not occur for any index n is at least $1 - Bc^s$. □

We now consider the following situation: we have chosen an index $0 \leqslant m \leqslant n$, and a constant $s \geqslant 0$. We wish to estimate the probability that there is a translate of a geodesic η of length $2s$ close to γ_n starting at $\gamma_n(t_m)$. In order to do this, it will be convenient to have information about the distribution of the nearest point projections of $w_k x_0$ to γ_n, and in particular, the sets P_{t_m+s} and P_{t_m+2s}. Proposition 8.6 below assembles the geometric information we need from all of the results above, and in particular shows that with high probability, there are linear bounds on the sizes of P_{t_m+s} and P_{t_m+2s}, and that these sets are disjoint.

Proposition 8.6. *Let G be a group which acts by isometries on the hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants $0 < L_1 \leqslant L_2$, such that for any $0 < \epsilon < 1$, there are constants $B \geqslant 0$ and $c < 1$ such that for any $0 \leqslant m \leqslant n$ and $s > 0$, the probability that all of the following events occur is at least $1 - Bc^s$:*

$$(x \cdot w_n x)_{w_m x} \leqslant \epsilon s \tag{8.6.1}$$
$$L_1 s \leqslant \min P_{t_m+s} \leqslant \max P_{t_m+s} \leqslant L_2 s \tag{8.6.2}$$
$$2L_1 s \leqslant \min P_{t_m+2s} \leqslant \max P_{t_m+2s} \leqslant 2L_2 s \tag{8.6.3}$$
$$(x \cdot w_n x)_{w_i x} \leqslant \epsilon s \text{ for all } i \in P_{t_m+s} \cup P_{t_m+2s} \tag{8.6.4}$$
$$\max P_{t_m+s} \leqslant \min P_{t_m+2s} \tag{8.6.5}$$

The proposition is illustrated in Figure 5 below, where the index $m + a$ belongs to P_{t_m+s}, and $m + b$ belongs to P_{t_m+2s}.

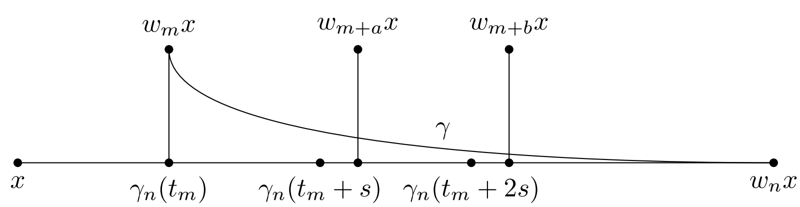

FIGURE 5. Nearest point projections relative to $\gamma_n(t_m)$.

Proof. We say that a function $\mathcal{E}(s) \colon \mathbb{R} \to \mathbb{R}$ is *exponential in s* if there are constants $B \geqslant 0$ and $c < 1$ such that $\mathcal{E}(s) \leqslant Bc^s$ for all $s \geqslant 0$. We observe that the sum of any two functions which are exponential in s is exponential in s, and if $p(s)$ is a polynomial in s, and $\mathcal{E}(s)$ is exponential in s, then $p(s)\mathcal{E}(s)$ is also exponential in s.

By exponential decay for Gromov products (Proposition 8.3), eq. (8.6.1) holds with probability at least $1 - \mathcal{E}_1(s)$, where $\mathcal{E}_1(s) = Bc^s$.

Let γ be a geodesic from $w_m x$ to $w_n x$, with unit speed parametrization, and write $\gamma(t_k)$ for a nearest point projection of $w_k x$ to γ. By the Markov property, we may apply Proposition 8.5 to γ, and so there are constants $B \geqslant 0$ and $c < 1$ such that the probability that

$$\{\gamma(t_k) \ : \ k \in \mathbb{Z}, k \leqslant m\} \subset [w_m x, \gamma(s/2)] \tag{14}$$

holds with probability at least $1 - \mathcal{E}_2(s)$, where $\mathcal{E}_2(s) = Bc^s$.

By thin triangles and assuming that $(x \cdot w_n)_{w_m x} \leqslant \epsilon s$, if the nearest point projection to γ_n of a location $w_{m+a}x$ lies in $[\gamma_n(t_m + s), \gamma_n(t_m + s) + D]$, then the nearest point projection of $w_{m+a}x$ to γ lies in $[\gamma(s), \gamma(s + \epsilon s + D + \delta)]$. Proposition 8.4 applied to each of the $(\epsilon s + \delta)/D$ subsegments of $[s, s + \epsilon s + D + \delta]$ of length D implies that $L_1 s \leqslant a \leqslant L_2(s + \epsilon s + D + \delta)$ with probability at least $1 - \mathcal{E}_3(s)$, where $\mathcal{E}_3(s) = ((\epsilon s + \delta)/D)Bc^s$. Therefore (8.6.2) holds (with a slightly larger value of L_2). Furthermore, by (14) there are no locations $w_k x$ with $k \leqslant m$ or $k \geqslant n$ which have nearest point projections in $[\gamma(s), \gamma(s + \epsilon s + D + \delta)]$.

The exact same argument works for (8.6.3), as long as $t_m + 5s/2 \leqslant |\gamma|$.

Exponential decay for Gromov products then implies (8.6.4) with probability at least $1 - \mathcal{E}_4(s)$, where $\mathcal{E}_4(s) = 3(L_2 - L_1)sBc^{\epsilon s}$. The constant $3(L_2 - L_1)s$ here derives from the cardinality of $P_{t_m+s} \cup P_{t_m+2s}$ when (6.8.2) and (6.8.3) hold.

Finally, if there is some $b < a$, then $(x \cdot w_{m+a}x)_{w_{m+b}x} \geqslant s - D + O(\delta)$, and so the probability that this does not occur for any a and b (i.e. (8.6.5) holds) is at least $1 - \mathcal{E}_5(s)$, where $\mathcal{E}_5(s) = 3(L_2 - L_1)sBc^{s-D+O(\delta)}$.

Therefore all equations (8.6.1)–(8.6.5) hold with probability at least $1 - \mathcal{E}(s)$, where $\mathcal{E}(s)$ is the sum of the functions $\mathcal{E}_1(s)$–$\mathcal{E}_5(s)$ above. All of these functions are exponential in s, so $\mathcal{E}(s)$ is also exponential in s, as required. □

We now show that for any fixed translate $g\eta$ of a geodesic η of length $2s$, if the first half of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$, then the probability that η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + 2s)]$ decays exponentially in s.

Proposition 8.7. *Let G be a group which acts by isometries on the hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants $B \geqslant 0$ and $c < 1$ such that for any geodesic segment η of length $2s$ with initial half-segment η_1 of length s, if there is an isometry $g \in G$ such that $g\eta_1$ is contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$, then the probability that $g\eta$ is contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + 2s)]$ is at most Bc^s.*

Proof. By Proposition 8.6, there are constants B_1 and $c_1 < 1$ such that (8.6.1)–(8.6.5) hold, with probability at least $1 - B_1 c_1^s$.

If $g\eta_1$ is contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$, then in order for η to be contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + 2s)]$, for any index $m + b \in P_{t_m+2s}$ the point $w_{m+b}x$ must lie in a shadow $S_{w_{m+a}x}(g\eta(2s), R)$, where R depends only on K and δ. As w_{m+a} and $w_{m+a}^{-1}w_{m+b}$ are independent, and there are at most $2(L_2 - L_1)s$ elements of P_{t_m+2s}, this happens with probability at most $2(L_2 - L_1)sB_2c_2^s$, by exponential decay for shadows. The result then follows for suitable choices of B and c. □

Proposition 8.7 above only holds for a *fixed* translate $g\eta$. We will use asymptotic acylindricality to extend this result to hold for *some* translate $g\eta$, where g runs over all elements of G. We start with a result from Calegari and Maher [CM15], which says that every point in γ_n is close to some location $w_k x_0$. We say that a point $\gamma(t) \in \gamma_n$ is *K-close* if $d(\gamma(t), w_i x) \leqslant K$ for some $0 \leqslant i \leqslant n$. We shall denote the set of K-close points by $\gamma_{n,K}$.

Lemma 8.8. [CM15, Lemma 5.13] *Given $\delta \geqslant 0$ and positive constants D, L and ϵ, there is a constant $K \geqslant 0$ such that for any sequence of points $x_0, x_1, \ldots x_n$ in* *a δ-hyperbolic space X, with $d(x_i, x_{i+1}) \leqslant D$, and $d(x_0, x_n) \geqslant Ln$, and for any geodesic γ_n from x_0 to x_n, the total length of $\gamma_{n,K}$ is at least*

$$|\gamma_{n,K}| \geqslant (1-\epsilon)|\gamma_n|.$$

Let U be the event that some translate of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + 2s)]$, and let V be the event that the first half of some translate of η is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$. We wish to estimate $\mathbb{P}(U)$. However, as $U \subset V$, the formula for conditional probability implies that $\mathbb{P}(U) \leqslant \mathbb{P}(U \mid V)$, so it suffices to estimate $\mathbb{P}(U \mid V)$.

Let U_g be the event that the translate $g\eta$ is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + 2s)]$, and let V_g be the event that the first half of the translate $g\eta$ is contained in a neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$. The set U is equal to the union of the U_g, and similarly V is equal to the union of the V_g. For each g, we have $\mathbb{P}(U_g \mid V_g) \leqslant Bc^s$, by Proposition 8.7. We wish to use this information to estimate $\mathbb{P}(U \mid V)$. The key property is that asymptotic acylindricality implies that with high probability each point of V is contained in a bounded number of sets V_g, and so exponential decay for the individual conditional probabilities $\mathbb{P}(U_g \mid V_g)$ gives exponential decay for $\mathbb{P}(U \mid V)$. We now give the details of this argument.

Let V and $\{V_i\}_{i \in I}$ be a collection of subsets of a probability space. We say that the collection of sets $\{V_i\}_{i \in I}$ *covers* the set V if $V \subset \bigcup_{i \in I} V_i$. We say that the *covering depth* of the $\{V_i\}_{i \in I}$ is $\sup_{v \in V} \#|\{i \in I \ : \ v \in V_i\}|$. If the covering depth of $\{V_i\}_{i \in I}$ is N, and all sets are measurable, then $\mathbb{P}(V) \leqslant \sum_{i \in I} \mathbb{P}(V_i) \leqslant N\mathbb{P}(V)$.

We will also make use of the following definition:

Definition 8.9. We say that a pair of points x and y are (K, N)-*stable* if

$$\#|\mathrm{Stab}_K(x) \cap \mathrm{Stab}_K(y)| \leqslant N.$$

We say that a geodesic segment η is (K, N)-*stable* if its endpoints are (K, N)-stable.

Proof (of Proposition 8.2). Let $s := |\eta|/2$. We wish to estimate the probability that a translate of η is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + 2s)]$. Let η_1 be the initial subsegment of η with length $|\eta_1| = |\eta|/2 = s$. By Proposition 8.6, we may assume that (8.6.1)–(8.6.5) hold, with probability at least $1 - Bc^s$.

Let us suppose now that a translate $g\eta$ is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + 2s)]$. By thin triangles, the geodesic $g\eta_1$ is contained in a $(K + 2\delta)$-neighbourhood of the geodesic $[w_m x, w_{m+a} x]$. By Lemma 8.8, choosing $\epsilon = 1/8$, there is a constant K_1 such that there are indices i and j, with $w_i x$ within distance $K_2 = K_1 + K + 2\delta$ of $[g\eta_1(0), g\eta_1(s/4)]$ and $w_j x$ within distance K_2 of $[g\eta_1(3s/4), g\eta_1(s)]$. In particular $d(w_i x, w_j x) \geqslant s/2 - 2K_2$, and so

$$|i - j| \geqslant (s/2 - 2K_2)/D. \tag{15}$$

Set $K_3 = \max\{K_2, 5K\}$ and $K_4 = K_3 + 2K + 2\delta$.

Let $U \subseteq (G, \mu)^{\mathbb{N}}$ be the set of sample paths for which a translate of η is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + 2s)]$, and let U_g be the set of sample paths for which $g\eta$ is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + 2s)]$. Let $V \subseteq (G, \mu)^{\mathbb{N}}$ be the set of sample paths for which a translate of η_1 is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + s)]$, and let V_g be the set of sample paths for which $g\eta_1$ is contained in a K-neighbourhood of $[\gamma(t_m), \gamma(t_m + s)]$. As $U \subseteq V$, the conditional probability $\mathbb{P}(U|V)$ satisfies $\mathbb{P}(U) \leqslant \mathbb{P}(U|V)$.

Proposition 8.7 shows that for any g the conditional probability $\mathbb{P}(U_g|V_g)$ decays exponentially in n. The sets $\{U_g\}_{g\in G}$ cover U, in fact $U = \bigcup_{g\in G} U_g$, and similarly $V = \bigcup_{g\in G} V_g$. The covering depth of $\{V_g\}$ is an upper bound on the covering depth of $\{U_g\}$. We now show that with high probability the covering depth of $\{V_g\}$ is bounded, i.e. there exists a set S of large measure such that the covering depth of $\{V_g \cap S\}$ is bounded.

We now have two cases. If η_1 is not $(K_3, N_{ac}(K_4))$-stable, then $w_i x$ and $w_j x$ are not $(K_4, N_{ac}(K_4))$-stable, where $N_{ac}(K)$ is the function from asymptotic acylindricality. Then by Theorem 7.2 the probability that, given i and j, the points $w_i x$ and $w_j x$ are not $(K_4, N_{ac}(K_4))$-stable is at most $Bc^{\sqrt{|j-i|}} \leqslant B_3 c_3^{\sqrt{s/2D}}$ for some constants B_3 and $c_3 < 1$, where we used eq. (15). Recall that by construction $m \leqslant i \leqslant j \leqslant m+a$, and by (8.6.2) we have $a \leqslant L_2 s$, hence there are at most $(L_2 s)^2$ such choices of i, j. Hence, the probability that there are such indices i and j is at most $2(L_2 s)^2 B_3 c_3^{\sqrt{s/2D}}$.

If η_1 is $(K_4, N_{ac}(K_4))$-stable, then by definition the covering depth of V_g is at most $N_{ac}(K_4)$. By Proposition 8.7, there are constants B_4 and $c_4 < 1$ such that $\mathbb{P}(U_g|V_g) \leqslant B_4 c_4^s$. As $U_g \subseteq V_g$, this implies $\mathbb{P}(U_g) \leqslant B_4 c_4^s \mathbb{P}(V_g)$. Therefore

$$\mathbb{P}(U) \leqslant \sum_{g\in G} \mathbb{P}(U_g) \leqslant B_4 c_4^s \sum_{g\in G} \mathbb{P}(V_g) \leqslant N_{ac}(K_4) B_4 c_4^s \mathbb{P}(V) \leqslant N_{ac}(K_4) B_4 c_4^s.$$

Therefore, the probability that a translate of η is contained in a K-neighbourhood of $[\gamma_n(t_m), \gamma_n(t_m + s)]$ is at most $Bc^s + 2(L_2 s)^2 B_3 c_3^{\sqrt{s/2D}} + N_{ac}(K_4) B_4 c_4^s$, which has square root exponential decay in s, as required. □

We are now interested in the particular case of matching between two subsegments of a given geodesic segment. We call this phenomenon a *self-match*. Here is the precise definition.

Definition 8.10. We say that a geodesic segment γ has an (L, K)-*self match* if there exist two disjoint subsegment $\eta, \eta' \subseteq \gamma$ of length L and an element $g \in G \setminus \{1\}$ such that the Hausdorff distance between $g\eta$ and η' is at most K.

Proposition 8.11. *Let G be a group acting by isometries on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G, such that the random walk generated by μ is asymptotically acylindrical with square root exponential decay. Then there is a constant K_0, depending only on δ, such that for any $K \geqslant K_0$, any $L \geqslant 0$ and any $n \geqslant 0$ the probability that γ_n has an (L, K)-self match is at most $n^3 Bc^{\sqrt{L}}$.*

Proof. Suppose that γ_n has an (L, K)-self-match. Then there is a subgeodesic $\eta = [\gamma_n(t), \gamma_n(t+L)]$ such that a translate $g\eta$ is contained in a K-neighbourhood of γ_n, and the nearest point projection of $g\eta$ to γ_n is disjoint from η. Without loss of generality, we may assume that the translate of η is contained in a K-neighbourhood of $[\gamma_n(t+L), \gamma_n(|\gamma_n|)]$.

There is a constant D such that the nearest point projection of the sample path $\{w_m x \colon 0 \leqslant m \leqslant n\}$ to γ_n is D-coarsely onto, and the diameter of the support of μ in X is at most D. Let $w_m x$ be a location of the random walk such that the nearest point projection $\gamma_n(t_m)$ lies within distance D of the interval of γ_n between η and the nearest point projection of $g\eta$.

Then η is contained in a $(K + D + \delta)$-neighbourhood of $[x, w_m x]$, and $g\eta$ is contained in a $(K + D + \delta)$-neighbourhood of $[w_m x, w_n x]$. We do not need to consider all possible subsegments of $[x, w_m x]$, as it suffices to consider those whose endpoints are integer distances from x. More precisely, there is a subsegment $\eta_- = [\gamma_n(a), \gamma_n(b)]$ of η, for integers $a \leqslant b$, with $|\eta_-| \geqslant |\eta| - 2$. If we set $K_1 := K + D + \delta + 1$, then the geodesic η_- K_1-matches $\gamma' = [w_m x, w_n x]$ at distance $\gamma'(c)$ from $w_m x$, where c is also an integer.

There are at most n choices for m, at most $d(x, w_m x) \leqslant Dm \leqslant Dn$ choices for a, and at most $d(w_m x, w_n x) \leqslant D(n - m) \leqslant Dn$ choices for c, so in total at most $D^2 n^3$ choices for the triple (m, a, c). Given a triple of choices m, a and c, and the constant K_1, Proposition 8.2 implies that there are constants B_1 and c_1 such that the probability that a translate of η_- is contained in a K_1-neighbourood of $[w_m x, w_n x]$ is at most $B_1 c_1^{\sqrt{L-2D}}$. Therefore the probability that γ_n has an (L, K)-self-match is at most $D^2 n^3 B_1 c_1^{\sqrt{L-2D}}$, and the result follows by suitable choices of B and c (since D is a constant). □

Finally, we record the following result, which is an immediate consequence of Propositions 8.11 and 7.5.

Corollary 8.12. *For any $\delta \geqslant 0$, there is a constant K_0 with the following properties. Let G be a group acting by isometries on a δ-hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G, such that the random walk generated by μ is asymptotically acylindrical with square root exponential decay. Let $\ell > 0$ be the drift constant for μ, and let p be a point on an axis for w_n.*

Then for any $K \geqslant K_0$ and $\epsilon > 0$, there are constants B and $c < 1$ such that for any $n \geqslant 0$ the probability that either $\gamma_n = [x, w_n x]$ or $[p, w_n p]$ has an $(\epsilon \ell n, K)$-self match is at most $Bc^{\sqrt{n}}$.

9. Asymmetric elements

We now use the non-matching results to show that a generic element is asymmetric in the following sense. This definition is a variation of the one used in [MS18].

Definition 9.1. We say that a loxodromic isometry $g \in G$ is (ϵ, L, K)-*asymmetric* if for any subsegment $[p, q] \subset \alpha_g$ of length at least $\epsilon d(p, gp)$, and any group element h, if $h[p, q]$ is contained in an L-neighbourhood of α_g, then there is an $i \in \mathbb{Z}$ such that $d(hp, g^i p) \leqslant K$ and $d(hq, g^i q) \leqslant K$.

Proposition 9.2. *Given a constant $\delta \geqslant 0$, for any constants $\epsilon > 0$ and $L \geqslant 0$, there is a constant K such that if G is a group acting on a δ-hyperbolic space X, and μ is a countable, non-elementary, bounded probability distribution on G, such that the random walk generated by μ is asymptotically acylindrical with square root exponential decay, then there are constants B and $c < 1$ such that the probability that w_n is (ϵ, L, K)-asymmetric is at least $1 - Bc^{\sqrt{n}}$.*

We first recall the following useful fact about isometries of Gromov hyperbolic spaces.

Proposition 9.3. *Given $\delta \geqslant 0$ there is a constant K_0 such that for any $K \geqslant K_0$, if X is a δ-hyperbolic space, and g is an isometry for which there is a point $x \in X$*

such that $d(x, gx) \geqslant 3K$ and $(x \cdot g^2x)_{gx} \leqslant K$, then g is loxodromic, and the axis α_g of g passes within distance $2K$ of gx.

Proof. This follows from the following estimate for the translation length of an isometry:

$$\tau(g) \geqslant d(x, gx) - 2(x \cdot g^2x)_{gx} - O(\delta),$$

see for example [MT16, Proposition 5.8]. As long as $\tau(g) \geqslant O(\delta)$, then any path $[x, gx]$ has a subsegment which is contained in an L_1-neighbourhood of α_g, and so by thin triangles, the distance from gx to α_g is at most $(x \cdot g^2x)_{gx} + L_1 + O(\delta)$. $\square$

Let γ_1 and $\gamma_2 = [x, y]$ be two $(1, K_1)$-quasigeodesics. Parameterizations $\gamma_1 \colon I_1 \to X$ and $\gamma_2 \colon I_2 \to X$ determine orientations of γ_1 and γ_2. Let $x' = \gamma_1(s)$ be a closest point on γ_1 to x, and let $y' = \gamma_1(t)$ be a closest point on γ_1 to y. We say these orientations agree if $s < t$ for any choice of closest points $x' = \gamma_1(s)$ and $y' = \gamma_1(t)$, and we say they disagree if $s > t$ for any choice of closest points $x' = \gamma_1(s)$ and $y' = \gamma_1(t)$. In any other case we say that the orientation of γ_2 is not well-defined with respect to γ_1. We omit the proof of the following basic fact.

Proposition 9.4. *Given constants δ, K_1 and L, there is a constant L' with the following properties. Let X be a δ-hyperbolic space, and let γ_1 and γ_2 be $(1, K_1)$-quasigeodesics in X such that γ_2 is contained in an L-neighbourhood of γ_1. If the length of γ_2 is at least L', then the orientation of γ_2 either agrees or disagrees with that of γ_1.*

Recall that we say a function $\mathcal{E}(n) \colon \mathbb{N} \to \mathbb{N}$ is exponential in n if there are constants B and $c < 1$ such that $\mathcal{E}(n) \leqslant Bc^n$ for all $s \geqslant 0$. Clearly, if $\mathcal{E}_1(n)$ is exponential in n, and $\mathcal{E}_2(n)$ is exponential in $\sqrt{n}$, then the sum of these two functions is exponential in $\sqrt{n}$.

We may now complete the proof of Proposition 9.2.

Proof of Proposition 9.2. If $L' \geqslant L$, then $N_L(\alpha_g) \subseteq N_{L'}(\alpha_g)$, so if the result holds for some K' and L', it also holds for K' and L. Therefore, without loss of generality we may assume that $L \geqslant 1 + \delta$.

Let α_{w_n} be an axis for w_n, and let x' be the nearest point projection of the basepoint x to α_{w_n}. If the result holds for some $\epsilon > 0$, it holds for any larger value of ϵ, so we may assume that $\epsilon \leqslant 1$. Furthermore, as α_{w_n} is w_n-invariant, after translating by a power of w_n, and possibly replacing ϵ by $\epsilon/2$, we may assume that $w_n^i[p, q]$ is contained in $[x', w_n x']$. By abuse of notation, we will relabel $w_n^i[p, q]$ as $[p, q]$.

If $h[p, q]$ is contained in a L-neighbourhood of α_{w_n}, then as α_{w_n} is w_n-invariant, then after replacing h by $w_n^k h$, we may assume that the nearest point projection of $h[p, q]$ to α_{w_n} is contained in $[x', w_n^2 x']$. By abuse of notation, we will relabel $w_n^k h$ as h.

Given L, let L' be the constant from Proposition 9.4. As $d(x', w_n x')$ tends to infinity almost surely as n tends to infinity, we may assume that $d(x', w_n x') \geqslant L'/\epsilon$, and so $d(p, q) \geqslant L'$. In particular, the orientation of $h[p, q]$ is well defined with respect to α_{w_n}, and either agrees, or disagrees with the orientation of α_{w_n}.

First consider the case in which h reverses the orientation of $[p, q]$ with respect to α_{w_n}, as illustrated below in Figure 6. We will show that if this occurs, it gives a self-match for γ_n which occurs with probability which is at most exponential in $\sqrt{n}$.

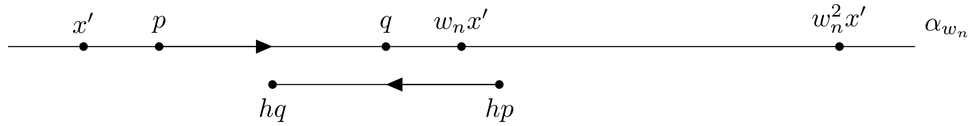

FIGURE 6. An orientation reversing translate of $[p, q]$ close to α_{w_n}.

By replacing $[p, q]$ by either its initial half, or terminal half, we may assume that either $[p, q]$ or $w_n^{-1}[p, q]$ has nearest point projection to α_{w_n} contained in $[x', w_n x']$. Again replacing $[p, q]$ by either its initial half, or terminal half, we may assume that $h[p, q]$ lies within distance K of a disjoint subsegment of $[x', w_n x']$ of length at least $\epsilon d(x', w_n x')/4$. This gives rise to an $(\epsilon d(x', w_n x')/4, K)$-self match for $[x', w_n x']$.

Let $\ell > 0$ be the linear progress constant for μ, and fix some $0 < \epsilon' < \min\{\ell, 1\}/2$.

The subsegment $[x', w_n x']$ of α_{w_n} is contained in an L_1-neighbourhood of $[x, w_n x]$, and by Proposition 7.5, given $\epsilon' > 0$, there are constants B_1 and $c_1 < 1$ such that the probability that the length of $[x', w_n x']$ is at least $(\ell - \epsilon')n$ is at least $1 - \mathcal{E}_1(n)$, where $\mathcal{E}_1(n) = B_1 c_1^n$, where ℓ is the linear progress constant for μ.

This gives an $(\epsilon(\ell - \epsilon')n/4, K)$-self match for $[x, w_n x]$, and by Proposition 8.11, there are constants B_2 and $c_2 < 1$ such that the probability that this occurs is at most $\mathcal{E}_2(n) = B_2 c_2^{\sqrt{n}}$.

Therefore, the existence of an orientation reversing translate of $[p, q]$ occurs with probability at most $\mathcal{E}_1(n) + \mathcal{E}_2(n)$, which is exponential in $\sqrt{n}$, as required.

We now consider the case in which the orientation of $h[p, q]$ agrees with that of α_{w_n}. We may replace $[p, q]$ by either its initial half or terminal half subinterval (in which case replace ϵ by $\epsilon/2$), and possibly replace h by $w_n^{-1}h$, to ensure that the nearest point projection of $h[p, q]$ to α_{w_n} is contained in $[x', w_n x']$. This is illustrated below in Figure 7.

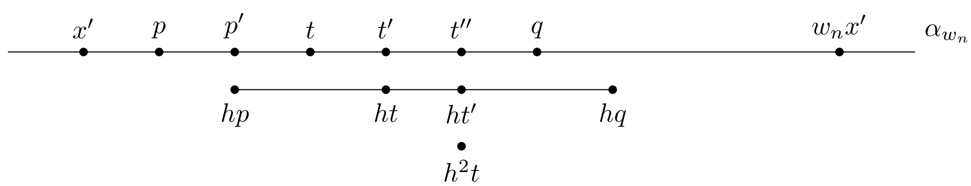

FIGURE 7. An orientation preserving translate of $[p, w_n p]$ close to α_{w_n}.

Let p' be a nearest point on α_{w_n} to hp. If $d(p, p') \geqslant \epsilon \ell n/10$, then this gives a linear size self-match of $[x, w_n x]$, and again by Proposition 8.11 there are constants B_3 and $c_3 < 1$ such that the probability that this occurs is at most $\mathcal{E}_3(n) = B_3 c_3^{\sqrt{n}}$.

We shall choose a constant $K = 4L + O(\delta)$, but in order to guarantee that there is no circularity in our choice of constants, we now recall some basic facts about Gromov hyperbolic spaces and give an explicit choice of the $O(\delta)$ term in terms of geometric constants which only depend on δ.

Recall that every axis is a $(1, K_1)$-quasigeodesic, where K_1 only depends on δ. Let L_1 be a Morse constant for $(1, K_1)$-quasigeodesics, i.e. any geodesic $[x, y]$ with endpoints in a $(1, K_1)$-quasigeodesic α is contained in an L_1-neighbourhood of α. As K_1 only depends on δ, the Morse constant L_1 also only depends on δ.

Given constants $\delta \geqslant 0$ and $K_1 \geqslant 0$ there are constants K_2 and K_3, such that for any $(1, K_1)$-quasigeodesic α, and any two points x and y in X, if x' is the nearest point projection of x to α and y' is the nearest point projection of y to α, then if x' and y' are distance at least K_2 apart, then the geodesic from x to y is Hausdorff distance at most K_3 from the piecewise geodesic path $[x, x'] \cup [x', y'] \cup [y', y]$. Furthermore

$$d(x', y') \geqslant d(x, y) - d(x, x') - d(y, y') - K_3. \tag{16}$$

As K_1 only depends on δ, the constants K_2 and K_3 also only depend on δ. We may now set $K = 4L + 2K_1 + 3K_2 + 3K_3 + 6\delta$.

Now suppose that p' is close to p and the length of $[p, p']$ is greater than K but less than $\epsilon \ell n/10$. Let t be any point in $[p', q]$. Let t' be a closest point on $[p, q]$ to ht, and let t'' be a closest point on $[p, q]$ to ht'.

Claim 9.5. We have chosen K sufficiently large such that $d(t, t') \geqslant K_2$.

Proof. By (16),

$$d(p', t') \geqslant d(hp, ht) - d(hp, p') - d(ht, t') - K_3.$$

As h is an isometry, and $d(hp, p')$ and $d(ht, t')$ are at most L, this gives

$$d(p', t') \geqslant d(p, t) - 2L - K_3.$$

The points p, p', t and t' all lie on the $(1, K_1)$-quasigeodesic α_{w_n}, which implies $d(p', t) + d(t, t') \geqslant d(p', t') - K_1$, and $d(p, t) \geqslant d(p, p') + d(p', t) - K_1$. This yields

$$d(t, t') \geqslant d(p, p') - 2L - 2K_1 - K_3.$$

Our choice of K therefore guarantees that $d(t, t') \geqslant K_2$, as required. In fact $d(t, t') \geqslant 2L + K_2 + K_3 \geqslant K_2$, and we will now use this stronger bound to obtain a bound on $d(t', t'')$. □

Claim 9.6. We have chosen K sufficiently large such that $d(t', t'') \geqslant K_2$.

Proof. By (16),

$$d(t', t'') \geqslant d(ht, ht') - d(ht, t') - d(ht', t'') - K_3.$$

as h is an isometry, and $d(ht, t')$ and $d(ht', t'')$ are at most L, this gives

$$d(t', t'') \geqslant d(t, t') - 2L - K_3.$$

Our choice of K then implies that $d(t', t'') \geqslant K_2$, as required. □

As $d(t', t'') \geqslant K_2 + L$, the geodesic from ht to $h^2 t$ passes within distance K_3 of $[t', t'']$, the Gromov product $(t \cdot h^2 t)_{ht}$ is at most $K_4 := L + K_2 + K_3 + 2\delta$. We have chosen K sufficiently large such that $d(t, ht) \geqslant 3K_4$, and so Proposition 9.3 implies that h is loxodromic, and the axis of h passes within distance $2K_4$ of α_{w_n}.

As we have assumed that $\tau(h) \leqslant \epsilon \ell n/10$, this gives a $(\epsilon \ell n/10, 2K_4)$-self match of $[x', w_n x']$, and hence of $\gamma_n = [x, w_n x]$, and so again by Proposition 8.11 there are constants B_4 and $c_4 < 1$ such that the probability that this occurs is at most $\mathcal{E}_4(n) = B_4 c_4^{\sqrt{n}}$.

Therefore, we have shown that the case of an orientation preserving translate of $[p, q]$ occurs with probability at most $\mathcal{E}_3(n) + \mathcal{E}_4(n)$, which is exponential in $\sqrt{n}$, as required. $\square$

10. Genericity of WPD elements

Let us now prove that WPD elements are generic for the random walk as long as there is one WPD element in the semigroup generated by the support of μ. This establishes Theorem 1.1 in the Introduction.

Theorem 10.1. *Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable, non-elementary, bounded, WPD probability distribution on G. Then there are constants B and $c < 1$ such that the probability that w_n is WPD satisfies*

$$\mathbb{P}(w_n \text{ is WPD}) \geqslant 1 - Bc^{\sqrt{n}}.$$

This has the following immediate corollary.

Corollary 10.2. *Let G be a group acting on a Gromov hyperbolic space X. Let μ be a countable, non-elementary, bounded, WPD probability distribution on G, and let Λ_μ be the limit set of Γ_μ in ∂X. Then endpoints $(\lambda^+(g), \lambda^-(g))$ of elements g which are WPD in G are dense in $\Lambda_\mu \times \Lambda_{\check{\mu}}$.*

Proof. By Borel-Cantelli, for almost every sample path ω there is an N such that $w_n(\omega)$ is WPD for all $n \geqslant N$. The distribution of endpoints of loxodromic elements $(\lambda^+(w_n(\omega)), \lambda^-(w_n(\omega)))$ converges in distribution to $\nu \times \check{\nu}$, the product of the hitting and reflected hitting measures, as n tends to infinity. This is shown in [Mah10a, Theorem 4.1] for the mapping class group acting on the curve complex, but using the convergence to the boundary result of [MT16, Theorem 1.1], the argument holds for a non-elementary random walk on a countable group acting on a Gromov hyperbolic space. For any pair of points $(\lambda, \lambda') \in \Lambda_\mu \times \Lambda_{\check{\mu}}$, any open set $U \times U'$ containing (λ, λ') has positive measure $\nu(U)\check{\nu}(U') > 0$ by [MT16, Proposition 5.4], and so there are infinitely many WPD elements with one endpoint in U and the other in U'. In particular, endpoints of WPD elements are dense in $\Lambda_\mu \times \Lambda_{\check{\mu}}$. $\square$

Given a loxodromic element g, its associated *maximal elementary subgroup* $E_G(g)$ is defined as the stabilizer of the two endpoints of the axis of g, i.e.

$$E_G(g) = \mathrm{Stab}^G(\{\lambda_g^+, \lambda_g^-\})$$

(note that elements of $E_G(g)$ may permute the two fixed points). We will use the following result due to Bestvina and Fujiwara [BF02, Proposition 6].

Theorem 10.3. *Let G act on X with a WPD element h, with axis α_h. Then $E_G(h)$ is the unique maximal virtually cyclic subgroup containing h. Furthermore, for any constant $K \geqslant 0$ there is a number L, depending on h, δ and K, such that if $g \in G$ is an element which K-coarsely stabilizes a subsegment of α_h of length L, then g lies in $E_G(h)$.*

That is, if α_h is an axis of h, then

$$E_G(h) = \{g \in G \ : \ d_{Haus}(g\alpha_h, \alpha_h) < \infty\}.$$

This is stated in [BF02] for a group action in which all loxodromic elements are WPD, but the proof works for a group acting non-elementarily on a Gromov hyperbolic space which has at least one WPD element.

If h is WPD, then as $E_G(h)$ is virtually cyclic, it contains $\langle h \rangle$ as a finite index subgroup. We now record the following elementary property of $E_G(h)$, that the image of this group in X under the orbit map intersects any bounded set in only finitely many points.

Proposition 10.4. *Let G be a group acting on a Gromov hyperbolic space X which contains a loxodromic isometry h, and let H be a subgroup of G which contains $\langle h \rangle$ as a finite index subgroup. Then for any $x \in X$ and $K \geqslant 0$, there is an N such that $\#|Hx \cap B_K(x)| \leqslant N$.*

Proof. As $\langle h \rangle$ is a finite index subgroup of H, there is a finite set of group elements F such that H is a finite union of right cosets $\langle h \rangle f$, for $f \in F$. In particular, any element $g \in H$ may be written as $g = h^k f$, for some $k \in \mathbb{N}$ and $f \in F$. By the triangle inequality, $d(x, gx) \geqslant d(x, h^k x) - d(x, fx)$. The distances $d(x, fx)$ have an upper bound depending on F and x, and $d(x, h^k x) \geqslant k\tau(h)$, so there are only finitely many group elements $g \in H$ with $d(x, gx) \leqslant K$. □

It is well known that the following (a priori weaker) definition, which we shall refer to as *axial WPD*, is equivalent to WPD.

Definition 10.5. Let G be a group acting on a δ-hyperbolic space X, and let h be a loxodromic isometry with an axis α_h. Then h is an *axial WPD* if there exists $p \in \alpha_h$ such that for any constant $K \geqslant 0$, there is an $M > 0$, such that

$$\#|\mathrm{Stab}_K(p) \cap \mathrm{Stab}_K(h^M p)| < \infty.$$

Proposition 10.6. *Let G be a group acting on a δ-hyperbolic space X, and let h be a loxodromic isometry. Then h is an axial WPD if and only if h is WPD.*

Proof. If h is WPD, then it is an axial WPD. We now show the other direction. By the triangle inequality, for any $x, y \in X$, $g \in G$, and $K \geqslant 0$

$$\mathrm{Stab}_K(y) \cap \mathrm{Stab}_K(h^M y) \subseteq \mathrm{Stab}_{K'}(x) \cap \mathrm{Stab}_{K'}(h^M x)$$

where $K' = K + 2d(x, y)$. □

We now complete the proof of Theorem 10.1.

Proof of Theorem 10.1. By Proposition 10.6 it suffices to show that w_n is an axial WPD.

By Theorem 2.5, there are constants B_1 and $c_1 < 1$ such that the element w_n is loxodromic with probability at least $1 - \mathcal{E}(n)$, where $\mathcal{E}_1(n) = B_1 c_1^n$, which is exponential in n. Let α_{w_n} be an axis for w_n.

Let L_1 be a fellow travelling constant for $(1, K_1)$-quasigeodesics, given by Proposition 2.3. Let δ be a constant of hyperbolicity for the space X. Fix $\epsilon = 1 > 0$, and then let $K_2 \geqslant 0$ be the constant from Proposition 9.2 given these values of δ, ϵ and L_1. As K_1 only depends on δ, then L_1, and hence K_2, also only depends on δ. Let p be a point on α_{w_n}, which, without loss of generality, we may choose to be a closest point projection of x to α_{w_n}. This is illustrated below in Figure 8.

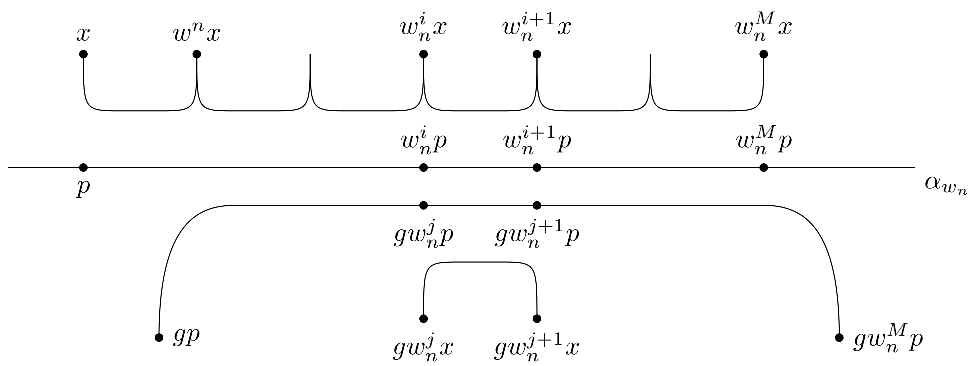

FIGURE 8. Fellow-traveling with a translate of α_h.

For any loxodromic w_n, and for any $K \geqslant 0$, there is an $M \geqslant 0$ sufficiently large such that $d(x, w_n^M x) \geqslant 2(K + \tau(w_n))$. Therefore, by Proposition 2.3, there is a translate $gw_n^j[p, w_n p]$ which is contained in an L_1-neighbourhood of α_{w_n}, with $0 \leqslant j \leqslant M$. By Proposition 9.2, there is an index i with $0 \leqslant i \leqslant M$ such that $d(w_n^i p, gw_n^j p) \leqslant K_2$ and $d(w_n^{i+1} p, gw_n^{j+1} p) \leqslant K_2$, where K_2 is the constant determined above which only depends on δ. In particular, $w_n^{-i} g w_n^j$ K_2-stabilizes both p and $w_n p$.

By Theorem 7.2, given K_2, there are constants $N = N_{ac}(K_2), B_2$ and $c_2 < 1$ such that

$$\mathbb{P}\left(\#|\mathrm{Stab}_{K_2}(p, w_n p)| \leqslant N\right) \geqslant 1 - B_2 c_2^{\sqrt{n}}.$$

Therefore, for any K, there are at most finitely many (in fact $N(M+1)^2$) choices for g, and so w_n is WPD, with probability at least $1 - \mathcal{E}_1(n) - \mathcal{E}_2(n)$, where $\mathcal{E}_2(n) = B_2 c_2^{\sqrt{n}}$, which is exponential in $\sqrt{n}$, as required. □

11. Small cancellation and normal closure

We will now prove results on the normal closure (Theorems 1.3 and 1.4 in the Introduction). In order to do so, we will use the following notions of *small cancellation* from [DGO17]. If $H \subseteq G$ is a subgroup, we define its *injectivity radius* as

$$\mathrm{inj}(H) := \inf\{d(gx, x) \ : \ g \in H \setminus \{1\}, x \in X\}.$$

Let $\mathcal{R}$ be a family of loxodromic elements which is closed under conjugation. We define its *injectivity radius* as

$$\mathrm{inj}(\mathcal{R}) := \inf_{g \in \mathcal{R}} \inf\{d(g^k x, x), k \in \mathbb{Z} \setminus \{0\}, x \in X\}.$$

In particular, if g is loxodromic and $\mathcal{R} := \{hgh^{-1}, h \in G\}$ is the set of conjugates of g, then

$$\mathrm{inj}(\mathcal{R}) \geqslant \tau(g).$$

Following [DGO17], for a loxodromic element g, let $\mathrm{Ax}(g)$ be the 20δ-neighbourhood of set of points x for which $d(x, gx) \leqslant \inf_{y \in X} d(y, gy) + \delta$. If $\tau(g)$ is sufficiently large, then this set is contained in a bounded neighbourhood of a quasi-axis α_g for g.

Proposition 11.1. *Given $\delta \geqslant 0$, there are constants A and K, such that if g is a loxodromic isometry of δ-hyperbolic space X with quasi-axis α_g and $\tau(g) \geqslant A$, then $\mathrm{Ax}(g) \subset N_K(\alpha_g)$. Furthermore, $\mathrm{Ax}(g)$ is 10δ-quasiconvex.*

Proof. Let x be a point in X, and let p be a closest point on α_g to x. As we may assume that α_g is g-invariant, gp is a closest point on α_g to gx, and $d(p, gp) \geqslant \tau(g)$. Given δ, there are constants A_1 and K_1 such that if $d(p, gp) \geqslant A_1$, then the union of the three geodesic segments $[x, p], [p, gp]$ and $[gp, gx]$ is contained in a bounded neighbourhood of a geodesic $[x, gx]$, and in particular,

$$d(x, gx) \geqslant d(x, p) + d(p, gp) + d(gp, gx) - K_1.$$

This is an elementary application of thin triangles, see for example [MT16, Proposition 2.3] for the geodesic case. As the quasigeodesics constants for the quasi-axis α_g only depend on δ, A_1 and K_1 may also be chosen to only depend on δ. Therefore, if $d(x, p) \geqslant B_1 + \delta$ then x does not lie in $\mathrm{Ax}(x)$, so we may choose $A = A_1$ and $K = K_1 + \delta$.

For the final statement, see for example Coulon [Cou13, Proposition 3.10]. □

We also define, for g and h loxodromic,

$$\Delta(g, h) := \mathrm{diam}\ \left(N_{20\delta}(\mathrm{Ax}(g)) \cap N_{20\delta}(\mathrm{Ax}(h))\right)$$

where $N_R(Y)$ denotes the R neighbourhood of the set Y in X.

Let g be a loxodromic element in G. We shall write $E^+_G(g)$ for the orientation preserving subgroup of $E_G(g)$, i.e. the subgroup which stabilizes λ^+_g and λ^-_g pointwise. This group is either equal to $E_G(g)$ or has index two in $E_G(g)$. There are elements g with $E_G(g) = E^+_G(g)$, and in fact they are generic.

Corollary 11.2. *Let G be a group acting by isometries on a δ-hyperbolic space X, and let μ be a countable, non-elementary, bounded probability distribution on G. Then there are constants B and $c < 1$ such that the probability that w_n is loxodromic with $E_G(w_n) = E^+_G(w_n)$ is at least $1 - Bc^{\sqrt{n}}$.*

Proof. If $E^+_G(w_n)$ is index two in $E_G(w_n)$, then there is an element f which reverses the orientation of α_{w_n}. This gives an $(\ell n/4, K)$-self match of $[p, w_n p]$, where $\ell > 0$ is the positive drift constant for μ, and K is the fellow travelling constant from Proposition 2.3. However by Corollary 8.12, there are constants B and $c < 1$ such that the probability that this occurs is at most $Bc^{\sqrt{n}}$. □

An essential feature of asymmetric elements is the following.

Proposition 11.3. *Given $\delta \geqslant 0$, there are constants K and L such that if g is a WPD element of G which is $(1, L, K)$-asymmetric, with translation length $\tau(g) > 3L + 2K$, then there is a surjective homomorphism $\phi\colon E^+_G(g) \to \mathbb{Z}$ with $\phi(g) = 1$. In particular,*

$$E^+_G(g) = \langle g \rangle \ltimes \ker \phi,$$

where $\ker \phi$ is finite and consists precisely of the elliptic elements of $E^+_G(g)$.

Note that the proposition is not true if one replaces $E^+_G(g)$ by $E_G(g)$, as the latter may contain infinitely many elliptic elements (think of the action of the infinite dihedral group on $\mathbb{Z}$).

Proof. Let p be a point on the axis α_g. Let L be the fellow travelling constant from Proposition 2.3. The axis α_g is L-coarsely preserved by $E^+_G(g)$. As g is $(1, L, K)$-asymmetric, the set $\{g^i p \colon i \in \mathbb{Z}\}$ is K-coarsely preserved by $E^+_G(g)$. As elements act by isometries, this gives an action of $E^+_G(g)$ on $\mathbb{Z}$, defined as follows. If $f \in E^+_G(g)$, $\phi(f)$ sends $g^i p$ to the closest $g^j p$ to $f g^i p$. As g is WPD, the group $E^+_G(g)$ is virtually cyclic, so $\ker \phi$ is finite. The element $g \in E^+_G(g)$ maps to $1 \in \mathbb{Z}$ and gives a splitting, so $E^+_G(g) = \langle g \rangle \ltimes \ker \phi$.

As $\ker \phi$ is a finite subgroup of G, all elements of $\ker \phi$ are elliptic. If $\phi(f) \neq 0$, then as $\tau(g) \geqslant 3L + 2K$, the three points p, fp and $f^2 p$ satisfy $d(p, fp) \geqslant 3L$, $d(fp, f^2 p) \geqslant 3L$ and $(p \cdot f^2 p)_{fp} \leqslant L$, and so f is loxodromic by Proposition 9.3. $\square$

Let G_{WPD} denote the set of WPD elements of G, and let $H \leqslant G$ be a subgroup of G which contains an element of G_{WPD}. Define

$$E^+_G(H) := \bigcap_{g \in H \cap G_{\mathrm{WPD}}} E^+_G(g).$$

and an equivalent definition holds for $E_G(H)$. We will also use the notation $E(G) := E_G(G)$ when G and H are equal.

Recall that two elements h_1, h_2 of G are *commensurable* if some power of h_1 is conjugate to some power of h_2, and *non-commensurable* otherwise. The result below follows from the arguments in [DGO17, Lemma 6.17], but we give the details for the convenience of the reader.

Proposition 11.4. *Let G be a group acting by isometries on a Gromov hyperbolic space X, and let H be a non-elementary subgroup of G which contains an element of G_{WPD}. Then there exist two independent, WPD elements h_1, h_2 in H such that*

$$E^+_G(h_1) \cap E^+_G(h_2) = E^+_G(H).$$

Moreover, for any $K \geqslant 0$ there exists an element f in H such that for any $z \in \alpha_f$ one has

$$Stab_K(z, fz) \subseteq E^+_G(H).$$

Proof. By [DGO17, Corollary 6.12], there exist two non-commensurable, loxodromic, WPD elements h_1, h_2 in H (pick h_1 as one such element, then apply Corollary 6.12 with the subgroup called G in Corollary 6.12 chosen to be H, the subgroup called H in the Corollary 6.12 chosen to be $E_G(h_1)$ and $a \in H \setminus E_G(h_1)$). Let N be the normalizer of H in G, i.e.

$$N := \{x \in G \;:\; xHx^{-1} = H\}$$

which contains the group H. Denote as $T(h_i)$ the set of finite order elements in $E^+_G(h_i)$. In $E^+_G(h_i)$ every conjugacy class is finite (since all conjugate elements have equal translation length), so a result of Neumann [Neu51] then implies that the set $T(h_i)$ of finite order elements is a finite group. Let us suppose that for any $x \in N$ we have

$$E^+_G(x h_1 x^{-1}) \cap E^+_G(h_2) \neq E^+_G(H).$$

Note moreover that

$$E^+_G(x h_1 x^{-1}) \cap E^+_G(h_2) = x T(h_1) x^{-1} \cap T(h_2).$$

Given $(s,t) \in P := T(h_1) \times (T(h_2) \setminus E^+(H))$, we pick $y \in N$ such $ysy^{-1} = t$, if it exists, and $y(s,t) = 1$ otherwise. Let $C_N(t)$ be the centralizer of t in N. Now, we claim that

$$N = \bigcup_{(s,t)\in P} y(s,t) C_N(t).$$

Indeed, let $x \in N$. Then since $xT(h_1)x^{-1} \cap T(h_2) \neq E^+_G(H)$, then there exists $s \in T(h_1)$ and $t \in T(h_2) \setminus E^+(H)$ such that $s = x^{-1}tx \in T(h_1)$. Thus if $y = y(s,t)$ then $s = x^{-1}tx = y^{-1}ty$, so $xy^{-1} \in C_N(t)$. This means that there is a finite collection of cosets of the subgroups $C_N(t)$, with $t \in T(h_2) \setminus E^+(H)$, which covers N, and a theorem of Neumann [Neu54] then implies that at least one of these subgroups has finite index in N. Therefore, there is a $t \in T(h_2) \setminus E^+_G(H)$ such that $C_N(t)$ has finite index in N. Hence, if $h \in N$ is a WPD element, then there exists $k > 0$ such that $h^k t = th^k$, hence $t \in E^+_G(h)$. Thus, $t \in E^+_G(N) \subseteq E^+_G(H)$, which is a contradiction. Finally, let us note that the claim implies that h_1 and h_2 are independent. In fact, as both h_1 and h_2 are WPD, the fixed point sets of h_1 and h_2 cannot have a common point. This is because in this case both h_1 and h_2 would coarsely stabilize a large segment of the axis of h_1, which by Theorem 10.3, would imply that $E^+_G(h_1) = E^+_G(h_2)$, contradicting the non-commensurability of h_1 and h_2.

We now prove the second claim. As h_1 and h_2 are independent loxodromic isometries, the ping-pong lemma implies that for any $n > 0$ sufficiently large, the orbit map gives a quasi-isometric embedding of the free group $\langle h_1^n, h_2^n \rangle$ in X. In particular, for all $m > 0$, the element $f := h_1^{nm} h_2^{nm}$ is loxodromic.

Fix some $K \geqslant 0$, and let L_1 be the fellow travelling constant for $(1, K_1)$-quasigeodesics from Proposition 2.4. Let L_2 be the constant given by Theorem 10.3 using the constant $K + 2\delta + L_1$. We may choose m sufficiently large so that there are two segments $\eta_1 \subseteq \alpha_{h_1}$ and $\eta_2 \subseteq \alpha_{h_2}$ of length $\geqslant L_2$, and a segment $\eta \subseteq \alpha_f$ such that

$$\eta_1 \cup \eta_2 \subseteq N_{L_1}(\eta).$$

Thus, if h belongs to $\mathrm{Stab}_K(z, fz)$, then for some $k \in \mathbb{Z}$ the isometry $f^k h f^{-k}$ $(K + 2\delta)$-coarsely stabilizes the segment η, hence it also $(K + 2\delta + L_1)$-coarsely stabilizes both η_1 and η_2, and preserves the orientation of the axes. Then by Theorem 10.3 it is contained in

$$E^+_G(h_1) \cap E^+_G(h_2) = E^+_G(H).$$

Thus, h belongs to $f^{-k} E^+_G(H) f^k = E^+_G(H)$, as required. □

From now on we shall assume that the probability distribution μ is reversible, so Γ_μ is a group. We will use the notation $E_\mu := E^+_G(\Gamma_\mu)$.

Corollary 11.5. *Given $\delta \geqslant 0$ there are constants K and L with the following properties. Let G be a group acting by isometries on a $\delta-$hyperbolic space X, and let μ be a countable, non-elementary, reversible, bounded, WPD probability distribution on G. Then there are constants B and $c < 1$ such that the probability that w_n is loxodromic, $(1, L, K)$-asymmetric, WPD with*

$$E_G(w_n) = E^+_G(w_n) = \langle w_n \rangle \ltimes E_\mu$$

is at least $1 - Bc^{\sqrt{n}}$. In particular, if E_μ is trivial, then $E_G(w_n)$ is cyclic with probability at least $1 - Bc^{\sqrt{n}}$.

Proof. We are left with proving the last claim. By Proposition 9.2, we know that there are constants B_1 and $c_1 < 1$ such that the probability that w_n is $(1, L, K)$-asymmetric is at least $1 - B_1 c_1^{\sqrt{n}}$, hence

$$E_G^+(w_n) = \langle w_n \rangle \ltimes \ker \phi$$

where $\phi : E_G^+ \to \mathbb{Z}$ is the homomorphism given in Proposition 11.3. Now, since w_n is asymmetric, we have that $\ker \phi$ is the (finite) set of elliptic elements in $E_G^+(w_n)$, hence it is contained in $\mathrm{Stab}_K(p, w_n p)$ where p is some point on the axis of w_n. Let $f \in \Gamma_\mu$ be given by Proposition 11.4. By Proposition 7.6, there are constants B_2 and $c_2 < 1$ such that the probability the axis of w_n has a (L, K)-match with a translate of the axis of f is at least $1 - B_2 c_2^{\sqrt{n}}$. Therefore, for $K' = 2K + 2\delta$ we get for some $z \in \alpha_f$

$$\ker \phi \subseteq \mathrm{Stab}_K(p, w_n p) \subseteq g \mathrm{Stab}_{K'}(z, fz) g^{-1} \subseteq E_G^+(\Gamma_\mu) = E_\mu.$$

The result then holds for suitable choices of B and $c < 1$. □

Given $g \in G$ a loxodromic element, let us define the *fellow travelling constant* for g as

$$\Delta(g) := \sup_{h \in G \setminus E(g)} \Delta(g, hgh^{-1})$$

where $E(g)$ is the maximal elementary subgroup which contains g.

Definition 11.6. ([DGO17, Definition 6.25]) Let X be a δ-hyperbolic space with $\delta > 0$, and let $\mathcal{R}$ be a family of loxodromic isometries of X which is closed under conjugation. Then we say that $\mathcal{R}$ satisfies the (A, ϵ)-small cancellation condition if the following holds:

(1) $\mathrm{inj}(\mathcal{R}) \geqslant A\delta$
(2) $\Delta(g, h) \leqslant \epsilon \cdot \mathrm{inj}(\mathcal{R})$ for all $g \neq h^{\pm 1} \in \mathcal{R}$.

We will now prove that the cyclic subgroup generated by a power of w_n satisfies the small cancellation condition. First of all, we show that the fellow travelling constant between translates of the axis is sublinear in n.

Proposition 11.7. *Let G be a group of isometries of a δ-hyperbolic metric space X, and μ a countable, non-elementary, reversible, bounded, WPD probability measure on G. Let $\ell > 0$ be the drift of the random walk. Then for any $0 < \epsilon < 1$, there are constants B and $c < 1$ such that for all n the fellow travelling constant of w_n satisfies*

$$\mathbb{P}(\Delta(w_n) \geqslant \epsilon \ell n) \leqslant B c^{\sqrt{n}}.$$

Proof. By Proposition 11.1, there is an L such that $N_{20\delta}(\mathrm{Ax}(w_n)) \subset N_{L/2}(\alpha_{w_n})$. Therefore, if $\Delta(w_n) \geqslant \epsilon \ell n$, there is a translate $h\alpha_{w_n}$, with $h \notin E(w_n)$, such that α_{w_n} and $h\alpha_{w_n}$ have a $(\epsilon \ell n, L)$-match. This by definition means that there is a segment $\eta = [p, q] \subseteq \alpha_{w_n}$ with $|\eta|$ equal to $\epsilon \ell n$, such that $h\eta$ is contained in an L-neighbourhood of α_{w_n}. By replacing η with $w_n^i \eta$ for some $i \in \mathbb{Z}$ and replacing ϵ by $\epsilon/2$, we can assume that $\eta \subseteq [x', w_n x']$ where x' is a closest point projection of the basepoint x to α_{w_n}.

By Proposition 9.2, there are constants B_1 and $c_1 < 1$ such that the element w_n is (ϵ, L, K)-asymmetric with probability at least $1 - B_1 c_1^{\sqrt{n}}$. Thus there is a K, depending on ϵ and L, such that up to replacing h by $w_n^j h$ for some $j \in \mathbb{Z}$, we may assume that $d(p, hp) \leqslant K$ and $d(q, hq) \leqslant K$.

Let f be given an in the second part of Proposition 11.4. As $[p, q]$ has length $\epsilon\ell n$ and is contained in $[x', w_n x']$, by Lemma 7.7 there are constants B_2 and $c_2 < 1$ such that the probability that it contains a match with a large subsegment of a translate $g\alpha_f$ of the axis α_f (where $g \in \Gamma_\mu$) is at least $1 - B_2 c_2^{\sqrt{n}}$.

As h K-coarsely stabilizes this subsegment, this implies that there exists $z \in \alpha_f$ such that by Proposition 11.4,

$$h \in \mathrm{Stab}_K(gz, gfz) = g\mathrm{Stab}_K(z, fz)g^{-1} \subseteq gE_G^+(\Gamma_\mu)g^{-1} = E_G^+(\Gamma_\mu),$$

hence, since by construction $E_G^+(\Gamma_\mu) \subseteq E_G^+(w_n)$ and, by Corollary 11.2, there are constants B_3 and $c_3 < 1$ such that the probability that $E_G^+(w_n) = E_G(w_n)$ is at least $1 - B_3 c_3^{\sqrt{n}}$. Therefore, by suitable choices of B and $c < 1$, any such h must lie in $E_G(w_n)$ with probability at least $1 - Bc^{\sqrt{n}}$. However, this contradicts our initial choice of h, and implies that $\Delta(w_n) \geqslant \epsilon\ell n$ with probability at most $Bc^{\sqrt{n}}$, as required. $\square$

11.1. **The structure of the normal closure.** The last step we need to understand the structure of the normal closure $\langle\langle w_n \rangle\rangle$ of w_n in G is to take care of the fact that the elementary subgroup $E_G^+(w_n)$ need not be cyclic, so we may have to pass to a power of w_n. However, the power may be chosen to be a constant which only depends on G and μ, as we now explain.

Let Γ_μ be the group generated by the support of μ, and let $E_\mu := E_G^+(\Gamma_\mu)$. By definition, E_μ is a normal subgroup of Γ_μ, hence one has the homomorphism

$$\varphi : \Gamma_\mu \to \mathrm{Aut}\ E_\mu \tag{17}$$

given by conjugation: $g \mapsto (k \mapsto gkg^{-1})$. We will denote as $H_\mu := \varphi(\Gamma_\mu)$ the image of φ.

Lemma 11.8. *The image of φ in* Aut E_μ *is trivial if and only if $E_\mu = Z(\Gamma_\mu)$.*

Proof. First note that $Z(\Gamma_\mu) \subseteq E_\mu$. In fact, let $g \in Z(\Gamma_\mu)$ and let $h \in \Gamma_\mu$ be a loxodromic, WPD element. Then $ghg^{-1} = h$, hence $\mathrm{Fix}(ghg^{-1}) = g\mathrm{Fix}(h) = \mathrm{Fix}(h)$, hence $g \in E_G(h)$. Since this is true for any h WPD, then $g \in E_\mu$.

Moreover, the kernel of φ is the set of g which commute with every element of E_μ, hence the image is trivial if and only if every element of E_μ commutes with every element of Γ_μ, which means that $E_\mu \subseteq Z(\Gamma_\mu)$. $\square$

Now, by Corollary 11.5, with probability which tends to 1, $E_G(w_n)$ is the semidirect product

$$E_G(w_n) = \langle w_n \rangle \ltimes E_\mu$$

and the group structure of $E_G(w_n)$ is determined by the map $\langle w_n \rangle \to \mathrm{Aut}\ E_\mu$, hence by the image $\varphi(w_n)$ in Aut E_μ.

Lemma 11.9. *Let K be a finite group, let $\psi \in$ Aut K, and consider the semidirect product*

$$H = \mathbb{Z} \ltimes_\psi K$$

where we denote as t a generator for $\mathbb{Z}$, so that $tkt^{-1} = \psi(k)$ for any $k \in K$. Then:

1. *for any $a \in \mathbb{Z} \setminus \{0\}$, if $\psi(t^a) = 1$, then the normal closure of t^a in H is cyclic and equal to $\langle t^a \rangle$;*
2. *if $\psi(t) \neq 1$, then the normal closure of t in H is not cyclic and not free;*

Proof. Let $u = t^a$, and suppose that $\psi(u) = 1$. Then for any $k \in K$ we have $kuk^{-1} = u$ and since by construction u commutes with t, then u commutes with H, hence the normal closure $\langle\langle u \rangle\rangle = \langle u \rangle$ is infinite cyclic.

Now, since H is virtually cyclic and the subgroup of a free group is free, then the normal closure $N := \langle\langle t \rangle\rangle$ is free if and only if it is infinite cyclic. Moreover, since t generates $\mathbb{Z}$, the only cyclic group which contains $\langle t \rangle$ is $\langle t \rangle$ itself. Hence $\langle\langle t \rangle\rangle$ is free if and only if it coincides with $\langle t \rangle$. If the image $\phi(t)$ is not trivial, then there exists $k \in K$ such that $ktk^{-1} \neq t$, hence the normal closure is larger than $\langle t \rangle$, hence not free. □

Lemma 11.10. *Let $h \in G$ be a loxodromic, WPD element, and let $g \in G$. Then if $ghg^{-1} \in E_G(h)$, then $g \in E_G(h)$.*

Proof. Suppose that $ghg^{-1} \in E_G(h)$, and let $\Lambda := \{\lambda^+, \lambda^-\}$ be the set of fixed points of h on ∂X. Then by the assumption ghg^{-1} also fixes Λ, hence by conjugating h fixes $g^{-1}\Lambda$. Since h fixes exactly two points on the boundary, then $\Lambda = g^{-1}\Lambda$, which implies that $g \in E_G(h)$. □

We are now ready to present the main Theorem (Theorems 1.4 and 1.3) and its proof.

Theorem 11.11. *Let G be a group acting on a Gromov hyperbolic space X, and let μ be a countable, non-elementary, reversible, bounded, WPD probability measure on G. Let us denote as H_μ the image of Γ_μ in $\mathrm{Aut}\ E_\mu$. Then:*

(1) the probability that the normal closure $\langle\langle w_n \rangle\rangle$ of w_n in G is free satisfies
$$\mathbb{P}(\langle\langle w_n \rangle\rangle \text{ is free}) \to \frac{1}{\# H_\mu}$$
as $n \to \infty$. As a corollary, this probability tends to 1 if and only if $E_\mu = Z(\Gamma_\mu)$.

(2) Moreover, if $k = \# H_\mu$, then
$$\mathbb{P}(\langle\langle w_n^k \rangle\rangle \text{ is free}) \to 1$$
as $n \to \infty$, and indeed there exist constant $B > 0, c < 1$ such that
$$\mathbb{P}(\langle\langle w_n^k \rangle\rangle \text{ is free}) \geqslant 1 - Bc^{\sqrt{n}}$$
for any n.

(3) Finally, if $N_n := \langle\langle w_n^k \rangle\rangle$, then for any $R > 0$ the injectivity radius of N_n satisfies for any n
$$\mathbb{P}(\mathrm{inj}(N_n) \geqslant R) \geqslant 1 - Bc^{\sqrt{n}}.$$

Proof. Let us choose $\alpha > 0$. Then by [DGO17, Proposition 6.23] there exist constants (A, ϵ) such that if a family $\{N_\lambda\}_{\lambda\in\Lambda}$ of subgroups, closed under conjugation, satisfies the small cancellation condition, then $\{N_\lambda\}$ is α-rotating on a hyperbolic graph X'. Note that X' is obtained from X in the following way. First, one chooses a hyperbolic graph X'' which is equivariantly quasi-isometric to X. This is chosen once and for all; let K be the Lipschitz constant of the map $X \to X''$. Now, the coned off space X' is obtained by coning off certain quasi-convex subsets of a rescaled copy $\lambda X''$. However, by looking at the proof one realizes that one can make sure that $\lambda \leqslant 1$ in all cases (indeed, in the language of [DGO17, Proposition 6.23], the correct choice is $\lambda = \min\left(\frac{\delta_c}{\delta}, \frac{\Delta_c}{\Delta}, 1\right)$, with $A = \max\left(\frac{\mathrm{inj}_c(r_0)}{\delta_c}, \frac{\mathrm{inj}_c(r_0)}{\delta}\right)$ and

$\epsilon = \frac{\Delta_c}{\mathrm{inj}_c(r_0)}$.) Thus, the map $X \to X'$ is K-Lipschitz, where K only depends on X and not on the constant α.

Let us fix $\alpha \geqslant 200$, and let (A, ϵ) chosen as above. Let $\ell > 0$ be the drift of the random walk. Then by Theorem 2.5 (3), there are constants B_1 and $c_1 < 1$ such that

$$\mathbb{P}\left(\tau(w_n) \geqslant \frac{\ell n}{2}\right) \geqslant 1 - B_1 c_1^n.$$

Moreover, by Proposition 11.7, there are constants B_2 and $c_2 < 1$ such that

$$\mathbb{P}\left(\Delta(w_n) \leqslant \frac{\epsilon \ell n}{2}\right) \geqslant 1 - B_2 c_2^{\sqrt{n}}.$$

Now by Corollary 11.5, there are constants B_3 and $c_3 < 1$ such that

$$\mathbb{P}\left(E_G^+(w_n) = \langle w_n \rangle \ltimes E_\mu\right) \geqslant 1 - B_3 c_3^{\sqrt{n}}.$$

Thus, for suitable choices of B_4 and $c_4 < 1$,

$$\mathbb{P}\left(\tau(w_n) \geqslant A\delta, \Delta(w_n) \leqslant \epsilon\tau(w_n) \text{ and } E_G^+(w_n) = \langle w_n \rangle \ltimes E_\mu\right) \geqslant 1 - B_4 c_4^{\sqrt{n}}. \quad (18)$$

In particular, with probability which tends to 1 we have

$$E_G(w_n) = \langle w_n \rangle \ltimes_{\varphi_n} E_\mu$$

where $\varphi_n = \varphi(w_n)$ is the image of w_n under the homomorphism

$$\varphi : \Gamma_\mu \to \mathrm{Aut}\ E_\mu.$$

Now, we have two cases.

(1) if $\varphi(w_n) = 1$, then all conjugates of w_n in G belong to different elementary subgroups.

In fact, suppose that there exists $g \in G$ such that $g w_n g^{-1} \in E_G(g)$. Then, by Lemma 11.10 one has $g \in E_G(w_n)$, and by Lemma 11.9 one has $g w_n g^{-1} = w_n$.

Now, consider the family of subgroups $\mathcal{R}_n := \{g w_n g^{-1}\}_{g \in G}$. Finally, let $N_n = \langle\langle H_n \rangle\rangle$ be the normal closure of H_n. By equation (18) above, with probability at least $1 - B_4 c_4^{\sqrt{n}}$, the family $\mathcal{R}_n$ satisfies the (A, ϵ)-small cancellation condition, hence it is an α-rotating family. Then by [DGO17, Corollary 5.4], the normal closure of w_n is the free product of conjugates of $\langle w_n \rangle$, hence it is free.

(2) if $\varphi(w_n) \neq 1$, then there exists $g \in \Gamma_\mu$ such that $g w_n g^{-1} \neq w_n$. This implies that the intersection

$$\langle\langle w_n \rangle\rangle \cap E_G(w_n)$$

is larger than $\langle w_n \rangle$, hence the normal closure $\langle\langle w_n \rangle\rangle$ cannot be a free group.

By the above discussion, the probability that the normal closure of w_n in G is free converges to the probability that w_n maps to the identity in E_μ. In order to compute such probability, note that under the map

$$\varphi : \Gamma_\mu \to \mathrm{Aut}\ E_\mu$$

the random walk on Γ_μ pushes forward to a random walk on Aut E_μ, which is a finite group. Hence, the random walk equidistributes on the elements of the image of φ into Aut E_μ, hence the probability that $\varphi(w_n) = 1$ converges to $\frac{1}{\#H_\mu}$, where $\#H_\mu$ is the cardinality of the image of φ. That is, the normal closure of w_n is free if and only if the image $\varphi(w_n) = 1$, and the probability of this happening tends to $\frac{1}{\# H_\mu}$, so

$$\mathbb{P}(\langle\langle w_n \rangle\rangle \text{ is free}) \to \frac{1}{\# H_\mu}.$$

Hence, this probability tends to 1 if and only if the image group $H_\mu = \varphi(\Gamma_\mu)$ is the trivial group, hence by Lemma 11.8 if and only if $E_\mu = Z(\Gamma_\mu)$.

To prove (ii), if $k = \# H_\mu$, then every element in the image of φ has order which divides k, hence $\varphi(w_n^k) = \varphi(w_n)^k = 1$. Thus, as in the previous argument, if one defines $H_n := \langle w_n^k \rangle$, the probability that the family $\mathcal{R}_n := \{g w_n^k g^{-1}\}_{g \in G}$ satisfies the small cancellation condition tends to 1, hence the probability that the normal closure $N_n := \langle\langle w_n^k \rangle\rangle$ is free satisfies

$$\mathbb{P}(\langle\langle w_n^k \rangle\rangle \text{ is free}) \geqslant 1 - Bc^{\sqrt{n}}$$

for suitable choices of $B > 0, c < 1$.

Now, to prove (iii), given $R > 0$ let α be such that $\frac{\delta\alpha}{K} = R$. Then one can choose (A, ϵ) as before for such α. Then with probability at least $1 - B_4 c_4^{\sqrt{n}}$, the family $\mathcal{R}_n$ is α-rotating. Hence, by [DGO17, Theorem 5.3], for each $g \in N_n$, either g belongs to some conjugate of H_n or is loxodromic on X' with translation length at least $\alpha\delta$. Then since the map $X \to X'$ is K-Lipschitz, such elements have translation length on X at least $\frac{\alpha\delta}{K}$. On the other hand, by Theorem 2.5 (3) we know that with probability at least $1 - B_1 c_1^n$, the isometry w_n^k is loxodromic on X with translation length $\geqslant R$. Therefore for suitable choices of B_5 and $c_5 < 1$, the probability that the injectivity radius of N_n is at least R is at least $1 - B_5 c_5^{\sqrt{n}}$. The stated result then follows for suitable choices of B and $c < 1$. $\square$

Corollary 11.12. *Let G be a group acting on a Gromov hyperbolic space, and let μ be a countable, non-elementary, reversible, bounded, WPD probability measure on G. Then there is a constant $k = \# H_\mu$ such that if $N_n(\omega) := \langle\langle w_n^k \rangle\rangle$ is the normal closure of w_n^k in G, then for almost every sample path ω, the sequence*

$$(N_1(\omega), N_2(\omega), \dots, N_n(\omega), \dots)$$

contains infinitely many different normal subgroups of G.

Proof. Fix $M > 0$, and consider the set

$$A_M := \{\omega \ : \ \sup_n \operatorname{inj}(N_n(\omega)) \leqslant M\}.$$

We claim that $\mathbb{P}(A_M) = 0$. Indeed, suppose $\mathbb{P}(A_M) = \epsilon > 0$. Then by Theorem 11.11, there exists n_0 such that for $n \geqslant n_0$

$$\mathbb{P}(\operatorname{inj}(N_n) \geqslant M + 1) > 1 - \epsilon$$

which is a contradiction because such a set must be disjoint from A_M. Then for almost every ω we have

$$\limsup_{n \to \infty} \operatorname{inj}(N_n(\omega)) = +\infty,$$

which implies the claim. $\square$

This completes the proof of Theorem 1.3 in the Introduction.

11.2. **Application to the mapping class group.** In the case of the mapping class group, we may answer [Mar18, Problem 10.11] and establish Theorem 1.11, as we now explain.

Corollary 11.13. *Let S be a surface of finite type whose mapping class group* $\mathrm{Mod}(S)$ *is infinite. Let μ be a probability distribution on* $\mathrm{Mod}(S)$ *such that the support of μ has bounded image in the curve complex under the orbit map, and for which $\Gamma_\mu = \mathrm{Mod}(S)$. Then there are constants $B > 0$ and $c < 1$ such that the probability that the normal closure $\langle\langle w_n \rangle\rangle$ is a free subgroup of* $\mathrm{Mod}(S)$ *is at least* $1 - Bc^{\sqrt{n}}$.

This follows immediately from Theorem 11.11, and the fact that for $G = \mathrm{Mod}(S_{g,n})$ the mapping class group of a finite type surface, the group $E^+_G(G)$ is equal to the center of G, as we now explain.

We shall write $S_{g,n}$ for the surface of genus g with n punctures. The mapping class groups $S_{0,n}$ with $n \leqslant 3$ are finite, and so the results of this paper do not apply to them, and we shall ignore them for the purposes of this section. If the mapping class group is infinite, then the center of the mapping class group $\mathrm{Mod}(S_{g,n})$ is trivial, unless $S_{g,n}$ is one of the following four surfaces: $S_{1,0}, S_{1,1}, S_{1,2}$ or $S_{2,0}$, in which case the center is $\mathbb{Z}/2\mathbb{Z}$, see [FM12, Section 3.4].

Proposition 11.14. *Let $S_{g,n}$ be a surface of genus g with n punctures, and suppose that its mapping class group $G = Mod(S_{g,n})$ is infinite. Then $E^+_G(G)$ is equal to the center of G.*

Proof. To simplify notation, we shall write $E(G)$ for $E^+_G(G)$. If S is a surface whose mapping class group has trivial center, then by the Nielsen realization theorem, due to Kerckhoff [Ker83], any non-trivial finite subgroup F of $G = \mathrm{Mod}(S)$ is realized by a group of isometries of a hyperbolic metric on S, giving rise to an orbifold cover $S \to \overline{S}$. The fixed point set of F in Teichmüller space $\mathcal{T}(S)$ is a totally geodesically embedded image of $\mathcal{T}(\overline{S})$, which has strictly smaller dimension. In particular, there are elements of G which do not preserve the fixed point set of F, and so F is not normal in G. This implies that $E(G)$ is trivial for those mapping class groups with trivial center.

We now consider the cases in which $G = \mathrm{Mod}(S)$ is infinite with non-trivial center. As the center of G is finite, it is contained in $E(G)$, so for these surfaces $E(G)$ always contains $\mathbb{Z}/2\mathbb{Z}$. We now show that $E(G)$ is in fact equal to $\mathbb{Z}/2\mathbb{Z}$. We consider the mapping class group of each surface with non-trivial center in turn.

The quotient of the genus two surface $S_{2,0}$ by the hyperelliptic involution as an orbifold is a sphere with six cone points of angle π, but as the mapping class group acts transitively on these marked points, the quotient mapping class group is isomorphic to $\mathrm{Mod}(S_{0,6})$. As $E(\mathrm{Mod}(S_{0,6}))$ is trivial, this implies that $E(\mathrm{Mod}(S_{2,0}))$ is equal to $\mathbb{Z}/2\mathbb{Z}$.

The quotient $\overline{S}$ of the twice punctured torus $S_{1,2}$ by the hyperelliptic involution as an orbifold is a sphere with four cone points of angle π, and one puncture. The quotient of the mapping class group acts transitively on the cone points, and preserves the puncture, and so is isomorphic to an index five subgroup G of $\mathrm{Mod}(S_{0,5})$. If F is a finite subgroup of G, then again by the Nielsen realization theorem, F may be realized as a group of isometries of a hyperbolic metric on $\overline{S}$, and the fixed point set of F in $\mathcal{T}(\overline{S})$ is isometric to the Teichmüller space of the quotient orbifold, which has strictly smaller dimension. As endpoints of pseudo-Anosov elements are dense in $\mathcal{PMF}(\overline{S}) \times \mathcal{PMF}(\overline{S})$, there is an element of $\mathrm{Mod}(\overline{S})$ which does not preserve the fixed point set, and so F is not normal. Therefore $E(G)$ is trivial, and so $E(\mathrm{Mod}(S_{0,5}))$ is equal to the center $\mathbb{Z}/2\mathbb{Z}$.

Finally, the mapping class groups of $S_{1,0}$ and $S_{1,1}$ are equal to $\mathrm{SL}(2,\mathbb{Z})$. Quotienting out by the center gives $\mathrm{PSL}(2,\mathbb{Z})$, which is isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \star (\mathbb{Z}/3\mathbb{Z})$. This free product has a trivial maximal normal finite subgroup, so for both of these surfaces $E(\mathrm{Mod}(S))$ is equal to the center, $\mathbb{Z}/2\mathbb{Z}$. □

References

[BBF15] Mladen Bestvina, Kenneth Bromberg, and Koji Fujiwara, *Constructing group actions on quasi-trees and applications to mapping class groups*, Publ. math. IHES **122** (2015).

[BF14] Mladen Bestvina and Mark Feighn, *Hyperbolicity of the complex of free factors*, Adv. Math. **256** (2014), 104-155.

[BF02] Mladen Bestvina and Koji Fujiwara, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. **6** (2002), 69–89 (electronic).

[Bow08] Brian H. Bowditch, *Tight geodesics in the curve complex*, Invent. Math. **171** (2008), no. 2, 281–300.

[BH99] Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999.

[CM15] Danny Calegari and Joseph Maher, *Statistics and compression of scl*, Ergodic Theory Dynam. Systems **35** (2015), no. 1, 64–110.

[CL13] Serge Cantat and Stéphane Lamy, *Normal subgroups in the Cremona group*, Acta Math. **210** (2013), no. 1, 31-94.

[Cou13] Rémi Coulon, *Partial periodic quotient of groups acting on a hyperbolic space* (2013), available at `arXiv:1311.0855`.

[DGO17] François Dahmani, Vincent Guirardel, and Denis Osin, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. **245** (2017), no. 1156, v+152.

[DH15] François Dahmani and Camille Horbez, *Spectral theorems for random walks on mapping class groups and Out(F_N)*, Int. Math. Res. Not. (to appear) (2015), available at `arXiv:1506.06790`.

[DP12] Thomas Delzant and Pierre Py, *Kähler groups, real hyperbolic spaces and the Cremona group*, Compositio Mathematica **148** (2012), no. 1, 153–184.

[DF01] Jeffrey Diller and Charles Favre, *Dynamics of Bimeromorphic Maps of Surfaces*, American Journal of Mathematics **123** (2001), no. 6, 1135–1169.

[FM12] Benson Farb and Dan Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Princeton, NJ, 2012.

[Fav08] Charles Favre, *Le groupe the Cremona et ses sous-groupes de type fini*, Séminaire Bourbaki **998** (2008).

[Fri95] Shmuel Friedland, *Entropy of algebraic maps*, Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993), 1995, pp. 215–228.

[Fur63] Harry Furstenberg, *Noncommuting random products*, Trans. Amer. Math. Soc. **108** (1963), 377–428.

[FL10] Jean-Philippe Furter and Stéphane Lamy, *Normal subgroup generated by a plane polynomial automorphism*, Transf. Groups **15** (2010), 577-610.

[Giz80] Marat H. Gizatullin, *Rational G-surfaces*, Izv. Akad. Nauk SSSR Ser. Mat. **44** (1980), no. 1, 110–144, 239.

[GST] Dominik Gruber, Alessandro Sisto, and Romain Tessera, *Gromov's random monsters do not act non-elementarily on hyperbolic spaces*, available at `arXiv:1506.06790`.

[HM13] Michael Handel and Lee Mosher, *The free splitting complex of a free group, I: hyperbolicity*, Geometry & Topology **17** (2013), no. 3, 1581–1672.

[HM14] Michael Handel and Lee Mosher, *The free splitting complex of a free group, II: loxodromic outer automorphisms* (2014). Available at `arXiv:1402.1886`.

[Hor16] Camille Horbez, *The Poisson boundary of Out(F_N)*, Duke Math. J. **165** (2016), no. 2, 341–369.

[Kai00] Vadim A. Kaimanovich, *The Poisson formula for groups with hyperbolic properties*, Ann. of Math. (2) **152** (2000), no. 3, 659–692.

[Ker83] Steven P. Kerckhoff, *The Nielsen realization problem*, Ann. of Math. (2) **117** (1983), no. 2, 235–265.

[KK14] Sang-Hyun Kim and Thomas Koberda, *The geometry of the curve graph of a right-angled Artin group*, Internat. J. Algebra Comput. **24** (2014), no. 2, 121–169.

[LP16] Stéphane Lamy and Piotr Przytycki, *Acylindrical hyperbolicity of the three-dimensional tame automorphism group*, Ann. Sci. Ec. Norm. Sup. (to appear) (2016), available at `arXiv:1610.05457[math.GR]`.

[Lon16] Anne Lonjou, *Non simplicité du groupe de Cremona sur tout corps*, Ann. Inst. Fourier **66** (2016), no. 5, 2021–2046.

[Mah10a] Joseph Maher, *Random Heegaard splittings*, Journal of Topology **3** (2010), no. 4, 997–1025.

[Mah10b] Joseph Maher, *Linear progress in the complex of curves*, Transactions of the American Mathematical Society **362** (2010), no. 6, 2963–2991.

[Mah11] Joseph Maher, *Random walks on the mapping class group*, Duke Math. J. **156** (2011), no. 3, 429–468.

[Mah12] Joseph Maher, *Exponential decay in the mapping class group*, Journal of the London Mathematics Society **86** (2012), no. 2, 366–386.

[MS18] Joseph Maher and Alessandro Sisto, *Random subgroups of acylindrical groups and hyperbolic embeddings*, Int. Math. Res. Not. (to appear) (2018).

[MT16] Joseph Maher and Giulio Tiozzo, *Random walks on weakly hyperbolic groups*, J. Reine Angew. Math. (published online, to appear in print) (2016).

[Mar18] Dan Margalit, *Problems, questions and conjectures about mapping class groups* (2018), available at `arXiv:1806:08773`.

[MM99] Howard A. Masur and Yair N. Minsky, *Geometry of the complex of curves. I. Hyperbolicity*, Invent. Math. **138** (1999), no. 1, 103–149.

[MO15] A. Minasyan and D. Osin, *Acylindrical hyperbolicity of groups acting on trees*, Math. Ann. **362** (2015), no. 3-4, 1055–1105.

[Neu51] Bernhard H. Neumann, *Groups with finite classes of conjugate elements*, Proc. London Math. Soc. (3) **1** (1951), 178–187.

[Neu54] Bernhard H. Neumann, *Groups covered by permutable subsets*, J. London Math. Soc. **29** (1954), 236–248.

[Osi16] Denis Osin, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. **368** (2016), 851-888.

[Riv08] Igor Rivin, *Walks on groups, counting reducible matrices, polynomials, and surface and free group automorphisms*, Duke Math. J. **142** (2008), no. 2, 353–379.

[Roh52] Vladimir A. Rohlin, *On the fundamental ideas of measure theory*, Amer. Math. Soc. Translation **1952** (1952), no. 71, 55.

[RS97] Alexander Russakovskii and Bernard Shiffman, *Value distribution for sequences of rational mappings and complex dynamics*, Indiana Univ. Math. J. **46** (1997), no. 3, 897–932.

[Sel97] Zlil Sela, *Acylindrical accessibility for groups*, Invent. Math. **129** (1997), no. 3, 527–565.

[SB13] N.I. Shepherd-Barron, *Some effectivity questions for plane Cremona transformations* (2013), available at `arXiv:1311.6608[math.AG]`.

[Sis16] Alessandro Sisto, *Contracting elements and random walks*, J. Reine Angew. Math. (published online, to appear in print) (2016).

[TT16] Samuel J. Taylor and Giulio Tiozzo, *Random free and surface group extensions are hyperbolic*, Int. Math. Res. Not. **2016** (2016), no. 1, 294–310.

[Ure18] Christian Urech, *Simple groups of birational transformations in dimension two* (2018), available at `arXiv:1802.09258[math.AG]`.

Joseph Maher
CUNY College of Staten Island and CUNY Graduate Center
`joseph.maher@csi.cuny.edu`

Giulio Tiozzo
University of Toronto

tiozzo@math.toronto.edu