CONTRACTING ISOMETRIES AND DIFFERENTIABILITY OF THE ESCAPE RATE

INHYEOK CHOI

Abstract. Let G be a countable group whose action on a metric space X involves a contracting isometry. This setting naturally encompasses groups acting on Gromov hyperbolic spaces, Teichmüller space, Culler-Vogtmann Outer space and CAT(0) spaces. We discuss continuity and differentiability of the escape rate of random walks on G. For mapping class groups, relatively hyperbolic groups, CAT(-1) groups and CAT(0) cubical groups, we further discuss analyticity of the escape rate. Finally, assuming that the action of G on X is weakly properly discontinuous (WPD), we discuss continuity of the asymptotic entropy of random walks on G.

Keywords. Random walk, escape rate, entropy, CAT(0) space, Outer space

MSC classes: 20F67, 30F60, 57M60, 60G50

1. Introduction

{section:intro}

Let G be a countable group acting on a metric space (X, d) with a basepoint $o \in X$ and let μ be a probability measure on G. The *first moment* and the *(time-one) entropy* of μ is defined respectively by

$$L(\mu) := \sum_{g \in G} \mu(g) d(o, go), \quad H(\mu) := \sum_{g \in G} \mu(g) \log \mu(g).$$

Since these quantities are subadditive with respect to convolution of measures, we can define two asymptotic quantities

$$l(\mu) := \lim_{n \to \infty} \frac{L(\mu^{*n})}{n}, \quad h(\mu) := \lim_{n \to \infty} \frac{H(\mu^{*n})}{n},$$

called the *escape rate* (or *drift*) and the *asymptotic* (or *Avez*) *entropy*. These quantities have been investigated in connection with the analytic properties of the ambient group. For example, a countable group G is nonamenable if and only if every admissible measure μ on G has strictly positive escape rate $l(\mu) > 0$ [KV83]. Moreover, for a probability measure μ on G with finite one-time entropy, its asymptotic entropy $H(\mu)$ vanishes if and only if the Poisson boundary for (G, μ) is trivial, i.e., there is no non-constant μ-harmonic bounded function on G (see [Ave72], [Der80] and [KV83]; see also [KL07] for the role of $L(\mu)$ in this problem).

Date: January 12, 2024.

Furthermore, these two quantities tell us how the random walk probes the G-orbit on X. When $L(\mu)$ is finite, almost every sample path $(Z_n)_{n>0}$ escapes to infinity with rate $l(\mu)$, i.e., $l(\mu) = \lim_n d(o, Z_n o)/n$. Similarly, when $H(\mu)$ is finite, almost every sample path $(Z_n)_{n>0}$ arrives at an $\exp(-nh(\mu))$-probable orbit point at time n, i.e., $h(\mu) = -\lim_n \log \mu^{*n}(Z_n)/n$. Finally, the volume growth $v := \limsup_n \ln(\#\{g \in G : d(o, go) < n\})/n$ tells us how many orbit points there are within a bounded distance. The Guivarc'h fundamental inequality ([Gui80], [BHM08]) tells us that

$$h \leq lv,$$

and for word hyperbolic groups, the equality holds if and only if the limiting distributions for the μ-random walk (the μ-harmonic measure) and the orbit counting (the Patterson-Sullivan measure) are equivalent [BHM11].

It is natural to ask the dependence of the quantities $h(\mu)$ and $l(\mu)$ on the underlying measure μ. Indeed, Gouëzel, Mathéus and Maucourant proved in [GMM18] that given a finite subset $S \subseteq G$ of a word hyperbolic group, there exists a constant $C < 1$ such that $h(\mu) \leq Cl(\mu)v$ holds for all symmetric probability measure μ supported in S. One ingredient of their argument is the continuity of $h(\mu)$ and $l(\mu)$ with respect to the underlying measure μ.

In this paper, we focus on group actions that involve strongly contracting isometries. We say that a probability measure μ on G is *non-elementary* if the semigroup generated by the support of μ contains two independent contracting isometries of X. We say that a sequence of measures $\{\mu_i\}_{i>0}$ on G *converges simply* to μ if $\mu_i(g) \to \mu(g)$ for each $g \in G$. We recall the continuity of the escape rate:

{thm:driftContinuity}

Theorem 1.1 ([Gou22, Proposition 5.15])**.** *Let G be a countable group acting on a metric space (X, d) that involves two independent contracting isometries. Let μ be a non-elementary probability measure on G, and let $\{\mu_i\}_{i>0}$ be a sequence of probability measures that converges simply to μ_0 with $L(\mu_i) \to L(\mu)$. Then $l(\mu_i)$ converges to $l(\mu)$.*

This result is originally due to Erschler and Kaimanovich for hyperbolic groups [KÈ13] and due to Mathieu and Sisto for acylindrically hyperbolic groups and Teichmüller space [MS20]. See also [Mas21] for another approach to the case of Teichmüller space. Gouëzel reproved this result for non-elementary isometry groups of Gromov hyperbolic spaces, but his argument applies to group actions involving contracting isometries (see [Cho22a] for necessary modification).

Hugely motivated by the works of Gouëzel [Gou22], Mathieu [Mat15] and Mathieu and Sisto [MS20], we aim to investigate higher regularity of the escape rate on various spaces. Given a signed measure η on G, we define

$$\|\eta\|_0 := \sum_{g\in G} |\eta(g)| \quad \text{(total variation)},$$

$$\|\eta\|_1 := \sum_{g\in G} d(o, go)|\eta(g)| \quad \text{(first moment)},$$

$$\|\eta\|_{0,1} := \|\eta\|_0 + \|\eta\|_1 = \sum_{g\in G} (d(o, go) + 1)|\eta(g)|.$$

Our main results are as follows.

{thm:driftLipschitz}

Theorem A. *Let G be a countable group acting on a metric space (X, d) that involves two independent contracting isometries, and let μ be a non-elementary probability measure on G with finite first moment. Then there exists $C, \epsilon > 0$ such that*

$$\left|l(\mu') - l(\mu)\right| \le C\|\mu' - \mu\|_{0,1}$$

for any probability measure μ' such that $\|\mu' - \mu\|_{0,1} < \epsilon$.

{thm:driftDiff}

Theorem B. *Let G be a countable group acting on a metric space (X, d) that involves two independent contracting isometries. Let μ be a non-elementary probability measure on G with finite first moment, let η be a signed measure that is absolutely continuous with respect to μ and such that $\|\eta\|_{0,1} < \infty$, and let $\{\mu_t : t \in [-1, 1]\}$ be a family of probability measures such that*

$$\|\mu_t - \mu - t\eta\|_{0,1} = o(t).$$

Then $l(\mu_t)$ is differentiable at $t = 0$ with the derivative

$$\lim_n \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\mu^{*i-1}*\eta*\mu^{*n-i}}\, d(o, go).$$

Moreover, this derivative is continuous with respect to μ and η.

Analyticity of the escape rate of random walks on free groups and free products was discussed independently by Ledrappier and Gilch (see [Led01], [Gil07], [LL12]). Haïssinsky, Mathieu and Müller proved analyticity of the escape rate of random walks on surface groups by using automatic structures [HMM18]. On Gromov hyperbolic groups, Lipschitz continuity and C^1-regularity of the escape rate is due to Ledrappier [Led13] and Mathieu [Mat15], respectively. These results are generalized to acylindrically hyperbolic groups and Teichmüller spaces by Mathieu and Sisto [MS20]. Finally, Gouëzel established analyticity of the escape rate on hyperbolic groups in [Gou17]. Theorem A and B generalize Mathieu-Sisto's differentiability of the escape rate to CAT(0) spaces and Culler-Vogtmann Outer space.

For certain examples, we have a different formula for the derivative of the escape rate.

{thm:driftDiffSqueeze}

Theorem C. *Let (X, G) be either:*

- *Teichmüller space (with the Teichmüller metric or the Weil-Petersson metric) and the mapping class group;*
- *Culler-Vogtman Outer space CV_N and $\mathrm{Out}(F_N)$;*
- *a CAT(-1) space and its countable isometry group;*
- *a CAT(0) cube complex and its countable isometry group;*
- *relatively hyperbolic group equipped with a Green metric;*
- *the standard Cayley graph of a surface group, or;*
- *the standard Cayley graph of a free product of nontrivial groups.*

Let μ be a non-elementary probability measure on G with finite first moment, let η be a signed measure on G such that $\|\eta\|_{0,1} < \infty$, and let $\{\mu_t : t \in [-1,1]\}$ be a family of probability measures such that

$$\|\mu_t - \mu - t\eta\|_{0,1} = o(t).$$

Then $l(\mu_t)$ is differentiable at $t = 0$, with derivative

$$\sigma_1(\mu,\eta) := \mathbb{E}_{\eta}\, d(o,go) - \lim_{n\to\infty} \mathbb{E}_{\mu^{*n}\times\eta\times\mu^{*n}} \big[d(o,g_1o)+d(o,g_2o)+d(o,g_3o)-d(o,g_1g_2g_3o)\big].$$

Moreover, this derivative is continuous with respect to μ and η.

When the underlying space exhibit a CAT(-1)-behaviour near an axis of a contracting isometry, we can promote this C^1-regularity of the drift to its analyticity. For simplicity, we here present a result about C^2-regularity.

riftDiffSqueezeSecond}

Theorem D. *Let (X, G) be either:*

- *the Weil-Petersson Teichmüller space and the mapping class group;*
- *a CAT(-1) space and its discrete isometry group;*
- *a CAT(0) cube complex and its countable isometry group;*
- *relatively hyperbolic group equipped with a Green metric;*
- *the standard Cayley graph of a surface group, or;*
- *the standard Cayley graph of a free product of nontrivial groups.*

Then the function $\sigma_1(\mu,\eta)$ is differentiable in the following sense. Let μ be a non-elementary probability measure on G with finite first moment and let η, η' be signed measures on G such that $\|\mu\|_{0,1}, \|\eta\|_{0,1} < \infty$. Let $\{\mu_t : t \in [-1,1]\}$ be a family of probability measures and $\{\eta_t : t \in [-1,1]\}$ be a family of signed measures such that

$$\|\mu_t - \mu - t\eta\|_{0,1},\ \|\eta_t - \eta - t\eta'\|_{0,1} = o(t).$$

Then $\sigma_1(\mu_t,\eta_t)$ is differentiable at $t = 0$, with derivative

$$\begin{aligned}
\sigma_2(\mu,\eta,\eta') &:= \mathbb{E}_{\eta'}\, d(o,go) - \lim_{n\to\infty} \mathbb{E}_{\mu^{*n}\times\eta'*\mu^{*n}} \big[d(o,g_1o)+d(o,g_2o)+d(o,g_3o)-d(o,g_1g_2g_3o)\big] \\
&- \sum_{i=1}^{\infty} \lim_{n\to\infty} \mathbb{E}_{\mu^{*n}\times\eta\times\left(\mu^{*(i-1)}*\eta*\mu^{*(n-i)}\right)} \big[d(o,g_1o)+d(o,g_2o)+d(o,g_3o)-d(o,g_1g_2g_3o)\big] \\
&- \sum_{i=1}^{\infty} \lim_{n\to\infty} \mathbb{E}_{\left(\mu^{*(n-i)}*\eta*\mu^{*(i-1)}\right)\times\eta\times\mu^{*n}} \big[d(o,g_1o)+d(o,g_2o)+d(o,g_3o)-d(o,g_1g_2g_3o)\big].
\end{aligned}$$

Moreover, this derivative is continuous with respect to μ, η and η'.

As mentioned, analyticity of the escape rate on Gromov hyperbolic groups is due to Gouëzel [Gou17]. Gouëzel's strategy is to utilize the contracting property of the Markov operator on a subset of the Busemann boundary of G. A crucial ingredient of his argument is that the values of a Busemann function h on a bounded ball can determine the values of h on a larger region ([Gou17, Lemma 3.4]). This phenomenon is expected if the underlying space is either a Gromov hyperbolic graph or a CAT(-1) space. It is not known whether such a strategy can be implemented for G-action on a general Gromov hyperbolic space.

Meanwhile, even restricted to the graph metric, our strategy requires much stronger (yet not global) rigidity of Busemann functions. Recall that for a hyperbolic group G, there is a continuous, (uniformly) finite-to-one surjection $\pi : \partial_B G \to \partial G$ from the Busemann boundary to the Gromov boundary. Theorem D applies to the Cayley graph of a hyperbolic group G that contains a loxodromic element g whose attracting point $g^+ \in \partial G$ is a 1-1 point of π, i.e., $\pi^{-1}(g^+)$ consists of a single horofunction. This holds on the standard Cayley graph of a surface group, for example, but the existence of such an element g in every hyperbolic group is not known.

Another difference between Gouëzel's method and ours is the condition on the underlying measure μ. Gouëzel's method investigates spectral property of the Markov operator which requires μ to be admissible (i.e., the support of μ generates entire G) and have finite support (or at least with exponential tail). In contrast, we do not make use of spectral theory and merely rely on the deviation inequality, our theorem applies to random walks that are not admissible (but non-elementary) nor finitely supported (but with finite first moment).

Our method provides the following effective bound:

{thm:quantitative}

Theorem E. *Let (X, G) be as in Theorem D and suppose that $\mu(g), \mu(h) > 0$ for some independent contracting isometries $g, h \in G$. Then for each $\epsilon_0 > 0$ there exists a threshold*

$$N = N\left(\epsilon_0, g, h, \mu(g), \mu(h), \|\mu\|_{0,1}, \|\eta\|_{0,1}\right)$$

such that

$$\left|\sigma_1(\mu, \eta) - \left(\mathbb{E}_\eta\, d(o, go) - \mathbb{E}_{\mu^{*N} \times \eta \times \mu^{*N}}[d(o, g_1 o) + d(o, g_2 o) + d(o, g_3 o) - d(o, g_1 g_2 g_3 o)]\right)\right| < \epsilon_0.$$

Also, when the other parameters are fixed, we have $N(\epsilon_0) = O(|\log \epsilon_0|)$.

Similarly, the second derivative $\sigma_2(\mu, \eta, \eta')$ in Theorem D can be calculated up to error ϵ by computing finitely many terms, where the threshold is determined by the nature of g and h, how μ puts nonzero weights on g and h, and the values of $\|\mu\|_{0,1}$, $\|\eta\|_{0,1}$ and $\|\eta'\|_{0,1}$.

We next turn to the continuity of the asymptotic entropy of μ. Note that the theorem deals with general non-elementary probability measures, without any moment condition. To the best of the author's knowledge, this is new even on the free groups.

{thm:entropyContinuity

Theorem F. *Let G be a countable group acting on a metric space (X, d) with a WPD contracting isometry $g \in G$, let μ be a probability measure such that $\mu^{*n}(g^m) > 0$ for some $n, m > 0$, and let $\{\mu_i\}_{i>0}$ be a sequence of probability measures that converges simply to μ with $H(\mu_i) \to H(\mu)$. Then $h(\mu_i)$ converges to $h(\mu_0)$.*

This continuity of the asymptotic entropy on hyperbolic groups was proven by Erschler and Kaimanovich [KÈ13] when the underlying measure μ has finite first moment, and later by Gouëzel, Mathéus and Maucourant when μ has finite logarithmic moment [GMM18]. See [Mas21] for the case of Teichmüller space. While preparing the current manuscript, the author was informed by Anna Erschler and Joshua Frisch of their independent proof of Theroem F.

For higher regularity (differentiability, analyticity, etc.) of the asymptotic entropy on hyperbolic groups, see [Gil11], [LL12], [Led13], [Mat15], [Gou17] and [MS20]. On hyperbolic groups and acylindrically hyperbolic groups, the known proofs of differentiability of the asymptotic entropy rely on the Martin boundary or the Green metric of G, and the Ancona inequality lies at the heart of the argument. Consequently, the underlying measures are subject to some moment condition (finite support or exponential tail). It seems very challenging to prove an analogue of Theorem B for asymptotic entropy, with finite entropy condition but without moment condition.

We prove Theorem F by combining Gouëzel's pivotal time technique in [Gou22] (modulo small modification as in [Cho22a]) and Chawla-Frisch-Forghani-Tiozzo's sublinear entropy growth of displacement [CFFT22].

In the first half of Section 2, we list the notions related to contracting isometries and random walks that are borrowed from [Cho22a]. In the latter half of Section 2, we summarize the geometric features of metric spaces possessing contracting isometries, including Teichmüller space, CAT(0) spaces, CAT(0) cube complices, CAT(-1) spaces, Culler-Vogtmann Outer space and (relatively) hyperbolic groups. In Section 3, we establish a single probabilistic estimate (Lemma 3.1) that leads to Theorem A and B. Once Lemma 3.1, the remaining argument for Theorem A and B has nothing to do with contracting isometries and works for general random walks on groups. In Section 4, similarly, we single out a probabilistic estimate (Proposition 4.4) that leads to Theorem C. In Section 5, a slightly more delicate estimate (Proposition 5.2) is stated and Theorem D is deduced. In Section 6, we establish Proposition 4.4 and 5.2 using Gouëzel's pivoting technique [Gou22], as sketched in [Cho22a]. In Section 7, we recall Chawla-Forghani-Frisch-Tiozzo's sublinear growth of entropy of displacement [CFFT22] and a more complicated version of pivotal time construction [Gou22, Section 5C]. By combining them, we establish Theorem F.

Acknowledgments. The author thanks Kunal Chawla, Joshua Frisch, Ilya Gekhtman and Hidetoshi Masai for helpful discussion. The author is also grateful to the American Institute of Mathematics and the organizers and the participants of the workshop "Random walks beyond hyperbolic groups" in April 2022 for helpful and inspiring discussions.

The author was supported by Samsung Science & Technology Foundation (SSTF-BA1301-51), and by a KIAS Individual Grant (SG091901) via the June E Huh Center for Mathematical Challenges at Korea Institute for Advanced Study.

2. Preliminaries {section:prelim}

In this section we review geometric and probabilistic preliminaries. Subsection 2.1, 2.2 and 2.3 summarize the essence of [Cho22a], regarding random walks and contracting isometries. These subsections are the only prerequisites for the proof of Theorem A, B and F. The remaining subsections deal with individual spaces that appear in Theorem C and D.

{subsection:contractin

2.1. **Contracting isometries.** Let X be a geodesic metric space, i.e., a metric space whose every pair of points can be connected by a geodesic. For a closed set $A \subseteq X$, we mean by $\pi_A(\cdot)$ the closest point projection onto A. We say that two paths γ and η are *K-fellow traveling* if $d_{\mathrm{Hauss}}(\gamma, \eta) < K$ and if their beginning points and ending points are pairwise K-near.

A K-quasigeodesic $\gamma : I \to X$ is called a *K-contracting axis* if

$$\mathrm{diam}(\pi_\gamma(\eta)) < K$$

holds for every geodesic η that does not enter the K-neighborhood of γ. An isometry g of X is called *K-contracting* if its orbit $\{g^i o\}_{i\in\mathbb{Z}}$ is a K-contracting axis. Two contracting isometries $g, h \in \mathrm{Isom}(X)$ are said to be *independent* if their orbits have bounded projections onto each other. A group $G \le \mathrm{Isom}(X)$ is said to be *non-elementary* if it contains two independent contracting isometries.

From now on, unless stated otherwise, we adopt:

{conv:main}

Convention 2.1. *(X, d) is a geodesic metric space with a basepoint o. For each $x, y \in X$, $[x, y]$ denotes an arbitrary geodesic connecting x to y. G is a non-elementary countable isometry group of X.*

This convention do not cover two important examples, namely, the Culler-Vogtmann Outer space and the Green metric on relatively hyperbolic groups. See Subsection 2.8 and 2.9 for their treatments.

For each isometry g of X, we define two norms

$$\|g\| := d(o, go), \quad \|g\|^{sym} := d(o, go) + d(go, o).$$

Note that these norms are subadditive. Clearly, $\|\cdot\|^{sym} = 2\|\cdot\|$ when d is a genuine metric. On Outer space, $\|\cdot\|^{sym}$ is only bi-Lipschitz to $\|\cdot\|$.

We say that a contracting isometry g of X is *WPD (weakly properly discontinuous)* if for each $K \ge 0$ there exists $N > 0$ such that

$$\mathrm{Stab}_K(o, g^N o) := \{h \in G : d(o, ho) \le K,\ d(g^N o, hg^N o) \le K\}.$$

is finite. For example, when the G-action on X is properly discontinuous, every contracting isometry of X in G becomes WPD.

{subsection:RW}

2.2. **Random walks.** Let $\{\mu_i\}_{i=1}^n$ be signed measures on G. By pushing forward the product measure $\mu_1 \times \cdots \times \mu_n$ on G by the composition map, we obtain the convolution measure $\mu_1 * \cdots * \mu_n$ on G. When $\mu_1 = \ldots = \mu_n = \mu$, we write by μ^n and μ^{*n} the n-product measure and the n-convolution measure of μ, respectively.

Let μ be a probability measure on G. The random walk on G generated by μ is a Markov chain with the transition probability $p(x, y) = \mu(x^{-1}y)$.

{notat:location}

Notation 2.2. *Given a sequence of isometries* $\mathbf{g} := (g_i)_{i \in \mathbb{Z}}$, *let*

$$Z_n := \begin{cases} g_1 \cdots g_n & n > 0 \\ id & n = 0 \\ g_0^{-1} \cdots g_{n+1}^{-1} & n < 0 \end{cases}$$

Given the i.i.d. input g_i, $Z_n(g_1, \ldots, g_n)$ models the n-th step position of the random walk starting from the identity. For instance, we can write the escape rate of μ as

$$l(\mu) := \lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\mu^n} \|Z_n\|.$$

For $p > 0$, we define the p-th moment of μ by

$$\mathbb{E}_\mu \|g\|^p = \sum_{g \in G} \mu(g)\, d(o, go)^p.$$

The *support* of μ is the set of elements $g \in G$ such that $\mu(g) > 0$.

We say that a probability measure μ on G is *non-elementary* if, for a suitable power n, $(\operatorname{supp} \mu^{*n})$ contains two independent contracting isometries.

Later, we will evaluate the same RV on $G^{\mathbb{Z}}$ with respect to different measures. For this purpose we introduce a notation.

{notat:prodExp}

Notation 2.3. *Let* $\{\mu_i\}_{i \in \mathbb{Z}}$ *be signed measures on* G, *all but finitely many being probability measures, and let* $F : G^{\mathbb{Z}} \to \mathbb{R}$ *be measurable function. When* g_i *are independently distributed according to* μ_i $(i \in \mathbb{Z})$, *we denote the expectation of* $F(\mathbf{g})$ *by*

$$\mathbb{E}_{\prod_{i \in \mathbb{Z}} \mu_i} F(\mathbf{g}),$$

and sometimes also by

$$\mathbb{E}_{\prod_{i=1}^\infty \mu_i \otimes \mu_0 \otimes \prod_{i=1}^\infty \mu_{-i}} F(\mathbf{g})$$

when we need to specify the indices associated with each measure.

{subsection:align}

2.3. **Alignment lemma.** In this subsection, we summarize the content of Subsection 3.2 and 3.3 of [Cho22a]. Refer to [Cho22a] for the proofs.

{dfn:alignment}

Definition 2.4. *For $i = 1, \ldots, n$, let γ_i be a path on X whose beginning and ending points are x_i and y_i, respectively. We say that $(\gamma_1, \ldots, \gamma_n)$ is* C-aligned *if*

$$\operatorname{diam}\big(y_i \cup \pi_{\gamma_i}(\gamma_{i+1})\big) < C, \quad \operatorname{diam}\big(x_{i+1} \cup \pi_{\gamma_{i+1}}(\gamma_i)\big) < C$$

hold for $i = 1, \ldots, n-1$.

{lem:1segment}

Lemma 2.5. *For each $C > 0$ and $K > 1$, there exists $D = D(C, K)$ that satisfies the following.*

Let γ, γ' be K-contracting axes whose beginning points are x and x', respectively. If (γ, x') and (x, γ') are C-aligned, then (γ, γ') is D-aligned.

{prop:BGIPWitness}

Proposition 2.6. *For each $D > 0$ and $K > 1$, there exist $E = E(D, K)$ and $L = L(D, K)$ that satisfy the following.*

Let $x, y \in X$ and let $\gamma_1, \ldots, \gamma_n$ be K-contracting axes whose domains are longer than L and such that $(x, \gamma_1, \ldots, \gamma_n, y)$ is D-aligned. Then the geodesic $[x, y]$ has subsegments $\eta_1, \ldots, \eta_n$, in order from left to right, that are longer than $100E$ and such that η_i and γ_i are $0.1E$-fellow traveling for each i.

We now define the notion of Schottky set, a collection of contracting axes (with uniform quality) that head into distinct directions. Given a sequence $s = (s_1, \ldots, s_n) \in G^n$, we employ the following notations:

$$\begin{aligned}\Pi(s) &:= s_1 s_2 \cdots s_n, \\ \Gamma(s) &:= \big(o, s_1 o, s_1 s_2 o, \ldots, \Pi(s) o\big).\end{aligned}$$

{dfn:Schottky}

Definition 2.7 (cf. [Gou22, Definition 3.11])**.** *Let $K_0 > 0$ and define:*

- *$D_0 = D(K_0, K_0)$ be as in Lemma 2.5,*
- *$E_0 = E(D_0, K_0)$, $L_0 = L(D_0, K_0)$ be as in Proposition 2.6.*

We say that $S \subseteq G^n$ is a long enough K_0-Schottky set *if:*

(1) $n > L_0$;
(2) $\Gamma(s)$ is K_0-contracting axis for all $s \in S$;
(3) for each $x \in X$ we have

$$\#\Big\{s \in S : \big(x, \Gamma(s)\big) \text{ and } \big(\Gamma(s), \Pi(s)x\big) \text{ are } K_0\text{-aligned}\Big\} \geq \#S - 1;$$

(4) for each $s, s' \in S$, $\big(\Gamma(s), \Pi(s)\Gamma(s')\big)$ is K_0-aligned.

Once a long enough Schottky set S is understood, its element s is called a Schottky sequence *and the translates of $\Gamma(s)$ are called* Schottky axes*. When μ is a probability measure on G such that $S \subseteq (\operatorname{supp} \mu)^n$, then we say that S is a long enough Schottky set for μ.*

{prop:Schottky}

Proposition 2.8. *Let μ be a non-elementary probability measure on G. Then for each $N > 0$ there exists a long enough Schottky set for μ with cardinality N.*

{dfn:semiAlign}

Definition 2.9. *Let S be a long enough Schottky set, let $D > 0$, let $x, y \in X$, let $(\gamma_1, \ldots, \gamma_N)$ be a sequence of Schottky axes, and let $(\gamma_{i(1)}, \ldots, \gamma_{i(n)})$ be its subsequence. If $(x, \gamma_1, \ldots, \gamma_N, y)$ is D-aligned, then we say that $(x, \gamma_{i(1)}, \ldots, \gamma_{i(n)}, y)$ is D-*semi-aligned.

{cor:semiAlign}

Corollary 2.10. *Let $n > 0$ and let $S \subseteq G^n$ be a long enough K_0-Schottky set, associated with constants D_0, E_0 as in Definition 2.7. If $x, y \in X$ and Schottky axes $\gamma_1, \ldots, \gamma_N$ are such that*

$$(x, \gamma_1, \gamma_2, \ldots, \gamma_N, y)$$

is D_0-semi-aligned, then the E_0-neighborhood of $[x, y]$ contains $\gamma_1, \ldots, \gamma_N$.

More precisely, $[x, y]$ has subsegments $\eta_1, \ldots, \eta_N$, in order from left to right, that are longer than $100E_0$ and such that η_i and γ_i are $0.1E_0$-fellow traveling for each i.

{subsection:Teich}

2.4. **Teichmüller space.** Let Σ be a finite-type hyperbolic surface. Its Teichmüller space $\mathcal{T}(\Sigma)$ carries two canonical metrics, namely, the Teichmüller metric $d_{\mathcal{T}}$ and the Weil-Petersson metric d_{WP}. The mapping class group $\mathrm{Mod}(\Sigma)$ of Σ acts on X isometrically with respect to both metrics.

A mapping class in $\mathrm{Mod}(\Sigma)$ either is cyclic, preserves a multicurve or exhibits Anosov dynamics on Σ away from finitely many points. The last ones are called pseudo-Anosov mapping classes, and they are contracting isometries with respect to both $d_{\mathcal{T}}$ and d_{WP} (see [Min96], [BF09]). Since the mapping class group action is properly discontinuous, all pseudo-Anosov mapping class is WPD. In this context, a probability measure μ on $\mathrm{Mod}(\Sigma)$ is non-elementary if the semigroup generated by $\operatorname{supp} \mu$ contains two independent pseudo-Anosov mapping classes.

A pseudo-Anosov mapping class has two fixed points on the Thurston boundary of $(\mathcal{T}(\Sigma, d_{\mathcal{T}})$, both uniquely ergodic. Masur's stability result [Mas80] tells us that these fixed points are connected by a unique Teichmüller geodesic, and moreover, that the following holds.

{fact:TeichSqueeze}

Fact 2.11. *Let $o \in \mathcal{T}(\Sigma)$, let φ be a pseudo-Anosov mapping class and let $\gamma : \mathbb{R} \to \mathcal{T}(\Sigma)$ be its invariant Teichmüller geodesic. Then for each $\epsilon > 0$ there exists $T > 0$ such that, if $x, y \in \mathcal{T}(\Sigma)$ satisfy*

$$\pi_\gamma(x) = \gamma(t_1), \ \pi_\gamma(y) = \gamma(t_2)$$

for some $t_2 > t_1 + 3T$, then $[x, y]$ contains a subsegment η that ϵ-fellow travels with $\gamma|_{[t_1+T, t_2-T]}$ (with respect to $d_{\mathcal{T}}$).

A similar conclusion for the Weil-Petersson geometry follows from the uniqueness of a recurrent Weil-Petersson geodesic connecting two ending laminations (Theorem 1.1, [BMM10]). In fact, due to the CAT(-1)-like behavior on the thick part of the Weil-Petersson Teichmüller geometry, we can say more:

{prop:WPSqueeze}

Proposition 2.12. *For each $K, \epsilon > 0$ there exists $L > 0$ such that the following holds.*

Let γ be a Weil-Petersson geodesic with ϵ-thick subsegments $\eta_1, \ldots, \eta_{2N+1}$, from left to right, whose lengths are at least L. Let γ' be another Weil-Petersson geodesic such that

$$d_{WP}(\gamma', \eta_1),\ d_{WP}(\gamma', \eta_{2N+1}) < K.$$

Then η_{N+1} is contained in the Ke^{-N}-neighborhood of γ'.

{subsection:CAT(0)}

2.5. **CAT(0) spaces.** CAT(0) spaces are geodesic spaces where geodesic triangles are not fatter than the Euclidean triangles with the same side lengths. A typical example is a complete Riemannian manifold with non-positive sectional curvature, such as, Euclidean spaces and n-dimensional hyperbolic spaces and their products.

If an isometry of a CAT(0) space preserves a bi-infinite geodesic that does not bound a flat half-plane, then we call it *rank-1*. Rank-1 isometries are essential for our theory due to:

{prop:bestvina2009Cont

Proposition 2.13 ([BF09, Theorem 5.3])**.** *Let X be a proper CAT(0) space. Then an isometry g of X is rank-1 if and only if it is contracting.*

We have plenty of contracting isometries in the following settings.

{prop:bestvina2009high

Fact 2.14 ([BF09, Theorem 6.5])**.** *Let X be a proper CAT(0) space. Suppose that the action of $\Gamma \le \mathrm{Isom}(X)$ satisfies WPD (see* [BF09, Definition 6.4]*). Then Γ is non-elementary.*

Fact 2.15 ([Ham09, Corollary 5.4])**.** *Let X be a proper CAT(0) space that admits a rank-1 isometry. Suppose that the limit set of $\mathrm{Isom}(X)$ on the visual boundary has at least 3 points and $\mathrm{Isom}(X)$ does not globally fix a point in ∂X. Then $\mathrm{Isom}(X)$ is non-elementary.*

Fact 2.16 ([CF10, Proposition 3.4])**.** *Let X be a proper CAT(0) space that admits a rank-1 isometry. Suppose that $\mathrm{Isom}(X)$ does not globally fix a point in ∂X nor stabilize a geodesic line. Then $\mathrm{Isom}(X)$ is non-elementary.*

{subsection:CAT(0)Cube

2.6. **CAT(0) cube complices.** CAT(0) cube complices are CAT(0) metric spaces made by gluing Euclidean cells. A cube complex is CAT(0) if and only if it is simply connected and each vertex exhibits a non-positively-curved behavior, namely, that its link is a flag complex. Originally proposed by Gromov, CAT(0) cube complices have become one of the central objects in geometric group theory. One instance of their importance lies in the recent proof of virtual Haken theorem, a major breakthrough in 3-manifold theory.

For basic terminologies about CAT(0) cube complices, we refer the readers to [CS11] and [CFI16]. A CAT(0) cube complex contains plenty of halfplanes, which cut the space into two (closed) halfspaces. Hence, each halfspace h is associated with a halfplane that we denote by $\hat{h}$, and also with its reflected copy $\overline{X \setminus h}$ that we denote by h^*.

{prop:capraceSageev}

Proposition 2.17 ([CS11, cf. Theorem A]). *Let X be an irreducible, finite-dimensional CAT(0) cube complex and $G \leq \mathrm{Aut}(X)$ be a group that does not globally fix a point nor stabilize a 1-dimensional flat in $X \cup \partial_\infty X$. Then X contains a convex G-invariant subcomplex Y on which the action of G is non-elementary.*

In particular, if G acts on X essentially, without fixing a point in $X \cup \partial X$, then G contains independent contracting isometries.

This result essentially follows from the Double Skewering Lemma of Caprace and Sageev in [CS11] that we describe now, with an additional information from [CFI16]. We say that two parallel hyperplanes of a CAT(0) cube complex are *strongly separated* if no hyperplane is transverse to both of them.

{fact:CAT(0)Bridge}

Fact 2.18 ([CS11, Proposition 3.2], [CFI16, Lemma 2.18, 2.24]). *Let G be a group that essentially acts on an irreducible, finite-dimensional CAT(0) cube complex X without a fixing a point. Then for any halfspace h of X associated with a hyperplane $\hat{h}$, there exists a rank-1 isometry $g \in G$ such that $\hat{h}$ and $g\hat{h}$ are strongly separated and such that $h \subseteq gh$. Moreover, there exists $p \in \hat{h}$, $q \in g\hat{h}$ such that*

$$d(x, y) = d(x, p) + d(p, q) + d(q, y) \tag{2.1}$$

holds for all $x \in h$ and $y \in (g\hat{h})^*$.

We now prove:

Lemma 2.19. *Let $D > 0$ be large enough, let $x, y \in X$ and suppose that the $10D$-neighborhood of $[x, y]$ contains $\{g^i o\}_{i=1}^{100D}$. Then we have*

$$d(x, y) = d(x, g^{50}p) + d(g^{50}p, y).$$

Proof. By translating by g if necessary, we may assume $o \in gh \setminus h$. Let g be K-contracting. Let $A = \{g^i o : i \in \mathbb{Z}\}$ and let $\pi_A(x) = g^m o$ and $\pi_A(y) = g^n o$; by swapping x and y if necessary, we may assume $m < n$. We claim that $[m, n] \supseteq [20D, 80D]$. If not and for example m is greater than $20D$, then we take a point $p \in [x, y]$ that is $10D$-close to go. Then both $[x, p]$ and $[p, q]$ has large projection on A that contains $\pi_A(x)$, so there exist points $p_1 \in [x, p]$, $p_2 \in [p, q]$ that are $3K$-close to p. This leads to a contradiction

$$20D \leq d(p_1, p) + d(p, p_2) = d(p_1, p_2) \leq 6K.$$

Similarly, $n < 80D$ leads to a contradiction and we establish the claim.

We now claim that $x \in g^{25D}h$. If not, then between x and $g^m o$ we need to cross at least 3 walls $g^{25D-1}\hat{h}, g^{25D-2}\hat{h}, g^{25D-3}\hat{h}$. Let p be the intersection point of $[x, g^m o]$ and $g^{25D-2}\hat{h}$. By [CS11, Lemma 6.1], the distance between p and $g^{25D-2}o$ is bounded by a constant that only depends on g and o, independent from D. Note also that

$$d(p, g^m o) \geq d(g^{25D-2}o, g^m o) - d(p, g^{25D-2}o)$$

is large. This leads to the contradiction that $g^{25D-2}o$ is much closer than $g^m o$. Hence, we obtain $x \in g^{25D}h \subseteq g^{50D}$, and similarly $y \in g^{75D}h^* \subseteq g^{50D+1}h^*$. We now apply Equation 2.1 to conclude. □

An important family of examples comes from right-angled Artin groups (RAAGs). Recall that a RAAG Γ is associated with a simply connected CAT(0) cube complex $\tilde{X}_\Gamma$; Γ acts properly and cocompactly on $\tilde{X}_\Gamma$, and the resulting quotient is called the *Salvetti complex* X_Γ of Γ. It is proved in [BC12] that if Γ is not a direct product, then the universal cover $\tilde{X}_\Gamma$ of the Salvetti complex admits a rank-1 isometry.

{section:CAT(-1)}

2.7. **CAT(-1) spaces.** CAT(-1) spaces are uniquely geodesic spaces where geodesic triangles are not fatter than the corresponding ones in the hyperbolic plane. Since the triangles in the hyperbolic plane are uniformly thin, so are the triangles in CAT(-1) spaces. In particular, CAT(-1) spaces are Gromov hyperbolic and all K-quasigeodesics are D-contracting for a constant $D = D(K)$ determined by K. Further, the following is true.

{fact:CAT(-1)Squeeze}

Fact 2.20. *There exists a constant $0 < c < 1$ such that the following holds. Let X be a CAT(-1) space and let $x, y, x', y' \in X$ be such that*

$$d(x, x'), d(y, y') < 0.02d(x, y).$$

Then $[x', y']$ is $c^{-d(x,y)}$-close to the midpoint of $[x, y]$.

{subsection:Outer}

2.8. **Culler-Vogtmann Outer space.** We now turn to the Culler-Vogtman Outer space equipped with the Lipschitz metric. We will not define most of the notions and only record the necessary facts; for precise definitions and proofs, we refer the readers to [Vog15], [FM11], [BF14] and [KMPT22a].

The Culler-Vogtman Outer space CV_N parametrizes projectivized marked minimal metric graphs with fundamental group F_N. The outer automorphism group of F_N, denoted by $\mathrm{Out}(F_N)$, naturally acts on CV_N by change of marking. Outer space has a natural simplicial structure that is preserved by the action of $\mathrm{Out}(F_N)$. The difference of two marked graphs are measured by the Lipschitz constant of the optimal map between them, which gives the Lipschitz metric on Outer space. This metric is not symmetric but is geodesic. It is necessary to define the neighborhoods of sets in terms of the symmetrization of d_{CV}, namely,

$$d^{sym}(x, y) := d_{CV}(x, y) + d_{CV}(y, x).$$

Hence, the R-neighborhood of a set $A \subseteq X$ is now defined by

$$\{x \in X : d^{sym}(x, A) \leq R\}.$$

From now on, when discussing Outer space with basepoint o, we only consider the standard geodesics between $\mathrm{Out}(F_N)$-orbit points of o that are proposed by Bestvina and Feighn in [BF14, Proposition 2.5]. For each $go, ho \in G \cdot o$ there exists a *standard geodesic* between go and ho, which is a concatenation $\gamma_1\gamma_2$ of two subsegments γ_1 and γ_2, where γ_1 is a rescaling segment on the simplex of x and γ_2 is a folding path. (For the definition of rescaling/folding path, refer to [BF14].) Note that the orbit points $G \cdot o$ are all ϵ-thick for some $\epsilon > 0$. This gives a uniform bound on the length of rescaling path: $l(\gamma_1) < \log(2/\epsilon)$ ([DT18, Lemma 2.6]).

We say that an outer automorphism $\phi \in \mathrm{Out}(F_N)$ is *reducible* if there exists a free product decomposition $F_N = C_1 * \cdots * C_k * C_{k+1}$, with $k \geq 1$ and $C_1, \ldots, C_{k+1} \neq \{e\}$, such that ϕ permutes the conjugacy classes of $C_1, \ldots, C_k$. If not, we say that ϕ is *irreducible.* We say that ϕ is *fully irreducible* if no power of ϕ is reducible. For us, an important point is that each fully irreducible automorphism $\phi \in \mathrm{Out}(F_N)$ has *bounded geodesic image property (BGIP)*, i.e., there exists $K > 0$

$$\mathrm{diam}(\pi_{\{\phi^i o : i \in \mathbb{Z}\}}(\eta)) < K$$

holds for all standard geodesics η that does not enter the K-neighborhood of $\{\phi^i o : i \in \mathbb{Z}\}$ ([KMPT22a, Theorem 7.8], cf. [Cho22b, Proposition 1.3]).

We now show a finer stability exhibited by certain outer automorphism. For each $N \geq 3$, $\mathrm{Out}(F_N)$ contains *principal* fully irreducible outer automorphism ([AKKP19, Example 6.1]). These automorphisms have lone axis on CV_N (cf. [MP16], [KMPT22b, Lemma 3.1]). Moreover, we have:

Fact 2.21 ([KMPT22a, Lemma 5.9])**.** *Let $\varphi \in \mathrm{Out}(F_N)$ be a fully irreducible outer automorphism with lone axis $\gamma : \mathbb{R} \to CV_N$ (parametrized by length), let $t \in \mathbb{R}$ and let $\epsilon \geq 0$. Then there exists $R > 0$ such that, if $x, y \in G \cdot o$ satisfies*

$$\pi_\gamma(x) \in \gamma((-\infty, t-R]), \quad \pi_\gamma(y) \in \gamma([t+R, +\infty)),$$

then $[x, y]$ passes through the ϵ-neighborhood of $\gamma(t)$.

{subsection:relHyp}

2.9. **Relatively hyperbolic groups.** Relatively hyperbolic groups are generalizations of free products of groups and geometrically finite Fuchsian and Kleinian groups. We call that a finitely generated group G is *word hyperbolic* if its Cayley graph is Gromov hyperbolic, or equivalently, if it acts geometrically on a Gromov hyperbolic space. More generally, let $\{P_\alpha\}_\alpha$ be a collection of subgroups of G. We say that G is *hyperbolic relative to the peripheral system* $\{P_\alpha\}_\alpha$ if it has a minimal, geometrically finite, convergence action on a compact space X in a way that the maximal parabolic subgroups are P_α's. Here, if G contains two hyperbolic elements with disjoint pair of fixed points, we say that G is a *non-elementary relatively hyperbolic group.*

Given a relatively hyperbolic group G, hyperbolic elements of G become contracting isometries on the Cayley graph of G with respect to any word metric. In particular, non-elementary relatively hyperbolic groups are subject to Theorem A, B and F.

There is another natural metric that can be put on G. Given a symmetric, finitely supported and admissible probability measure μ on G, we define the Green function associated to μ by

$$G^\mu(g) = \sum_{n=0}^{\infty} \mu^{*n}(g)$$

and the Green metric with respect to μ by

$$d_G^{\mu}(g,h) := \log G^{\mu}(id) - \log G^{\mu}(g^{-1}h).$$

This metric is not geodesic in general but is roughly geodesic. That means, there exists a uniform constant $C \geq 0$ such that for each $g, h \in G$ there is a C-coarse geodesic connecting g to h, which is a path $\gamma : [0, L] \to G$ such that $\gamma(0) = g$, $\gamma(L) = h$ and

$$|t_1 - t_2| - C \leq d_G^{\mu}\big(\gamma(t_1), \gamma(t_2)\big) \leq |t_1 - t_2| + C$$

holds for all $t_1, t_2 \in [0, L]$. For this metric, we define that a K-quasigeodesic $\gamma : I \to G$ is a K-contracting axis if $\mathrm{diam}(\pi_{\gamma}(\eta)) < K$ holds for all C-coarse geodesic η that does not enter the K-neighborhood of γ. Furthermore, $[x, y]$ denotes an arbitrary C-coarse geodesic between x and y. Then all the Alignment lemmata in Subsection 2.3 hold. Moreover, hyperbolic elements of G are indeed contracting isometries with respect to d_G^{μ}.

Let f be a hyperbolic element of G and let $\epsilon > 0$ be a small enough constant. Then the axis $\{f^n o\}_{n \in \mathbb{Z}}$ cannot stay long in any ϵ-neighborhood of a translate of a peripheral subgroup gP_{α}. That means, there exists $L > 0$ such that $\{f^n o, f^{n+1} o, \ldots, f^{n+L} o\}$ cannot be contained in $N_{\epsilon}(gP_{\alpha})$ for any $n \in \mathbb{Z}$, $g \in G$ and any peripheral P_{α}. Now we record the strong Ancona inequality along a geodesic that does not go deep into any translates of peripheral subgroups, due to Dussaule and Gekhtman.

{prop:greenDecay}

Proposition 2.22 ([DG21, Theorem 3.14]). *Let G be a non-elementary relatively hyperbolic group, let f be a hyperbolic element of G and let $K > 0$. Then there exists $L > 0$ such that, if $\gamma_1, \ldots, \gamma_N$ are paths on G of the form*

$$\gamma_i = (go, gfo, \ldots, gf^L o) \quad (g \in G)$$

and if $x, y, x', y' \in G$ satisfy that $(x, \gamma_1, \ldots, \gamma_N, y)$, $(x', \gamma_1, \ldots, \gamma_N, y')$ are K-aligned, then we have

$$|d_G^{\mu}(x,y) + d_G^{\mu}(x',y') - d_G^{\mu}(x,y') - d_G^{\mu}(x',y)| \leq e^{-N}. \tag{2.2}$$
{eqn:greenDecay}

In some cases, we have an estimate analogous to Inequality 2.2 on the Cayley graph of G with the word metric d_G. Two notable examples are surface groups and free products. First consider the Cayley graph of $G = \langle a_1, \ldots, a_g, b_1, \ldots, b_g | \prod_{i=1}^{g} [a_i, b_i] \rangle$. Note that:

{lem:corridor}

Lemma 2.23 (cf. [Sam02, Lemma 2.3]). *Let $w = a_i^k b_{i+1}^k$, let $n > 2$ and let $x, y, z \in G$ be such that:*

1. *there exists a geodesic γ from x to y containing w^n as a subword, and*
2. *$\pi_{\gamma}(z)$ is left to the w^n-subword of γ by more than $4g$.*

Then every geodesic on G from z to y must contain the middle subsegment w^{n-2} of w^n.

In [Sam02], the author describes the situation $z = x$. One can reduce the setting of Lemma 2.23 to that of [Sam02, Lemma 2.3] by considering one pod of the geodesic triangle $\triangle xyz$, whose one side contains w^n as subword.

Using this lemma, we can deduce the following:

{cor:corridor}

Corollary 2.24. *Let $w = a_i^k b_{i+1}^k$ and let $x, y \in G$ be such that*

$$\pi_{\{w^i : i \in \mathbb{Z}\}}(x) = w^n,\ \pi_{\{w^i : i \in \mathbb{Z}\}}(y) = w^m \quad (n < m - 5).$$

Then any geodesic on G from x to y must pass through the vertex w^{n+2}.

In particular, if we consider another pair of points $x', y' \in G$ such that

$$\pi_{\{w^i : i \in \mathbb{Z}\}}(x') = w^{n'},\ \pi_{\{w^i : i \in \mathbb{Z}\}}(y') = w^{m'} \quad (n' \le n, m' \ge n + 4),$$

Then we have

$$d_G(x, y) - d_G(x', y) = d_G(x, w^{n+2}) - d_G(x', w^{n+2}) = d_G(x, y') - d_G(x', y').$$

A similar bottlenecking phenomenon happens in the free product of nontrivial groups $G = *_\alpha G_\alpha$. Indeed, if we take two distinct free factors G_α, G_β and pick nontrivial elements $a \in G_\alpha$ and $b \in G_\beta$, then $w = ab$ does the same job with $w = a_i^k b_i^k$ as in Lemma 2.23.

{section:diff}

3. Differentiability of the escape rate

In this section, we establish Lipschitz continuity and differentiability of the escape rate.

{subsection:deficit}

3.1. Defect.

We first define the *defect* arising from the addition of displacements along a random path. Let:

$$\begin{aligned} R_{n,m}(\mathbf{g}) &:= \|Z_{-(m+1)}^{-1} Z_{-1}\| + \|Z_{-1}\| + \|Z_n\| - \|Z_{-(m+1)}^{-1} Z_n\| \\ &= \|g_{-m} \cdots g_{-1}\| + \|g_0\| + \|g_1 \cdots g_n\| - \|g_{-m} \cdots g_n\|. \end{aligned}$$

{lem:conditionedDist}

Lemma 3.1. *Let μ be a non-elementary probability measure on G. Then there exists $K, \epsilon > 0$ such that*

$$\mathbb{P}_{\mu_1 \times \cdots \times \mu_k}\left((g_1, \ldots, g_k) : \begin{array}{c} \text{there exists } 1 < l < k \text{ such that} \\ d\left(Z_l o, [o, Z_k o]\right) < K \end{array} \middle| \, g_1, g_j, g_k \right) \ge 1 - K e^{-k/K}$$

for all $1 < j < k$, for all choice of g_1, g_j and g_k, and for all sequence of probability measures $\{\mu_i\}_i$ such that $\|\mu_i - \mu\|_0 < \epsilon$ for each i.

Proof. Recall that G contains two independent contracting isometries of X. In this setting, we have defined when a set of sequences $S \subseteq G^n$ is called a long enough Schottky set ([Cho22a, Definition 3.15]). A long enough Schottky set S comes with constants $D_0, E_0 > 0$. Given an element $s = (s_1, \ldots, s_n)$ of S, we define its axis by the path

$$\Gamma(s) := (o,\, s_1 o,\, s_1 s_2 o,\, \ldots,\, s_1 \cdots s_n o)$$

and employ the notation $\Pi(s) := s_1 \cdots s_n$. Translates of $\Gamma(s)$ for $s \in S$ are called the Schottky axes.

Once a long enough Schottky set S is understood, we can define the D_0-*semi-alignment* of

$$(x, \gamma_1, \gamma_2, \ldots, \gamma_N, y),$$

for $x, y \in X$ and Schottky axes γ_i's. For us, what only matters is the conclusion of Corollary 2.10, namely, that $\gamma_1, \ldots, \gamma_N$ are contained in the E_0-neighborhood of $[x, y]$.

Let us fix a long enough Schottky set S (together with constants D_0, E_0 as in Definition 2.7). Given a sequence $(w_i)_{i=0}^{\infty}$ in G, we draw $a_i, b_i, c_i, d_i, \ldots$ independently from S with the uniform distribution, and construct a word

$$W_n = w_0 \Pi(a_1)\Pi(b_1)\Pi(c_1)\Pi(d_1) w_1 \cdots \Pi(a_n)\Pi(b_n)\Pi(c_n)\Pi(d_n) w_n.$$

In [Cho22a, Subsection 5.1], we defined the *set of pivotal times* P_n that is determined by the values of $\{a_i, b_i, c_i, d_i, v_i\}_{i=1}^{n}$ and $\{w_i\}_{i=0}^{n}$. We have proven:

{fact:pivotAlign}

Fact 3.2. *If a choice $(a_i, b_i, c_i, d_i, w_{i-1})_{i>0} \in (S^{\mathbb{N}})^4 \times G^{\mathbb{N}}$ has the set of pivotal times $P_n = \{j(1) < \ldots < j(m)\}$, then the sequence*

$$(o, W_{j(1)-1}\Gamma(a_{j(1)}), \ldots, W_{j(m)-1}\Gamma(a_{j(m)}), W_n o)$$

is D_0-semi-aligned. In particular, $W_{j(1)-1}o, \ldots, W_{j(m)-1}o$ are contained in the E_0-neighborhood of $[o, W_n o]$.

We also have:

{fact:pivotIncrease}

Fact 3.3. *Given a long enough Schottky set S of cardinality at least 100, there exists a constant $K' > 0$ such that the following holds.*

Let $(w_i)_{i=0}^{\infty}$ be a sequence in G. Then

$$\mathbb{P}\left(\#P_n(\{a_i, b_i, c_i, d_i, w_{i-1}\}_{i>0}) \le n/K'\right) \le K e^{-n/K'}.$$

for each n, where a_i, b_i, c_i, d_i are chosen independently from S with the uniform distribution.

We are now ready to prove the main statement. Since μ is non-elementary, there exists $M_0 > 0$ such that $(\operatorname{supp} \mu^{M_0})$ contains a long enough Schottky set S with cardinality 400 (Proposition 2.8). This means that

$$\epsilon := 0.5 \cdot \min\{\mu(a_1), \ldots, \mu(a_{M_0}) : (a_1, \ldots, a_{M_0}) \in S\}$$

is positive. If we take any probability measures μ_i's with $\|\mu - \mu_i\| < \epsilon$, then $\mu_i(a_1)$, $\ldots$, $\mu_i(a_{M_0})$ are greater than ϵ for each $(a_1, \ldots, a_{M_0}) \in S$. Consequently, for each i, the product measure $\mu_{i+1} \times \cdots \times \mu_{i+M_0}$ dominates $\beta \cdot (\text{uniform measure on } S)^4$, where $\beta = \epsilon^{4M_0}$.

Let D_0, E_0 be the constants for S as in Corollary 2.10. Let K be a constant larger than E_0 and such that the conclusion of Fact 3.3 holds. Recall that our ambient probability space is $(G^k, \prod_{i=1}^{k} \mu_i)$, where $\|\mu - \mu_i\|_0 < \epsilon$ for each i. Hence, for each l, we have the decomposition

$$\mu_{4M_0 l+1} \times \cdots \times \mu_{4M_0(l+1)} = \beta \cdot (\text{uniform measure on } S)^4 + (1-\beta)\nu_l$$

for some probability measure ν_l. Let

$$I_1 = [4M_0k_1 + 1, 4M_0(k_1+1)],\ I_2 = [4M_0k_2+1, 4M_0(k_2+1)],\ \ldots$$

be disjoint intervals that does not contain j: we can ensure that $[2, k-1]$ contains at least $k/5M_0$ such intervals. We now independently pick each interval I_i with probability β to make a collection $\mathcal{A}$:

$$\mathbb{P}(I_i \in \mathcal{A}) = \beta \text{ for each } i.$$

Then for probability at least $1 - e^{-\beta k/8}$ we have $\#\mathcal{A} \geq 0.1\beta k/M_0$. Once $\mathcal{A}$ is determined, we decide the values of $g_i : i \notin I_1 \cup I_2 \cup \cdots$ using the distribution of μ_i and the values of $(g_{4M_0k_i+1}, \ldots, g_{4M_0(k_i+1)}) : I_i \notin \mathcal{A}$ using the distribution of ν_{k_l}. Note that g_0 and g_j are pre-determined. Conditioned on these choices, the remaining choices $(g_{4M_0k_i+1}, \ldots, g_{4M_0(k_i+1)}) : I_i \in \mathcal{A}$ are independent and distributed according to (uniform measure on $S)^4$. As a result, we obtain measurable partition $\{\mathcal{F}_\alpha\}_\alpha$ of the ambient space:

$$\{(g_1, \ldots, g_k) \in G^k : g_1, g_j, g_k \text{ fixed}\} = \sqcup_\alpha \mathcal{F}_\alpha$$

such that each $\mathcal{F}_\alpha$ is associated with a set $\mathcal{A}$ of intervals, fixed isometries $\{w_i\}_i$, and sequences $(a_i, b_i, c_i, d_i)_{i=1}^{\#\mathcal{A}}$ chosen independently from S. On each $\mathcal{F}_\alpha$ we have

$$g_1 \cdots g_k = W_{\#\mathcal{A}} = w_0\Pi(a_1)\Pi(b_1)\cdots\Pi(c_{\#\mathcal{A}})\Pi(d_{\#\mathcal{A}}) \cdot w_{\#\mathcal{A}},$$

and

$$W_1, W_2, \ldots, W_{\#\mathcal{A}-1} \in \{g_0, g_0g_1, \ldots, g_0\cdots g_{k-1}\}.$$

Moreover, we have observed

$$\mathbb{P}_{\mu_1\times\cdots\times\mu_k}(\#\mathcal{A} \geq 0.1\beta k/M_0) \geq 1 - e^{-\beta k/8}. \tag{3.1}$$

Now Fact 3.2 and Fact 3.3 tell us that

$$\begin{aligned} &\mathbb{P}\left(d(g_1\cdots g_l o, [o, g_1\cdots g_k o]) < K \text{ for some } l < k \,\middle|\, \mathcal{F}_\alpha\right) \\ &\leq \mathbb{P}\left(d(W_l o, [o, W_{\#\mathcal{A}} o]) < K \text{ for some } l \leq \#\mathcal{A} \,\middle|\, \mathcal{F}_\alpha\right) \\ &\leq \mathbb{P}\left(\#P_{\#\mathcal{A}}(\mathbf{g}) \leq \#\mathcal{A}/K \,\middle|\, \mathcal{F}_\alpha\right) \leq K' e^{-\#\mathcal{A}/K'}. \end{aligned} \tag{3.2}$$

Combining Inequality 3.1 and 3.2 yields the conclusion. □

{cor:bddDefect}

Corollary 3.4. *Let μ be a non-elementary probability measure on G with finite first moment. Then there exists $K_1, \epsilon > 0$ such that*

$$\mathbb{E}_{\prod_{i=1}^\infty \mu_i \otimes 1_{g_0=g} \otimes \prod_{i=1}^\infty \mu_{-i}}[R_{n,m}(\mathbf{g})] \leq K_1$$

for every $n, m > 0$, for every $g \in G$, and for every probability measures $\{\mu_i\}_{i\neq 0}$ such that $\|\mu_i - \mu\|_{0,1} < \epsilon$ for each i.

Proof. We define auxiliary RVs

$$\begin{aligned} R'(\mathbf{g}) &:= \|g_{-m}\cdots g_{-1}g_0\| + \|g_1\cdots g_n\| - \|g_{-m}\cdots g_{-1}\cdot g_0 \cdot g_1\cdots g_n\|, \\ R''(\mathbf{g}) &:= \|g_{-m}\cdots g_{-1}\| + \|g_0\| - \|g_{-m}\cdots g_{-1}g_0\|. \end{aligned}$$

Clearly, we have $R(\mathbf{g}) = R'(\mathbf{g}) + R''(\mathbf{g})$.

Let $K, \epsilon > 0$ be as in Lemma 3.1 and set

$$\begin{aligned} E_j &:= \big\{\mathbf{g} : \text{there exists } i \in \{1, \ldots, j\} \text{ such that } d\big(Z_i o,\, [Z_{-m} o, Z_n o]\big) < K\big\} \\ &= \left\{\mathbf{g} : \begin{array}{c} \text{there exists } i \in \{1, \ldots, j\} \text{ such that} \\ d\big(ho, [o, hg_1 \cdots g_j h' o]\big) < K \end{array} \right\} \quad (h = g_{-m} \cdots g_0,\, h' = g_{j+1} \cdots g_n). \end{aligned}$$

By the previous lemma, we have

$$\mathbb{P}_{\prod_i \mu_i} \left(E_{j-1}^c \,\middle|\, g_{-m}, \ldots, g_0, g_j, \ldots, g_n \right) < K e^{-(j-1)/K}$$

for each j and each prescribed choices of $g_{-m}, \ldots, g_0, g_j, \ldots, g_n$. Moreover, if $Z_l o$ is K-close to $[Z_{-m} o, Z_n o] < K$ for some $1 \le l < j$, then

$$\begin{aligned} R'(\mathbf{g}) &= \|g_{-m} \cdots g_{-1} g_0\| + \|g_1 \cdots g_m\| - \|g_{-m} \cdots g_{-1} \cdot g \cdot g_1 \cdots g_n\| \\ &\le \|g_{-n} \cdots g_{-1} g_0\| + \|g_1 \cdots g_m\| - (\|g_{-m} \cdots g_{-1} g_0 g_1 \cdots g_l\| + \|g_l \cdots g_n\| - K) \\ &\le \|g_1 \cdots g_l\|^{sym} + K \le \sum_{i=1}^{j} \|g_i\|^{sym} + K. \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathbb{E}[R'] &\le \mathbb{E}\left[\sum_{i=1}^{\min\{j : \mathbf{g} \in E_j\}} \|g_i\|^{sym} + K \right] \le \sum_{i=1}^{m} \mathbb{E}\left[\|g_i\|^{sym} 1_{E_{i-1}^c} \right] + K \\ &\le K + \sum_{i=1}^{m} \mathbb{E}\left[\|g_i\|^{sym}\, \mathbb{P}(E_{i-1}^c | g_i) \right] \\ &\le K + \sum_{i=0}^{m-1} K e^{-k/K} \|\mu_i\|_1 \le K + \frac{K}{1 - e^{-1/K}} (\|\mu\|_1 + \epsilon). \end{aligned}$$

Similarly, if we define

$$F_j := \left\{\mathbf{g} : \exists k \in \{1, \ldots, j\} \text{ s.t. } d\left(Z_{-(i+1)} o, [Z_{-m} o, o]\right) < K\right\},$$

then $\mathbb{P}(F_{j-1}^c | g_{-m}, \ldots, g_{-j}, g_0) < K e^{-(j-1)/K}$ holds for each j and $g_{-m}, \ldots, g_{-j}, g_0$. Moreover, we similarly have

$$R'' \le K + \sum_{i=1}^{\min\{j : \mathbf{g} \in F_j\}} \|g_{-i}\|^{sym}.$$

Based on the same computation, we conclude

$$\mathbb{E}[R_{n,m}] = \mathbb{E}[R'] + \mathbb{E}[R''] \le 2K + \frac{2K}{1 - e^{-1/K}} (\|\mu_1\| + \epsilon). \qquad \square$$

We now prove Theorem A.

Proof. Let μ be a non-elementary probability measure with finite first moment, and let $K_1, \epsilon > 0$ be the constants for μ as in Corollary 3.4. Now

consider a measure μ' with $\|\mu' - \mu\| < \epsilon$. Then we have
(3.3) {eqn:driftLipschitz}

$$\begin{aligned}
L(\mu'^{*n}) - L(\mu^{*n}) &= \sum_{i=1}^{n} \left(\mathbb{E}_{\mu^{*(i-1)} \times \mu' \times \mu'^{*(n-i)}} \|g_1 g_2 g_3\| - \mathbb{E}_{\mu^{*(i-1)} \times \mu \times \mu'^{*(n-i)}} \|g_1 g_2 g_3\| \right) \\
&= \sum_{i=1}^{n} \left(\mathbb{E}_{\mu'} \|g\| + \mathbb{E}_{\mu^{*(i-1)}} \|g\| + \mathbb{E}_{\mu'^{*(n-i)}} \|g\| - \mathbb{E}_{\mu'^{i-1} \otimes \mu' \otimes \mu^{n-i}} [R_{n-i,i-1}(\mathbf{g})] \right) \\
&\quad - \sum_{i=1}^{n} \left(\mathbb{E}_{\mu} \|g\| + \mathbb{E}_{\mu^{*(i-1)}} \|g\| + \mathbb{E}_{\mu'^{*(n-i)}} \|g\| - \mathbb{E}_{\mu'^{n-1} \otimes \mu \otimes \mu^{i-1}} [R_{n-i,i-1}(\mathbf{g})] \right) \\
&= n \, \mathbb{E}_{\mu' - \mu} \|g\| + \sum_{i=1}^{n} \sum_{g \in G} \big(\mu'(g) - \mu(g)\big) \, \mathbb{E}_{\mu'^{n-i} \otimes 1_{g_0 = g} \otimes \mu^{i-1}} [R_{n-i,i-1}(\mathbf{g})].
\end{aligned}$$

Here, $\mathbb{E}_{\mu'-\mu} \|g\|$ is bounded by $\|\mu' - \mu\|_1$. Moreover, Corollary 3.4 tells us that $\mathbb{E}_{\mu'^k \otimes 1_{g_0 = g} \otimes \mu^l} [R_{l,k}(\mathbf{g})] < K_1$ for any $k, l \geq 0$ and $g \in G$. Hence, the second term is bounded by $nK_1 \|\mu_1 - \mu\|_0$. As a result, we have

$$\left| \frac{L(\mu'^{*n}) - L(\mu^{*n})}{n} \right| \leq \|\mu' - \mu\|_1 + K_1 \|\mu' - \mu\|_0.$$

By sending n to infinity, we obtain the same bound for $|l(\mu') - l(\mu)|$. $\square$

We record some lemmata before proceeding to Theorem B. Their proofs follow from simple computations so we omit them.

{lem:driftDiffMean0}

Lemma 3.5. *Let μ, η be signed measures and $\{\mu_t\}_{t \in [-1,1]}$ be a family of signed measures such that*

$$\|\mu\|_0 = \|\mu_t\|_0 < +\infty \text{ for all } t, \quad \|\eta\|_0 < +\infty,$$

and such that $\|\mu_t - \mu - t\eta\|_0 = o(t)$. Then η is balanced, i.e., $\sum_{g \in G} \eta(g) = 0$.

{lem:driftConv}

Lemma 3.6. *Let μ, μ', η, η' be signed measures with finite $\|\cdot\|_{0,1}$-norm, and let $\{\mu_t\}_{t \in [-1,1]}$, $\{\mu'_t\}_{t \in [-1,1]}$ be families of signed measures such that*

$$\|\mu_t - \mu - t\eta\|_{0,1}, \|\mu'_t - \mu' - t\eta'\|_{0,1} = o(t).$$

Then we have

$$\big\| \mu_t * \mu'_t - \mu * \mu' - t(\eta * \mu' + \mu * \eta') \big\|_{0,1} = o(t)$$

Proof of Theorem B. Again, let $K_1, \epsilon > 0$ be the constants for μ as described in Corollary 3.4.

In the previous proof, we basically proved

$$\left| \frac{1}{n} [L(\mu_t^{*n}) - L(\mu^{*n})] - \mathbb{E}_{\mu_t - \mu} \|g\| \right| \leq K(\mu_t, \mu) \|\mu_t - \mu\|_0,$$

for small enough μ_t (so that $\|\mu_t - \mu\|_0 < \epsilon$), where

$$K(\mu_t, \mu) := \sup_{g \in G} \sup_{n,m > 0} \mathbb{E}_{\mu_t^n \otimes 1_{g_0 = g} \otimes \mu^m} [R_{n,m}(\mathbf{g})] \leq K_1.$$

Let us now replace μ_t with μ_t^{*k} and μ with μ^{*k}. We have

$$\left|\frac{1}{n}[L(\mu_t^{*nk}) - L(\mu^{*nk})] - \mathbb{E}_{\mu_t^{*k}-\mu^{*k}} \|g\|\right| \leq K(\mu_t^{*k}, \mu^{*k})\|\mu_t^{*k} - \mu^{*k}\|_0.$$

Here, note that $K(\mu_t^{*k}, \mu^{*k}) \leq K(\mu_t, \mu) \leq K$. Hence, after dividing by k and t we have

$$\frac{1}{t}\left|\frac{1}{nk}[L(\mu_t^{*nk}) - L(\mu^{*nk})] - \frac{1}{k}\,\mathbb{E}_{\mu_t^{*k}-\mu^{*k}} \|g\|\right| \leq \frac{1}{t}\frac{K_1}{k}\|\mu_t^{*k} - \mu^{*k}\|_0.$$

for arbitrary $k > 0$. By sending n to infinity, we have

$$\left|\frac{1}{t}[l(\mu_t) - l(\mu)] - \frac{1}{t}\frac{1}{k}\,\mathbb{E}_{\mu_t^{*k}-\mu^{*k}} \|g\|\right| \leq \frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu_0^{*k}\|_0}{k}.$$

By Lemma 3.6, we have

$$\lim_{t\to 0} \frac{1}{t}\frac{1}{k}\,\mathbb{E}_{\mu_t^{*k}-\mu^{*k}} \|g\| = \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\|$$

and

(3.4) {eqn:driftDiff1}

$$\left|\limsup_{t\to 0} \frac{1}{t}[l(\mu_t) - l(\mu)] - \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\|\right| \leq \limsup_{t\to 0}\left|\frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu_0^{*k}\|_0}{k}\right|.$$

The inequality still holds if we replace $\limsup_t \frac{l(\mu_t)-l(\mu)}{t}$ with $\liminf_t \frac{l(\mu_t)-l(\mu)}{t}$. Hence, to conclude that $\lim_{t\to 0} \frac{l(\mu_t)-l(\mu)}{t} = \lim_k \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\|$, it suffices to prove that

$$\lim_{k\to\infty} \limsup_{t\to 0}\left|\frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu_0^{*k}\|_0}{k}\right| = 0.$$

Again, Lemma 3.6 implies

(3.5) {eqn:driftDiff2}

$$\begin{aligned}\limsup_{t\to 0}\left|\frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu^{*k}\|_0}{k}\right| &= \limsup_{t\to 0} \frac{K_1}{kt} \cdot \left(t\left\|\sum_{i=1}^{k} \mu^{*(i-1)} * \eta * \mu^{(k-i)}\right\|_0 + o(t)\right) \\ &= \frac{K_1}{k} \cdot \left\|\sum_{i=1}^{k} \mu^{*(i-1)} * \eta * \mu^{(k-i)}\right\|_0.\end{aligned}$$

Let $f(g) := \eta(g)/\mu(g)$ for $g \in \operatorname{supp}\mu$. Then f is a μ-integrable function (since $\sum_g |f(g)|\mu(g) = \sum_g |\eta(g)| < \infty$) with mean 0 (since $\sum_g f(g)\mu(g) = \sum_g \eta(g) = 0$). Hence, we have

$$\begin{aligned}\frac{1}{k}\left\|\sum_{i=1}^{k} \mu^{*(i-1)} * \eta * \mu^{(k-i)}\right\|_0 &= \frac{1}{k}\sum_{g_i\in G} |f(g_1) + \cdots + f(g_k)|\,\mu(g_1)\cdots\mu(g_k) \\ &\leq \mathbb{E}_{\mu^k}\left|\frac{f(g_1) + \cdots + f(g_k)}{k}\right|.\end{aligned}$$

The final term tends to 0 as k tends to infinity, by the recurrence of integrable balanced random walk on $\mathbb{R}$ and the subadditive ergodic theorem. This concludes the differentiability of $l(\mu_t)$.

Let us now combine Inequality 3.4 and 3.5: under the assumptions on $\mu, \eta, \{\mu_t\}$, we have

$$\left| \frac{d}{dt} l(\mu_t)\Big|_{t=0} - \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\| \right| \le \frac{K}{k} \cdot \left\| \sum_{i=1}^{k} \mu^{*(i-1)} * \eta * \mu^{(k-i)} \right\|_0 . \tag{3.6}$$

Here, the constant K depends on μ, and can be kept the same even if we replace μ with μ' such that $\|\mu' - \mu\| < \epsilon/2$.

Let us now show the the continuity of the derivative. Fix $\delta > 0$. Thanks to the previous argument, there exists $k > 0$ such that

$$\frac{1}{k}\left\| \sum_{i=1}^{k} \mu^{*(i-1)} * \eta * \mu^{(k-i)} \right\|_0 < \frac{\delta}{3K_1}. \tag{3.7}$$

Note that this finite combination of convolutions of μ and η is continuous with respect to μ and η (under the $\|\cdot\|_0$-topology). For the same reason,

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\|$$

is continuous with respect to μ and η (under the $\|\cdot\|_{0,1}$-topology). Hence, we can take $0 < \epsilon_1 < \epsilon/2$ such that, for all probability measure μ' and for all signed measure η' such that $\|\mu' - \mu\|_{0,1}, \|\eta' - \eta\|_{0,1} < \epsilon_1$, we have:

$$\begin{aligned} \frac{1}{k}\left\| \sum_{i=1}^{k} \mu'^{*(i-1)} * \eta' * \mu'^{(k-i)} \right\|_0 &< \frac{\delta}{2K_1}, \\ \left| \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu'^{*(i-1)}*\eta'*\mu'^{*(k-i)}} \|g\| - \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{\mu^{*(i-1)}*\eta*\mu^{*(k-i)}} \|g\| \right| &< \frac{\delta}{2}. \end{aligned} \tag{3.8}$$

Now let $\{\mu'_t\}_t$ be another family of probability measures such that $\|\mu'_t - \mu'\|_{0,1} \to 0$, $\frac{1}{t}\|\mu'_t - \mu' - t\eta'\|_{0,1} \to 0$. By plugging Inequality 3.7 and 3.8 into Inequality 3.6, we conclude that $\left|\frac{d}{dt} l(\mu_t) - \frac{d}{dt} l(\mu'_t)\right| < 2\delta$ as desired. □

4. Squeezing isometries and escape rate

{section:squeezing}

As we did in Lemma 3.1, we describe another simple geometric situation. Given that this situation happens with high probability, Theorem C can be deduced follow from elementary arguments.

We first collect directions along which geodesics are brought close to each other as much as we want.

{dfn:squeezing}

Definition 4.1. *Let (X, d) be a geodesic metric space and let $\epsilon \ge 0$. A geodesic $\gamma : [0, L] \to X$ is said to be* ϵ-squeezing *if there exists $p \in \gamma$ such that $d(p, \gamma') < \epsilon$ for all geodesic γ' whose $0.01L$-neighborhood contains γ.*

On CAT(-1) spaces, any geodesic longer than $e^{100/\epsilon}$ is ϵ-squeezing. For Teichmüller space, given any pseudo-Anosov φ, a sufficiently long subsegment of the φ-invariant geodesic (with respect to either Teichmüller metric or the Weil-Petersson metric) is ϵ-squeezing. The standard Cayley graph of a surface group or a free product even contains 0-squeezing geodesics.

For CAT(0) cube complices and Green metrics on relatively hyperbolic groups, we use the following version:

{dfn:squeezingNon}

Definition 4.2. *Let (X, d) be a metric space and let $\epsilon \geq 0$. A pair of points $(p, q) \in X^2$ is said to be* ϵ-squeezing *if, for each $x, y, x', y' \in X$ such that*

$$(x, q)_p, (x', q)_p, (y, p)_q, (y', p)_q < 0.01 d(p, q),$$

then

$$\big|d(x, y) + d(x', y') - d(x, y') - d(x', y)\big| < \epsilon.$$

We say that an isometry g of X is ϵ-squeezing *if (o, go) is ϵ-squeezing.*

On CAT(0) cube complex, the bridge between two strongly separated halfspaces is always 0-squeezing. In relatively hyperbolic groups, a geodesic $\gamma : I \to X$ is ϵ-squeezing if it contains $|\log \epsilon|$-many (ϵ, η)-transition points $\gamma(t_i)$'s, i.e., where $\gamma([t_i - \eta, t_i + \eta])$ does not lie in the ϵ-neighborhood of a single parabolic subgroup.

In the remaining of this section, we assume that:

{conv:squeeze}

Convention 4.3. *(X, G) is either:*

- *Teichmüller space (with the Teichmüller metric or the Weil-Petersson metric) and the mapping class group;*
- *Culler-Vogtman Outer space CV_N and* $\mathrm{Out}(F_N)$*;*
- *a CAT(-1) space and its countable isometry group;*
- *a CAT(0) cube complex and its countable isometry group;*
- *relatively hyperbolic group equipped with a Green metric;*
- *the standard Cayley graph of a surface group, or;*
- *the standard Cayley graph of a free product of nontrivial groups.*

Let $A_{k;\epsilon}$ be the collection of step paths $\mathbf{g} \in G^{\mathbb{Z}}$ for which there exist ϵ-squeezing geodesics γ, γ' (whose length is denoted by L) satisfying the following for each $m, n \geq k$:

(1) γ and γ' are contained in the $0.01L$-neighborhood of $[Z_{-m}o, Z_n o]$;
(2) γ is contained in the $0.01L$-neighborhood of $[o, Z_n o]$, and;
(3) γ' is contained in the $0.01L$-neighborhood of $[Z_{-m}o, Z_{-1}o]$.

Our main probabilistic estimate is as follows. We will defer the proof of Proposition 4.4 to Section 6.

{prop:squeeze}

Proposition 4.4. *Let (X, G) be as in Convention 4.3 and let μ be a non-elementary probability measure on G. Then there exists $\epsilon > 0$, and for each $\epsilon' > 0$ there exists $K > 0$, such that for every sequence of probability* *measures* $\{\mu_i\}_i$ *with* $\|\mu_i - \mu\|_{0,1} < \epsilon$*, for every* $j \in \mathbb{Z}\setminus\{0\}$ *and* $k \in \mathbb{Z}_{>0}$*, and for every prescribed choices* $g_0, g_j \in G$*, we have*

$$\mathbb{P}_{\prod_i \mu_i}\left(\mathbf{g} \in A_{k;\epsilon'} \,\middle|\, g_0, g_j\right) \geq 1 - Ke^{-k/K}.$$

Our notion of $A_{k,\epsilon'}$ is justified the following lemma.

{lem:squeezeCompare}

Lemma 4.5. *Let* $\mathbf{g} \in A_{k;\epsilon'}$*. Then for each* $n, m, n', m' \geq k$*, we have*

$$|R_{n,m}(\mathbf{g}) - R_{n',m'}(\mathbf{g})| < 10\epsilon'.$$

Proof. We will prove that $|R_{n,m} - R_{n',m}| < 5\epsilon'$. Together with its symmetric counterpart, this leads to the conclusion.

Since $\mathbf{g} \in A_{k;\epsilon'}$, there exists an ϵ-squeezing geodesic γ, with length L, that is contained in the $0.01L$-neighborhoods of $[Z_{-(m+1)}o, Z_n o]$, $[Z_{-(m+1)}o, Z_{n'}o]$, $[o, Z_n o]$ and $[o, Z_{n'}o]$. Since γ is ϵ-squeezing, there exists $p \in \gamma$ that is ϵ'-close to these 4 geodesics. Then we have

$$\begin{aligned} |d(Z_{-(m+1)}o, Z_n o) - d(Z_{-(m+1)}o, p) - d(p, Z_n o)| \leq \epsilon', \\ |d(o, Z_n o) - d(o, p) - d(p, Z_n o)| \leq \epsilon'. \end{aligned}$$

It follows that

$$\left| R_{n,m} - \Big(\|Z_{-(m+1)}^{-1} Z_{-1}\| + \|Z_{-1}\| + d(o,p) - d(Z_{-(m+1)}o, p) \Big) \right| \leq 2\epsilon'.$$

The same inequality holds if we replace n with n'. Hence, $R_{n,m}$ and $R_{n',m}$ differ by at most $4\epsilon'$. ☐

Note that for each ϵ', we have

$$A_{1;\epsilon'} \subseteq A_{2;\epsilon'} \subseteq \ldots, \quad \mathbb{P}(A_{k;\epsilon'}) \nearrow 1.$$

It follows that $\cup_{n>0} \cap_{k>0} A_{k;1/n}$ has measure zero. Moreover, Lemma 4.5 says that $\{R_{n,m}(\mathbf{g})\}_{n,m}$ is Cauchy for each $\mathbf{g} \in \cup_{n>0} \cap_{k>0} A_{k;1/n}$. Hence, we have:

{cor:squeezeRV}

Corollary 4.6. *In the setting of Proposition 4.4,*

$$R(\mathbf{g}) := \lim_{n,m \to +\infty} R_{n,m}(\mathbf{g})$$

is well-defined almost everywhere.

We also have the L^1-convergence of $R_{n,m}$.

{lem:squeezeL1}

Lemma 4.7. *Let* (X, G) *be as in Convention 4.3 and let* μ *be a non-elementary probability measure on* G *and let* $\epsilon > 0$ *be as in Proposition 4.4. Then for each* $\epsilon' > 0$ *there exists* $N > 0$ *such that, for every sequence of probability measures* $\{\mu_i\}_i$ *with* $\|\mu_i - \mu\| < \epsilon$ *for each* i*, we have*

$$\mathbb{E}_{\prod_{i=1}^\infty \mu_i \otimes 1_{g_0 = g} \otimes \prod_{i=1}^\infty \mu_{-i}} \left| R(\mathbf{g}) - R_{n,m}(\mathbf{g}) \right| < \epsilon'$$

for every $n, m > N$ *and for every* $g \in G$.

Proof. Observe a simple inequality: for 4 points $x, y, x', y' \in X$, we have

$$|d(x,y) - d(x',y')| \le d(x,x') + d(y,y').$$

It follows that for any $0 \le n \le n'$ and $0 \le m \le m'$, we have

$$|R_{n',m'}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \le 2 \sum_{i=n+1}^{n'} \|g_i\| + 2 \sum_{i=m+1}^{m'} \|g_{-i}\|. \tag{4.1}$$

Let $\eta = 0.01\epsilon'$. For each $n, m > N$ we claim that

(4.2) {eqn:CAT(-1)SqueezeL1}

$$|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| \le 10\eta + 2 \sum_{1 \le |i| \le N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{N-1;\eta}} + 2 \sum_{|i|>N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}.$$

To prove this, let $k = \sup\{i : \mathbf{g} \notin A_{i-1;\eta}\}$. Lemma 4.5 tells us that $|R(\mathbf{g}) - R_{n',m'}(\mathbf{g})| < 10\eta$ for any $n', m' \ge k$. In particular, Inequality 4.2 holds when $n, m \ge k$.

If $n, m < k$, then we have

$$\begin{aligned}
|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| &\le |R(\mathbf{g}) - R_{k,k}(\mathbf{g})| + |R_{k,k}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \\
&\le 10\eta + 2 \sum_{1 \le |i| \le k} \|g_i\| && (\because \text{Inequality 4.1}) \\
&\le 10\eta + 2 \sum_{1 \le |i| \le N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{N-1;\eta}} + 2 \sum_{|i|>N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}. && (\because N < n < k)
\end{aligned}$$

If $n < k \le m$, then we have

$$\begin{aligned}
|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| &\le |R(\mathbf{g}) - R_{k,m}(\mathbf{g})| + |R_{k,m}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \\
&\le 10\eta + 2 \sum_{i=n+1}^{k} \|g_i\| && (\because \text{Inequality 4.1}) \\
&\le 10\eta + 2 \sum_{|i|>N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}. && (\because N < n < k)
\end{aligned}$$

The case of $m < k \le n$ can be handled in a similar way.

We now estimate the expectation of the RHS of Inequality 4.2. For each $j \in \mathbb{Z}$ and $k > 0$ we have

$$\begin{aligned}
\mathbb{E}_{\prod_{i=1}^{\infty} \mu_i \otimes 1_{g_0 = g} \otimes \prod_{i=1}^{\infty} \mu_{-i}} \left[\|g_j\| \cdot 1_{\mathbf{g} \notin A_{k;\eta}} \right] &= \mathbb{E}_{\mu_j} \left[\|g_j\| \, \mathbb{P}(A_{k;\eta}^c | g_0, g_j) \right] \\
&\le \mathbb{E}_{\mu_j} \|g_j\| \cdot K e^{-k/K} && (\because \text{Proposition 4.4}) \\
&\le K(1+\epsilon) \|\mu\|_1 \cdot e^{-k/K}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\mathbb{E} \left[10\eta + 2 \sum_{1 \le |i| \le N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{N-1;\eta}} + 2 \sum_{|i|>N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}} \right] \\
&\le 10\eta + 2K(1+\epsilon)\|\mu\|_1 \cdot \left(N e^{-(N-1)/K} + \frac{1}{1 - e^{-1/K}} e^{-(N-1)/K} \right).
\end{aligned}$$

By taking suitably large N, we can guarantee $\mathbb{E}\,|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| < 20\eta$ for each $n, m > N$. $\square$

We now prove Theorem C that we state again:

hm:driftDiffSqueezeRe}

Theorem 4.8. *Let (X, G) be as in Convention 4.3, let μ be a non-elementary probability measure on G with finite first moment, let η be a signed measure on G such that $\|\eta\|_{0,1} < \infty$, and let $\{\mu_t : t \in [-1, 1]\}$ be a family of probability measures such that*

$$\|\mu_t - \mu - t\eta\|_{0,1} = o(t).$$

Then $l(\mu_t)$ is differentiable at $t = 0$, with derivative

$$\sigma_1(\mu, \eta) := \mathbb{E}_\eta\, d(o, go) - \mathbb{E}_{\prod_{i=1}^\infty \mu \otimes \eta \otimes \prod_{i=1}^\infty \mu}[R(\mathbf{g})]$$

Moreover, this derivative is continuous with respect to μ and η.

Proof of Theorem 4.8. Let $\epsilon > 0$ be as in Proposition 4.4. Without loss of generality we suppose $\|\mu_t - \mu\|_{0,1} \le \epsilon$ for $t \in [-1, 1]$.

Recall first that Corollary 3.4 guarantees a uniform bound K_1 such that

$$\mathbb{E}_{\mu_t^i \otimes 1_{g_0 = g} \otimes \mu^j}\, R_{i,j}(\mathbf{g}) < K_1$$

for any $i, j > 0$ and $g \in G$. Given $n > 0$ and $t \in [-1, 1]$ we define

$$E_{n,t} := \frac{1}{n} \sum_{g \in G} (\mu_t(g) - \mu(g)) \cdot \sum_{i=1}^n \mathbb{E}_{\mu_t^{n-i} \otimes 1_{g_0 = g} \otimes \mu^{i-1}}\, \mathbb{E}[R_{n-i,i-1}(\mathbf{g})].$$

Then Equation 3.3 reads

$$\frac{L(\mu_t^{*n}) - L(\mu^{*n})}{n} = \mathbb{E}_{\mu_t - \mu}\, \|g\| + E_{n,t}.$$

Now for $(g, u) \in G \times [0, 1]$ we define

$$F_{n,t}(g, u) := (\mu_t(g) - \mu(g)) \cdot \mathbb{E}_{\mu_t^{n-i} \otimes 1_{g_0 = g} \otimes \mu^{i-1}}[R_{n-i,i-1}(\mathbf{g})] \quad (i := n\lceil u/n \rceil).$$

Then $\sum_{g \in G} \int_0^1 F_{n,t}(g, u)\, du = E_{n,t}$, where du is the Lebesgue measure on $[0, 1]$. Moreover, $|F_{n,t}(g, u)|$ is pointwise dominated by $2K_1\mu(g)$, which is integrable: $\sum_{g \in G} 2K_1\mu(g) = 2K_1\|\mu\|_0 < +\infty$. Finally, for all $u \in (0, 1)$ and for all $g \in G$, we have

$$\lim_{n \to +\infty} F_{n,t}(g, u) = (\mu_t(g) - \mu(g)) \cdot \mathbb{E}_{\prod_{i=1}^\infty \mu_t \otimes 1_{g_0 = g} \otimes \prod_{i=1}^\infty \mu}\, R(\mathbf{g})$$

by Lemma 4.7. Hence, by DCT, we deduce that

$$\begin{aligned} l(\mu_t) - l(\mu) &= \mathbb{E}_{\mu_t - \mu}\, \|g\| + \lim_{n \to +\infty} E_{n,t} = \mathbb{E}_{\mu_t - \mu}\, \|g\| + \lim_{n \to +\infty} \sum_{g \in G} \int_0^1 F_{n,t}(g, u)\, du \\ &= \mathbb{E}_{\mu_t - \mu}\, \|g\| + \sum_{g \in G} \int_0^1 \left(\lim_n F_{n,t}(g, u) \right) du \\ &= \mathbb{E}_{\mu_t - \mu}\, \|g\| + \sum_{g \in G} (\mu_t(g) - \mu(g)) \cdot \mathbb{E}_{\prod_{i=1}^\infty \mu_t \otimes 1_{g_0 = g} \otimes \prod_{i=1}^\infty \mu}\, R(\mathbf{g}). \end{aligned}$$

We now divide this by t and claim

$$\lim_{t\to 0}\frac{1}{t}\big(l(\mu_t)-l(\mu)\big)=\mathbb{E}_\eta\,\|g\|-\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes\eta\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g}). \tag{4.3}$$

First, we have

$$\left|\frac{1}{t}(\mathbb{E}_{\mu_t-\mu}\,\|g\|)-\mathbb{E}_\eta\,\|g\|\right|=\frac{1}{t}\left|\mathbb{E}_{\mu_t-\mu-t\eta}\,\|g\|\right|\le\frac{1}{t}\|\mu_t-\mu-t\eta\|_1\to 0.$$

Next, note that

$$\begin{aligned}&\left|\frac{1}{t}\sum_{g\in G}(\mu_t(g)-\mu(g))\cdot\mathbb{E}_{\prod_{i=1}^\infty\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})-\sum_{g\in G}\eta(g)\cdot\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})\right|\\&\le\sum_{g\in G}\left|\frac{\mu_t(g)-\mu(g)}{t}-\eta(g)\right|\left|\mathbb{E}_{\prod_{i=1}^\infty\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})\right|\\&+\sum_{g\in G}\eta(g)\left|\mathbb{E}_{\prod_{i=1}^\infty\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})-\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})\right|.\end{aligned}$$

We want to show that both summations tend to 0. The first summation is bounded by $K_1\cdot\|\frac{\mu_t-\mu}{t}-\eta\|_0$, which tends to 0 as $t\to 0$. The second summation is over G and the summand is bounded by $2K\eta(g)$ for any t, which is integrable. Hence, we can conclude by DCT once we show for each $g\in G$ that

$$\text{\{eqn:squeezeDiffFinal\}}\lim_{t\to 0}\mathbb{E}_{\prod_{i=1}^\infty\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})=\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g}). \tag{4.4}$$

Let $\epsilon>0$ and let N be as in Lemma 4.7. Then we have

$$\begin{aligned}&\left|\mathbb{E}_{\prod_{i=1}^\infty\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})-\mathbb{E}_{\prod_{i=1}^N\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^N\mu}R(\mathbf{g})\right|<\epsilon\quad(t\in[-1,1]),\\&\left|\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g})-\mathbb{E}_{\prod_{i=1}^N\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^N\mu}R(\mathbf{g})\right|<\epsilon,\\&\lim_{t\to 0}\mathbb{E}_{\prod_{i=1}^N\mu_t\otimes 1_{g_0=g}\otimes\prod_{i=1}^N\mu}R(\mathbf{g})=\mathbb{E}_{\prod_{i=1}^N\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^N\mu}R(\mathbf{g}).\end{aligned}$$

Combining these, we obtain that the two expectations in Equation 4.4 differ by at most 2ϵ eventually. We now send $\epsilon\to 0$ to establish Equation 4.4.

We now discuss the continuity of $\sigma_1(\mu,\eta)$ with respect to μ and η. For this let μ_n,η_n be probability measures and signed measures, respectively, such that $\|\mu_n-\mu\|_{0,1},\|\eta_n-\eta\|_{0,1}\to 0$. Without loss of generality we assume $\|\mu_n-\mu\|<\epsilon$ for each n. For notational purpose we temporarily introduce

$$\begin{aligned}D_n(g)&:=\mathbb{E}_{\prod_{i=1}^\infty\mu_n\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu_n}R(\mathbf{g}),\\D(g)&:=\mathbb{E}_{\prod_{i=1}^\infty\mu\otimes 1_{g_0=g}\otimes\prod_{i=1}^\infty\mu}R(\mathbf{g}).\end{aligned}$$

In fact, the above argument for Equation 4.4 did not require the differentiability of $\{\mu_t\}_{t\in[-1,1]}$. Namely, the exactly same argument implies that

$$\lim_{n\to\infty}D_n(g)=D(g)$$

holds for each $g \in G$. Moreover, Corollary 3.4 gives the bound $D_n(g), D(g) \le K$ for all $g \in G$ and $n \in \mathbb{Z}_{>0}$.

Now given $\epsilon > 0$, let A be a finite set of G such that $\|\eta 1_{G \setminus A}\|_0 < \epsilon$. We then have

$$\begin{aligned}
|\sigma_1(\mu_n, \eta_n) - \sigma_1(\mu, \eta)| &= \left| \sum_{g \in G} \eta_n(g) D_n(g) - \sum_{g \in G} \eta(g) D(g) \right| \\
&\le \sum_{g \in A} |\eta_n(g) D_n(g) - \eta(g) D(g)| \\
&+ \sum_{g \notin A} |\eta_n(g) - \eta(g)| |D_n(g)| + \sum_{g \notin A} |\eta(g)| |D_n(g) - D(g)|.
\end{aligned}$$

Here, the first summation tends to 0 as n goes to infinity. The second summation is bounded by $K_1 \|\eta_n - \eta\|_0$, which also tends to 0. The third summation is bounded by $2K_1 \cdot \|\eta 1_{G \setminus A}\|_0 < 2K\epsilon$. We now bring ϵ down to zero and conclude. $\square$

5. Exponentially squeezing geodesics

{section:expSqueeze}

In this section, we discuss higher regularity of the escape rate on CAT(-1) spaces and their ilk. Throughout the section we assume that:

{conv:expSqueeze}

Convention 5.1. *(X, G) is either:*

- *Teichmüller space with the Weil-Petersson metric and the mapping class group;*
- *a CAT(-1) space and its countable isometry group;*
- *a CAT(0) cube complex and its countable isometry group;*
- *relatively hyperbolic group equipped with a Green metric;*
- *the standard Cayley graph of a surface group, or;*
- *the standard Cayley graph of a free product of nontrivial groups.*

{prop:expSqueeze}

Proposition 5.2. *Let (X, d) be as in Convention 5.1 and let μ be a non-elementary probability measure on G. Then there exists $\epsilon, K > 0$ such that for every sequence of probability measures $\{\mu_i\}_i$ with $\|\mu_i - \mu\|_{0,1} < \epsilon$, for every $k \in \mathbb{Z}_{>0}$, and for every combination of N integers $\{j_1, \ldots, j_N\} \subseteq \mathbb{Z} \setminus \{0\}$, we have*

$$\mathbb{P}_{\prod_i \mu_i}\left(\mathbf{g} \in A_{k; e^{-k/K}} \,\middle|\, g_0, g_{j_1}, g_{j_2}, \ldots, g_{j_N}\right) \ge 1 - Ke^{-(k-N)/K}.$$

{lem:expSqueezeL1}

Lemma 5.3. *Let (X, d) be as in Convention 5.1, let μ be a non-elementary probability measure on G and let $\epsilon, K > 0$ be as in Proposition 5.2. Let $\{\mu_i\}_i$ be a sequence of probability measures such that $\|\mu_i - \mu\| < \epsilon$ for each i. Then we have*

$$\mathbb{E}_{\prod_i \mu_i}\left[\left| R(\mathbf{g}) - R_{n,m}(\mathbf{g}) \right| \,\middle|\, g_0, g_{j_1}, g_{j_2}, \ldots, g_{j_N} \right] < K_2 e^{-(\min(n,m) - 2N)/2K} \left(1 + \sum_{l=1}^{N} \|g_{j_l}\| \right)$$

for every $n, m > k$ and for every combination of $g_{j_1}, \ldots, g_{j_N} \in G$, where K_2 is a constant that is determined by the values of ϵ, K and $\|\mu\|_1$.

Proof. For notational convenience, let $u = \min(n, m)$. As in the proof of Lemma 4.7, we will first bound $R - R_{n,m}$ with a summation of RVs related to $A_{k;e^{-k/K}}$'s and estimate the latter summation. One subtle point is that $\{A_{k;e^{-k/K}}\}_{k=1}^{\infty}$ is not nested, whereas $\{A_{k;\epsilon}\}_{k=1}^{\infty}$ is. Hence, we define

$$B_k := \cap_{i \geq k} A_{i;e^{-i/K}}.$$

This time, we claim

(5.1) {eqn:expSqueezeL1}

$$|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| \leq 10Ke^{-u/K} + \sum_{1 \leq |i| \leq u} \|g_i\| 1_{g \notin B_{u-1}} + \sum_{|i| > u} \|g_i\| \cdot 1_{g \notin B_{|i|-1}}.$$

The proof is almost identical to the one in the proof of Lemma 4.7 so we omit it.

We now estimate the expectation of the RHS of Inequality 5.1. For each $j \in \mathbb{Z} \setminus \{0\}$ and $k > 0$ we have

$$\begin{aligned}
\mathbb{E}\left[\, \|g_j\| \cdot 1_{\mathbf{g} \notin B_k} \middle|\, g_{j_1}, \ldots, g_{j_N} \right] &\leq \mathbb{E}\left[\|g_j\|\, \mathbb{P}\left(\cup_{i \geq k} A^c_{i;e^{-i/K}} \,\middle|\, g_{j_1}, \ldots, g_{j_N}, g_j \right) \right] \\
&\leq \mathbb{E}_{\mu_j} \|g_j\| \cdot \frac{K}{1 - e^{-1/K}} e^{-(k-N)/K} \qquad (\because \text{Proposition 5.2}) \\
&\leq \frac{K}{1 - e^{-1/K}} e^{-(k-N)/K} (1 + \epsilon) \|\mu\|_1
\end{aligned}$$

if $j \notin \{j_1, \ldots, j_m\}$, and we get $\leq \frac{K}{1-e^{-1/K}} e^{-(k-N)/K} \|g_{j_l}\|$ for each $j = j_l$.

Hence, we have

$$\begin{aligned}
&\mathbb{E}\left[10Ke^{-u/K} + 2 \sum_{1 \leq |i| \leq u} \|g_i\| \cdot 1_{\mathbf{g} \notin B_{u-1}} + 2 \sum_{|i| > u} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\epsilon}} \right] \\
&\leq 10Ke^{-u/K} + \frac{2K}{1 - e^{-1/K}} (1 + \epsilon) \|\mu\|_1 \cdot \left(ue^{-(u-1-N)/K} + \frac{1}{1 - e^{-1/K}} e^{-(u-1-N)/K} \right) \\
&+ \frac{K}{1 - e^{-1/K}} e^{-(u-1-N)/K} \sum_{l=1}^{N} \|g_{j_l}\| \\
&\leq K_2 e^{-(u-2N)/2K} \left(1 + \|g_{j_1}\| + \ldots + \|g_{j_N}\|\right)
\end{aligned}$$

for $K_2 = 100K(1 + \epsilon)e^{1/K}(2K/e)\|\mu\|_1(1 - e^{-1/K})^{-2}$. □

In fact, the constant ϵ, K we obtained in Proposition 5.2 only depends on how much weight μ^{*n} puts on independent contracting isometries $g, h \in G$. Together with this fact, Lemma 4.7 will imply Theorem E. We will come back to this point after the proof of Proposition 5.2.

We now prove Theorem D. We invite the readers to recall Notation 2.3.

Proof. Let ϵ, K, K_1, K_2 be the constants that satisfy the conclusions of Corollary 3.4, Proposition 5.2, and Lemma 5.3. Without loss of generality, we assume that $\|\mu_t - \mu\|_{0,1} < \epsilon$ for all $t \in [-1, 1]$.

We will first show that

(5.2) {eqn:secondDiff}

$$\begin{aligned}\sigma_1(\mu_t, \eta_t) - \sigma_1(\mu, \eta) &= \mathbb{E}_{\eta_t - \eta}\, d(o, go) - \mathbb{E}_{\prod_{i=1}^{\infty} \mu_t \otimes (\eta_t - \eta) \otimes \prod_{i=1}^{\infty} \mu_t}[R(\mathbf{g})] \\ &\quad - \sum_{k=1}^{\infty} \mathbb{E}_{\left(\prod_{i=1}^{k-1} \mu \times (\mu_t - \mu) \times \prod_{i=k+1}^{\infty} \mu_t\right) \otimes \eta \otimes \left(\prod_{i=1}^{k-1} \mu \times \prod_{i=k}^{\infty} \mu_t\right)}[R(\mathbf{g})] \\ &\quad - \sum_{k=1}^{\infty} \mathbb{E}_{\left(\prod_{i=1}^{k} \mu \times \prod_{i=k+1}^{\infty} \mu_t\right) \otimes \eta \otimes \left(\prod_{i=1}^{k-1} \mu \times (\mu_t - \mu) \times \prod_{i=k}^{\infty} \mu_t\right)}[R(\mathbf{g})]\end{aligned}$$

and that the summands of the two summations are exponentially decreasing with respect to k. For this let us bring finite truncations of these summations: we have

(5.3)

$$\begin{aligned}\sigma_1(\mu_t, \eta_t) - \sigma_1(\mu, \eta) &= \mathbb{E}_{\eta_t - \eta}\, d(o, go) - \mathbb{E}_{\prod_{i=1}^{\infty} \mu_t \otimes (\eta_t - \eta) \otimes \prod_{i=1}^{\infty} \mu_t}[R(\mathbf{g})] \\ &\quad - \sum_{k=1}^{N} \mathbb{E}_{\left(\prod_{i=1}^{k-1} \mu \times (\mu_t - \mu) \times \prod_{i=k+1}^{\infty} \mu_t\right) \otimes \eta \otimes \left(\prod_{i=1}^{k-1} \mu \times \prod_{i=k+1}^{\infty} \mu_t\right)}[R(\mathbf{g})] \\ &\quad - \sum_{k=1}^{N} \mathbb{E}_{\left(\prod_{i=1}^{k} \mu \times \prod_{i=k+1}^{\infty} \mu_t\right) \otimes \eta \otimes \left(\prod_{i=1}^{k-1} \mu \times (\mu_t - \mu) \times \prod_{i=k}^{\infty} \mu_t\right)}[R(\mathbf{g})] \\ &\quad - \mathbb{E}_{\left(\prod_{i=1}^{N} \mu \times \prod_{i=N+1}^{\infty} \mu_t\right) \otimes \eta \otimes \left(\prod_{i=1}^{N} \mu \times \prod_{i=N+1}^{\infty} \mu_t\right)}[R(\mathbf{g})] \\ &\quad + \mathbb{E}_{\left(\prod_{i=1}^{N} \mu \times \prod_{i=k+1}^{\infty} \mu\right) \otimes \eta \otimes \left(\prod_{i=1}^{N} \mu \times \prod_{i=N+1}^{\infty} \mu\right)}[R(\mathbf{g})].\end{aligned}$$

Here, Lemma 5.3 guarantees for each $\epsilon' > 0$ a threshold $N_{\epsilon'}$ such that

$$\left| \mathbb{E}_{\prod_{i=1}^{\infty} \mu \otimes 1_{g_0 = g} \otimes \prod_{i=1}^{\infty} \mu}[R(\mathbf{g})] - \mathbb{E}_{\prod_{i=1}^{N_{\epsilon'}} \mu \otimes 1_{g_0 = g} \otimes \prod_{i=1}^{N_{\epsilon'}} \mu}[R(\mathbf{g})] \right| \le \epsilon'$$

for every $g \in G$. Moreover, for any $N > N_{\epsilon'}$ we have

$$\left| \mathbb{E}_{\left(\prod_{i=1}^{N} \mu \times \prod_{i=N+1}^{\infty} \mu_t\right) \otimes 1_{g_0 = g} \otimes \prod_{i=1}^{N} \mu \times \prod_{i=N+1}^{\infty} \mu_t}[R(\mathbf{g})] - \mathbb{E}_{\prod_{i=1}^{N_{\epsilon'}} \mu \otimes 1_{g_0 = g} \otimes \prod_{i=1}^{N_{\epsilon'}} \mu}[R(\mathbf{g})] \right| \le \epsilon'$$

for every $g \in G$. From these estimates, we deduce that the final two terms of Equation 5.3 differ by $2\epsilon'$ whenever $N > N_{\epsilon'}$. By sending ϵ' to zero we establish Equation 5.2, given that the summands in the summations are summed up in the correct order. It remains to show that the summands are absolutely summable.

Note that

(5.4)

$$\begin{aligned}
&\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times(\mu_t-\mu)\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}[R(\mathbf{g})]\\
&=\sum_{g\in G}(\mu_t(g)-\mu(g))\\
&\quad\cdot\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}\left[\left(R(\mathbf{g})-R_{k-1,k-1}(\mathbf{g})\right)+R_{k-1,k-1}(\mathbf{g})\right].
\end{aligned}$$

Here, note that the RV $R_{k-1,k-1}(\mathbf{g})$ is independent from g_k. That means,

$$\begin{aligned}
&\sum_{g\in G}\mu_t(g)\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}R_{k-1,k-1}(\mathbf{g})\\
&=\mathbb{E}_{\prod_{i=1}^{k-1}\mu\otimes\eta\otimes\prod_{i=1}^{k-1}\mu}R_{k-1,k-1}(\mathbf{g})\\
&=\sum_{g\in G}\mu(g)\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}R_{k-1,k-1}(\mathbf{g})
\end{aligned}$$

holds, Hence, the LHS of Equation 5.4 is in fact bounded by

$$\begin{aligned}
&\sum_{g\in G}|\mu_t(g)-\mu(g)|\cdot\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}|R(\mathbf{g})-R_{k-1,k-1}(\mathbf{g})|\\
&\le\sum_{g\in G}|\mu_t(g)-\mu(g)|\cdot K_2e^{-(k-1)/K}(1+\|\eta\|_{0,1}+\|g\|) && (\because\text{ Lemma 4.}\\
&\le K_2e^{-(k-1)/K}(\|\mu_t-\mu\|_0(1+\|\eta\|_{0,1})+\|\mu_t-\mu\|_1).
\end{aligned}$$

This decays exponentially and the summations in Equation 5.2 are absolutely summable.

We now divide both hand sides of Equation 5.2 by t and take the limit. First note that we still have absolutely summable summands:

(5.5) {eqn:squeeze2ndDerivUniform}

$$\begin{aligned}
&\frac{1}{t}\left|\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times(\mu_t-\mu)\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}R(\mathbf{g})\right|\\
&\le K_2e^{-(k-1)/K}\left(\frac{\|\mu_t-\mu\|_0}{t}(1+\|\eta\|_{0,1})+\frac{\|\mu_t-\mu\|_1}{t}\right),
\end{aligned}$$

and note that

$$\lim_{t\to 0}\frac{\|\mu_t-\mu\|_i}{t}=\|\eta\|_i$$

for $i=0,1$. Given this summable bounds, we now have to show the termwise convergence, i.e., that

$$\begin{aligned}
&\lim_{t\to 0}\frac{1}{t}\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times(\mu_t-\mu)\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)}R(\mathbf{g})\\
&=\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu\right)}R(\mathbf{g}).
\end{aligned}$$

Note that

$$
\begin{aligned}
&\frac{1}{t}\left|\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times(\mu_t-\mu-t\eta)\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} R(\mathbf{g})\right| \\
&\le \sum_{g\in G}\frac{1}{t}|\mu_t(g)-\mu(g)-t\eta(g)|\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} R(\mathbf{g}) \\
&\le \sum_{g\in G}\frac{1}{t}|\mu_t(g)-\mu(g)-t\eta(g)|\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} |R(\mathbf{g})-R_{k-1,k-1}(\mathbf{g})| \\
&+ \sum_{g\in G}\frac{1}{t}|\mu_t(g)-\mu(g)-t\eta(g)|\,\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times 1_{g_k=g}\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} |R_{k-1,k-1}(\mathbf{g})| \\
&\le \sum_{g\in G}\frac{1}{t}|\mu_t(g)-\mu(g)-t\eta(g)|K_2e^{-(k-3)/K}(1+\|g\|) \\
&+ \sum_{g\in G}\frac{1}{t}|\mu_t(g)-\mu(g)-t\eta(g)|K_1,
\end{aligned}
$$

and the RHS tends to 0 as $t\to 0$. It remains to show:

$$
\lim_{t\to 0}\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} R(g) = \mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu\right)\otimes\eta\otimes\left(\prod_{i=1}^{\infty}\mu\right)} R(g). \tag{5.6}
$$

Again, given $\epsilon' > 0$, Lemma 4.7 guarantees $N > 0$ such that

$$
\begin{aligned}
&\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{\infty}\mu_t\right)} |R(g)-R_{N,N}(g)| < \epsilon', \\
&\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu\right)\otimes\eta\otimes\left(\prod_{i=1}^{\infty}\mu\right)} |R(g)-R_{N,N}(g)| < \epsilon'.
\end{aligned}
$$

Note that $\|\mu_t-\mu\|_{0,1}\to 0$ as $t\to 0$ and $R_{N,N}$ continuously depends on the distribution of $(g_{-N},\ldots,g_N)$. Thus, the expectations of $R_{N,N}(\mathbf{g})$ with respect to $\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{N}\mu_t\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\times\prod_{i=k}^{N}\mu_t\right)$ and $\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{N}\mu\right)\otimes\eta\otimes\left(\prod_{i=1}^{N}\mu\right)$ differ by at most ϵ' for small enough t. By decreasing ϵ' down to 0, we conclude Equation 5.6.

We now discuss the continuity of $\sigma_2(\mu,\eta)$ with respect to μ and η.

For this let μ_n be probability measures and η_n,η_n' be signed measures such that

$$
\lim_{n\to\infty}\|\mu_n-\mu\|_{0,1} = \lim_{n\to\infty}\|\eta_n-\eta\|_{0,1} = \lim_{n\to\infty}\|\eta_n'-\eta\|_{0,1} = 0.
$$

Without loss of generality we assume $\|\mu_n-\mu\|_{0,1} < \epsilon$ for all n. Note that $\sigma_2(\mu_n,\eta_n)$ is a linear combination of two terms

$$
\mathbb{E}_{\eta_n'}\|g\|, \quad \mathbb{E}_{\prod_{i=1}^{\infty}\mu_n\otimes\eta_n'\otimes\prod_{i=1}^{\infty}\mu_n} R(\mathbf{g})
$$

and countably many terms

$$
\mathbb{E}_{\left(\prod_{i=1}^{k-1}\mu_n\times\eta_n\times\prod_{i=k+1}^{\infty}\mu_n\right)\otimes\eta_n\otimes\left(\prod_{i=1}^{\infty}\mu_n\right)}[R(\mathbf{g})] \quad (k=1,2,\ldots)
$$

and their symmetric counterparts. Here, we have an exponentially decaying bounds for the countably many terms: Inequality 5.5 gives

$$\mathbb{E}_{(\prod_{i=1}^{k-1}\mu_n\times\eta_n\times\prod_{i=k+1}^{\infty}\mu_n)\otimes\eta_n\otimes(\prod_{i=1}^{k-1}\mu_n)}[R(\mathbf{g})] < K_2 e^{-(k-1)/K}\left(\|\eta_n\|_0(1+\|\eta_n\|_{0,1})+\|\eta_n\|_1\right)$$

which is uniformly summable. Given this, the convergence $\sigma_2(\mu_n,\eta_n)\to\sigma_2(\mu,\eta)$ follows from termwise convergence. To show this, let us fix $k>0$. Given $\epsilon'>0$, Lemma 5.3 provides a number N such that

$$\mathbb{E}_{\prod_i \mu_n}\left[|R(\mathbf{g})-R_{N,N}(\mathbf{g})|\,\Big|\, g_0, g_k\right] < \epsilon'(1+\|g_k\|)$$

for every $n\in\mathbb{Z}_{>0}$ for every $n,m\geq N$. This implies that

$$\left|\mathbb{E}_{(\prod_{i=1}^{k-1}\mu_n\times\eta_n\times\prod_{i=k+1}^{\infty}\mu_n)\otimes\eta_n\otimes(\prod_{i=1}^{\infty}\mu_n)} R(\mathbf{g}) - \mathbb{E}_{(\prod_{i=1}^{k-1}\mu_n\times\eta_n\times\prod_{i=k+1}^{\infty}\mu_n)\otimes\eta_n\otimes(\prod_{i=1}^{\infty}\mu_n)} R_{N,N}(\mathbf{g})\right|$$
$$< \epsilon'(\|\eta_n\|_0+\|\eta_n\|_1).$$

Moreover, $R_{N,N}(\mathbf{g})$ is continuous with respect to the underlying measure. Combining these facts, we deduce that the expectations of $R(\mathbf{g})$ with respect to $\left(\prod_{i=1}^{k-1}\mu_n\times\eta_n\times\prod_{i=k+1}^{\infty}\mu_n\right)\otimes\eta_n\otimes\left(\prod_{i=1}^{k-1}\mu_n\right)$ and $\left(\prod_{i=1}^{k-1}\mu\times\eta\times\prod_{i=k+1}^{\infty}\mu\right)\otimes\eta\otimes\left(\prod_{i=1}^{k-1}\mu\right)$ eventually differ by at most $2\epsilon'\|\eta\|_{0,1}$. We now bring ϵ' down to zero and conclude the termwise convergence. □

6. Probablistic estimates for alignments

{section:squeezeProb}

In this section, we give proofs of Proposition 4.4 and 5.2.

{subsection:squeezeSch

6.1. Squeezing Schottky sets.

First, we recall a more precise version of Proposition 2.8. Given sequences $s_i=(a_{n(i-1)+1},\ldots,a_{ni})\in G^n$ for $i=1,\ldots,k$, by abuse of notation, we denote $(a_1,\ldots,a_{nk})$ by $(s_1,\ldots,s_k)$. We also denote the k-self-concatenation of $s\in G^n$ by $s^{(k)}$.

{fact:SchottkyRefine}

Fact 6.1. *Let (X,d) be a metric space and let $\alpha,\beta\in G^n$ be sequences of isometries of X such that $\Pi(\alpha)$ and $\Pi(\beta)$ are independent contracting isometries of X. Then there exist $K,M>0$ such that for each $k,N>0$,*

$$S := \left\{(s_1,\ldots,s_k,\underbrace{\alpha,\ldots,\alpha}_{N\ copies},s_{k+1},\ldots,s_{2k}) : s_i\in\{\alpha^{(M)},\beta^{(M)}\}\right\}\subseteq G^{Mn(2k+N)}$$

is a kK-Schottky set.

For a proof, refer to [Cho22a, Proposition 3.18]. We now say that a Schottky set S is *ϵ-squeezing* if $[o,\Pi(s)]$ is ϵ-squeezing for each $s\in S$.

On CAT(-1) space, if a K-Schottky set S is contained in G^n for $n>Ke^{100/\epsilon}+K$, then $[o,\Pi(s)]$ is longer than $e^{100/\epsilon}$ and hence ϵ-squeezing for each $s\in S$. In view of Fact 6.1, for all non-elementary probability measure μ on G there exists such an ϵ-squeezing, long enough Schottky set for μ.

Next, let (X,d) be Teichmüller space with either metric, and let μ be a non-elementary probability measure on the mapping class group. Then for some $n>0$, there exist $\alpha=(a_1,\ldots,a_n)$ and $\beta=(b_1,\ldots,b_n)$ in $(\operatorname{supp}\mu)^n$

that result in independent pseudo-Anosov mapping classes $\Pi(\alpha)$ and $\Pi(\beta)$. Let K, M be as in Fact 6.1 for α and β andconsider the set

$$S_N := \left\{ (s_1, \ldots, s_{10}, \underbrace{\alpha, \ldots, \alpha}_{N \text{ copies}}, s_{11}, \ldots, s_{20}) : s_i \in \{\alpha^{(M)}, \beta^{(M)}\} \right\}.$$

By Fact 6.1, S_N is a $10K$-Schottky set for each N. Moreover, there exists a uniform fellow-traveling constant K' that only depends on K, M, α, β and not on N, such that: for each element $s \in S_N$, the geodesic $[o, \Pi(s)o]$ is K'-fellow traveling with $[go, g\Pi(\alpha^N)o]$ for some $g \in G$. Recall that suitably long pseudo-Anosov axes are ϵ-squeezing. Hence, for each $\epsilon > 0$ there exists N such that each geodesic $[o, \Pi(s)o]$ made with $s \in S_N$ is ϵ-squeezing.

This construction does not only apply to Teichmüller space. If the isometry group G of X contains a contracting isometries g such that for each $\epsilon > 0$ $[o, g^N o]$ is ϵ-squeezing for sufficiently large N, then the construction works the same. This construction equally applies to non-geodesic spaces that possess ϵ-squeezing isometries as in Definition 4.2. Hence, we have:

{prop:SchottkySqueeze}

Proposition 6.2. *Let (X, G) be as in Convention 4.3. Then for each $\epsilon > 0$ and $N \in \mathbb{Z}_{>0}$, there exists an ϵ-squeezing long enough Schottky set for μ with cardinality N.*

In particular, CAT(0) cube complices, surface groups and free products even contain 0-squeezing isometries. Using them, we obtain:

prop:Schottky0Squeeze}

Proposition 6.3. *Let (X, G) be a CAT(0) cube complex with its automorphism group or the standard Cayley graph of a surface group or a free product of nontrivial groups. Then for each $N \in \mathbb{Z}_{>0}$, there exists a 0-squeezing long enough Schottky set for μ with cardinality N.*

subsection:expSqueeze}

6.2. **Exponential squeezing.** On spaces as in Convention 5.1, we expect further squeezing along random paths.

Let X be a CAT(-1) space, with the constant $c > 1$ as in Fact 2.20. Let $\gamma_1 = [x_1, y_1], \ldots, \gamma_N = [x_N, y_N]$ be geodesics of length L and let γ be geodesics with subsegments $\eta_1, \ldots, \eta_N$, ordered from left to right, such that γ_i and η_i are $0.01L$-fellow traveling for each i. Then γ contains points x, y such that $d(x, x_1) < 0.01L$, $d(y, y_N) < 0.01L$ and $d(x_1, y_N) > 0.9NL$. By Fact 2.20, γ is $c^{-0.9NL}$-close to the midpoint of $[x_1, y_N]$. From this, we conclude:

lem:CAT(-1)expSqueeze}

Lemma 6.4. *Let X be a CAT(-1) space and S be a long enough K-Schottky set in the isometry group G of X. Then there exists $K' > 0$ such that the following holds.*

Let $x, y \in X$ and let $\gamma_1 = [x_1, y_1], \ldots, \gamma_N = [x_N, y_N]$ be Schottky axes such that $(x, \gamma_1, \ldots, \gamma_N, y)$ is D_0-semi-aligned. Then $[x, y]$ is $e^{-N/K'}$-close to the midpoint of $[x_1, y_N]$.

A similar argument works for Weil-Petersson Teichmüller space and Green metrics on relatively hyperbolic group, with help of Proposition 2.12 and Proposition 2.22 respectively. Hence, we have:

{lem:otherexpSqueeze}

Lemma 6.5. *Let X be a Teichmüller space with the Weil-Petersson metric or a relatively hyperbolic group with a Green metric, and let S be a long enough K-Schottky set in the isometry group G of X. Then there exists $K' > 0$ such that the following holds.*

Let $x, y \in X$ and let $\gamma_1 = [x_1, y_1], \ldots, \gamma_N = [x_N, y_N]$ be Schottky axes such that $(x, \gamma_1, \ldots, \gamma_N, y)$ is D_0-semi-aligned. Then $[x_1, y_N]$ is $e^{-N/K}$-squeezing.

For CAT(0) cube complices, surface groups and free products, we do not even have to consider alignment with several squeezing segments, because they possess 0-squeezing isometries: the alignment with a single Schottky segment guarantees 0-bottlenecking.

{subsection:probEst}

6.3. **Probabilistic estimates.** Given Proposition 6.2, Proposition 6.3 Lemma 6.4 and Lemma 6.5, we now introduce the final probabilistic estimate that leads to Proposition 4.4 and 5.2.

Let (X, d) be a metric space, let G be a non-elementary isometry group of X and let S be a long enough K_0-Schottky set in G^n for some n, associated with constants D_0, E_0 as in Definition 2.7. We now define B_k to be the collection of step paths $\mathbf{g} \in G^{\mathbb{Z}}$ for which there exist Schottky axes $\Gamma_1, \ldots, \Gamma_k$, $\Gamma'_1, \ldots, \Gamma'_k$ satisfying the following for each $m, n \geq k$:

(1) $(Z_{-m}o, \Gamma'_k, \ldots, \Gamma'_1, \Gamma_1, \ldots, \Gamma_k, Z_n o)$ is D_0-semi-aligned, and
(2) $(\Gamma'_1, Z_{-1}o)$ and (o, Γ_1) are each D_0-semi-aligned.

{prop:conditionedDistS

Proposition 6.6. *Let μ be a non-elementary probability measure on G and let S be a long enough Schottky set for μ with cardinality 400. Then there exist $\epsilon, K > 0$ such that for all sequence of probability measures $\{\mu_i\}_{i \in \mathbb{Z}}$ with $\|\mu_i - \mu\|_{0,1} < \epsilon$ and for all combination of m integers $\{j_1, \ldots, j_m\} \in \mathbb{Z} \setminus \{0\}$, we have*

$$\mathbb{P}_{\prod_{i \in \mathbb{Z}} \mu_i} \left(\mathbf{g} \in B_{k; \lfloor k/K \rfloor} \,|\, g_{j_1}, g_{j_2}, \ldots, g_{j_m} \right) \geq 1 - K e^{-(k-m)/K}.$$

Proof. Let M_0 be the length of sequences in S, i.e., $S \subseteq G^{M_0}$. Recall that for fixed isometries $w_i \in G$ and $a_i, b_i, c_i, d_i, \ldots$ chosen independently from a long enough Schottky set S with the uniform distribution, we constructed a word

$$W_n = w_0 \Pi(a_1) \Pi(b_1) \Pi(c_1) \Pi(d_1) w_1 \cdots \Pi(a_n) \Pi(b_n) \Pi(c_n) \Pi(d_n) w_n.$$

and defined the *set of pivotal times* $P_n(\{w_{i-1}, a_i, \ldots, d_i\}_i)$. This time, we will use a variant of Fact 3.2:

{fact:pivotAlign2}

Fact 6.7 ([Cho22a, Proposition 5.1])**.** *If a choice $(a_i, b_i, c_i, d_i, w_{i-1}, v_i)_{i>0} \in (S^{\mathbb{N}})^4 \times G^{\mathbb{N}}$ has the set of pivotal times $P_n = \{j(1) < \ldots < j(m)\}$, then the sequence*

$$(o, U_{j(1)} \Gamma(b_{j(1)}), \ldots, U_{j(m)} \Gamma(b_{j(m)}), W_n o)$$

is D_0-semi-aligned, where $U_{j(i)} := W_{j(i)-1} \Pi(a_{j(i)})$.

We now list some more facts on the set of pivotal times.

{fact:listPivot} **Fact 6.8.** [Cho22a, Lemma 5.4, Lemma 5.5, Proposition 5.6, Lemma 4.4]

(1) $P_n(\{w_i\}_{i=0}^n, \{a_i, b_i, c_i, d_i\}_{i=1}^n)$ *is either* $P_n \cup \{n\}$ *or an initial section of* P_n*, i.e.,* $P_n \cap \{1, \ldots, k\}$ *for some* $k < n$*.*

(2) Fixing $\{w_i\}_{i=0}^n \in G^n$ *and* $\{a_i, c_i, d_i\}_{i=1}^n \in (S^3)^n$*, let* $\sim$ *be a relation on* S^n*:*

$$\{b_i\}_{i=1}^n \sim \{b_i'\}_{i=1}^n \;\Leftrightarrow\; b_i = b_i' \text{ unless } i \notin P_n(\{w_{i-1}, a_i, b_i, c_i, d_i\}_i).$$

Then $\sim$ *is an equivalence relation, each equivalence class endowed with the uniform measure* $S^{\#P_n}$*.*

(3) Fixing $(w_i)_i$ *and choosing* $(a_i, b_i, c_i, d_i)_i$ *from* $(S^4)^{\mathbb{N}}$ *with i.i.d. uniform measure, we have*

$$\mathbb{P}\left(\#P_n(\{a_i, b_i, c_i, d_i, w_i\}_i) = \#P_{n-1}(\{a_i, b_i, c_i, d_i, w_i\}_i) + 1\right) \geq 0.9$$

and

$$\mathbb{P}\left(\#P_n(\{a_i, b_i, c_i, d_i, w_i\}_i) < \#P_{n-1}(\{a_i, b_i, c_i, d_i, w_i\}_i) - j\right) \leq (0.1)^{j+1}$$

for each $j > 0$*.*

(4) Let $(u_i, u_i')_{i=0}^n$ *be a sequence in* G^{2n+2} *and pick* $(b_i, b_i')_{i=1}^n$ *from* S^{2n} *with uniform measure. Then we have*

$$\mathbb{P}\left(\begin{array}{c} \left(\Gamma(b_i'),\, \Pi(b_i')u_{(i-1)}' \cdots \Pi(b_1')u_0' \cdot u_0\Pi(b_1) \cdots u_{i-1} \cdot \Gamma(b_i)\right) \\ \textit{is } D_0\textit{-aligned for some } i < k \end{array} \right) \geq 1 - 0.1^k.$$

We now proceed to the proof. Since S is a long enough Schottky set for μ, there exist ϵ and β such that

$$\mu_{4M_0k+1} \times \cdots \times \mu_{4M_0(k+1)} = \beta(\text{uniform measure on } S)^4 + (1-\beta)\nu_k$$

for probability measures μ_i's with $\|\mu_i - \mu\| < \epsilon$. By taking large enough K, we may only prove for the case $k > 20M_0m$. This time, we take disjoint closed intervals

$$I_1 = [4M_0k_1 + 1, 4M_0(k_1+1)],\; I_2 = [4M_0(k_2+1), 4M_0(k_2+1)],\; \ldots$$

of length $4M_0 - 1$ on $\mathbb{Z}_{>0} \setminus \{j_1, \ldots, j_m\}$. We can ensure that $[1, n]$ contains at least $n/4M_0 - m$ such intervals for each n. We again independently pick each interval I_i with probability β to make a collection $\mathcal{A}$ and do the construction as in the proof of Lemma 3.1. Then we get a measurable partition $\{\mathcal{F}_\alpha\}_\alpha$ of $(G^{\mathbb{Z}}, \mu^{\mathbb{Z}})$ such that, each $\mathcal{F}_\alpha$ is associated with a set $\mathcal{A}$ of intervals, fixed isometries $\{w_i\}_i$, and sequences $(a_i, b_i, c_i, d_i)_i$ chosen independently from S, and such that

$$g_1 \cdots g_n = w_0\Pi(a_1)\Pi(b_1)\Pi(c_1)\Pi(d_1)w_1 \cdots \Pi(a_{T(n)})\Pi(b_{T(n)})\Pi(c_{T(n)})\Pi(d_{T(n)}) \cdot w^{(n)},$$

for each n; here, $T(n) = \#\{I \in \mathcal{A} : I \subseteq [0, n]\}$ and $w^{(n)}$ is a word depending on the index n:

$$w^{(n)} := g_{4M_0(k_{T(n)}+1)+1} \cdots g_{n-1}g_n.$$

We now define events:
$E'_n = \{(g_i)_{i>0} : T(n) \geq 0.1\beta n/M_0\}$,
$E''_n = \Big\{(g_i)_{i>0} : \#P_{n+1}\big(\{w_i\}_{i=0}^n, \{a_i, \ldots, d_i\}_{i=1}^n\big) \geq 0.1n\Big\}$,

$$E'''_n = \left\{ (g_i)_{i>0} : \begin{array}{c} \#P_{T(n)}\left(\{w_0, \ldots, w_{T(n)-1}, w^{(n)}\}, \{a_i, \ldots, d_i\}_{i=1}^{T(n)}\right) \\ \geq \#P_{T(n)-1}\left(\{w_i\}_{i=0}^{T(n)-1}, \{a_i, \ldots, d_i\}_{i=1}^{T(n)-1}\right) - T(n)/2K' \end{array} \right\}.$$

First, note that $[0, n]$ contains at least $n/4M_0 - m \geq n/5M_0$ intervals I_i's, since $n \geq k \geq 20M_0 m$. Each interval is then collected by $\mathcal{A}$ for probability β, so Hoeffding's inequality tells us that

$$\mathbb{P}(E'_n) \geq 1 - e^{-\beta k/40M_0} \quad (n \geq k).$$

Next, by Fact 3.3 we have

$$\mathbb{P}(E''_n) \geq 1 - K' e^{-n/K'}.$$

Finally, by Fact 6.8 (3), we have

$$\mathbb{P}(E'_n \cap E'''_n) \geq 1 - 0.1^{0.1\beta n/2K'M_0}.$$

In summary, we have

$$\mathbb{P}\left(E := \left(\cap_{n \geq k}(E'_n \cap E'''_n)\right) \cap \left(\cap_{n \geq 0.1\beta k/M_0} E''_n\right)\right) \geq 1 - K'' e^{-k/K''}$$

for some suitable $K'' > 0$. Now for a path $\mathbf{g} \in E$, $\#P_{n+1}(\{w_i\}_i, \{a_i, d_i\}_i)$ for $n \geq k$ share the initial section of length k/K'. Moreover, for each $n \geq k$, $\#P_{T(n)}(\{w_0, \ldots, w_{T(n)-1}, w^{(n)}\}, \{a_i, \ldots, d_i\}_{i=1}^{T(n)}$ and $P_{T(n)-1}(\{w_i\}_i, \{a_i, \ldots, d_i\}_i)$ share the initial section of length $T(k)/2K' \geq 0.1\beta k/2K'M_0$. In summary, if we denote the first $0.1\beta k/2K'M_0$ elements of $P_{0.1\beta k/K'M_0}(\mathbf{g})$ by $j(1), \ldots, j(0.1\beta k/2K'M_0)$, then

$$(o, \gamma_1, \ldots, \gamma_{0.1\beta k/2K'M_0}, Z_n o)$$

is D_0-semi-aligned for each $n \geq k$, where $\gamma_i := U_{j(i)}\Gamma(b_{j(i)})$. Let us now consider an equivalence relation $\sim$ as in Fact 6.8 (2), for $n = 0.1\beta k/2K'M_0$. On each equivalence class $\mathcal{E}$, $(b_{j(1)}, \ldots, b_{j(0.1\beta k/2K'M_0)})$ are chosen from the uniform measure on $S^{0.1\beta k/2K'M_0}$, and we have

$$\gamma_i = u_0 \Pi(b_{j(1)}) u_1 \cdots \Pi(b_{j(i-1)}) u_{i-1} \cdot \Gamma(b_i),$$

where $u_i := U_{j(i)}^{-1} U_{j(i+1)}$'s are fixed on $\mathcal{E}$.

For the very same reason, by considering the pivotal times for the backward path, we can construct a measurable set $E' \subseteq G^{\mathbb{Z}_{<0}}$ with $\mathbb{P}(E') \geq 1 - K'' e^{-k/K''}$ such that for each $\mathbf{g} \in E'$ there exists Schottky axes $\gamma'_1, \ldots, \gamma_{0.1\beta k/2K'M_0}$ for which

$$(Z_{-m} o, \gamma'_{0.1\beta k/2K'M_0}, \ldots, \gamma'_1, Z_{-1} o)$$

is D_0-semi-aligned for each $m \geq k$. Furthermore, we have an equivalence relation on $G^{\mathbb{Z}_{<0}}$ such that each equivalence class $\mathcal{E}'$ is associated with fixed isometries u_i''s and RVs $(b'_{j(1)}, \ldots, b'_{j(0.1\beta k/2K'M_0)}) \in S^{0.1\beta k/2K'M_0}$ following the uniform distribution, and such that

$$\gamma_i' = (\Pi(b_i')u_{i-1}' \cdots \Pi(b_{-1})u_0')^{-1}\Gamma(b_i')$$

for each i. Now Fact 6.8(4) tells us that on each pair of equivalence classes $\mathcal{E} \times \mathcal{E}'$, $(\gamma'_{0.1\beta k/4K'M_0}, \gamma_{0.1\beta k/4K'M_0})$ is D_0-semi-aligned for probability at least $1 - 0.1^{0.1\beta k/4K'M_0}$. Summing up the conditional inequality, we finally observe that

$$(\gamma'_{0.1\beta k/4K'M_0}, \ldots, \gamma'_{0.1\beta k/2K'M_0}, \gamma_{0.1\beta k/4K'M_0}, \ldots, \gamma_{0.1\beta k/2K'M_0})$$

is D_0-semi-aligned for probability $1 - 0.1^{0.1\beta k/4K'M_0}$. □

We are now ready to prove Theorem E.

Proof. From Lemma 5.3, we have

$$\mathbb{E}_{\prod_{i>0}\mu\times\eta\times\prod_{i>0}}[R(\mathbf{g})] - \mathbb{E}_{\prod_{i>0}\mu\times\eta\times\prod_{i>0}}[R_{n,n}(\mathbf{g})] \leq K_2 e^{-(n-1)/K}(1 + \|\eta\|_1).$$

Fixing the constants K_2, K and ϵ, $n = O(|\log \epsilon_1|)$ gives error bounded by ϵ_1. Also recall that K_2 is determined by ϵ, K and $\|\mu\|_1$.

It remains to show that ϵ and K that come from Proposition 6.6 can be determined by ϵ, g, h and $\mu(g)$, $\mu(h)$. In the proof of Proposition 6.6, ϵ and K are determined by β, K' and M_0, which are in turn determined by:

(1) the long enough Schottky set S for μ, including the length M_0 of sequences in S;
(2) the minimum weight $\min\{\mu(a_1)\cdots\mu(a_{M_0}) : (a_1, \ldots, a_{M_0}) \in S\}$ that μ^{M_0} puts on each $s \in S$.

Since we can construct such a Schottky set S by composing g and h, these values are controlled by the norm and the contracting power of g and h, and the weights $\mu(g)$, $\mu(h)$ that μ puts on them. This ends the proof. □

{section:entropy}

7. Continuity of the entropy

In this section, we prove Theorem F. We first recall a result regarding sublinear growth of entropy of displacement.

prop:sublinearEntropy}

Proposition 7.1 ([CFFT22, Proposition 4.5, 4.6])**.** . *Let (X, d) be a metric space involving contracting isometries and let G be a countable, non-elementary isometry group of X. Let μ and ν be non-elementary probability measures on G with finite one-time entropy. Then for each $\epsilon > 0$ there exist $\epsilon_1, N_1 > 0$ such that for every probability measures μ_i on G that*

$$\Big[\|\mu_i - \mu\|_0 < \epsilon_1,\ |H(\mu_i) - H(\mu)| < \epsilon_1\Big] \text{ or } \Big[\|\mu_i - \nu\|_0 < \epsilon_1,\ |H(\mu_i) - H(\nu)| < \epsilon_1\Big] \quad (\forall i > 0),$$

we have

$$H_{\prod_i \mu_i}(\lfloor \|Z_n\| \rfloor) \leq \epsilon n.$$

for each $n > N_1$.

This proposition is proved in [CFFT22] when the underlying space is Gromov hyperbolic, by means of pivoting technique. Using the pivotal time construction in [Cho22a] for metric spaces with contracting isometries, their proof applies to the setting as in Proposition 7.1.

We next describe a more complicated version of pivotal time construction in [Gou22, Section 5C]. Recall the notation $Z_n = g_1 \cdots g_n$ in 2.2. We also employ the notation $\mathbf{Y}_i := (Z_i o, \ldots, Z_{i+M_0} o)$.

{dfn:pivotalEquivLDP}

Definition 7.2 ([Cho22a, Definition 6.1]). *Let μ and ν be non-elementary probability measures on G and $(\Omega, \mathbb{P})$ be a probability space for μ. Let $0 < \epsilon < 1$, $K_0, N > 0$ and let $S \subseteq (\operatorname{supp} \mu)^{M_0}$ be a long enough K_0-Schottky set.*

*A subset $\mathcal{E}$ of Ω is called an (n, N, ϵ, ν)-*pivotal equivalence class *for μ, associated with the* set of pivotal times

$$\mathcal{P}^{(n,N,\epsilon,\nu)}(\mathcal{E}) = \{j(1) < j'(1) < \ldots < j(\#\mathcal{P}/2) < j'(\#\mathcal{P}/2)\} \subseteq M_0 \mathbb{Z}_{>0},$$

if the following hold:

(1) for each $\omega \in \mathcal{E}$ and $k \geq 1$,

$$s_k(\omega) := \big(g_{j(k)-M_0+1}(\omega),\ g_{j(k)-M_0+2}(\omega),\ \ldots,\ g_{j(k)}(\omega)\big),$$
$$s'_k(\omega) := \big(g_{j'(k)-M_0+1}(\omega),\ g_{j'(k)-M_0+2}(\omega),\ \ldots,\ g_{j'(k)}(\omega)\big)$$

are Schottky sequences;

(2) for each $\omega \in \mathcal{E}$, $\big(o, \mathbf{Y}_{j(1)}, \mathbf{Y}_{j'(1)}, \ldots, \mathbf{Y}_{j(\#\mathcal{P}/2)}, \mathbf{Y}_{j'(\#\mathcal{P}/2)}, Z_n o\big)$ is D_0-semi-aligned;

(3) for the RV defined as

$$r_k := g_{j(k)+1} g_{j(k)+2} \cdots g_{j'(k)-M_0},$$

$(s_k, s'_k, r_k)_{k>0}$ on $\mathcal{E}$ are i.i.d.s and r_k's are distributed almost according to $\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{M_0}-1-2N}$ in the sense that

$$(1-\epsilon)(\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{M_0}-1-2N})(g) \leq \mathbb{P}(r_k = g) \leq (1+\epsilon)(\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{M_0}-1-2N})(g).$$

for each $g \in G$.

{prop:gouezelRWLDP}

Proposition 7.3 ([Cho22a, Proposition 6.2]). *Let μ be a non-elementary probability measure on G, let $0 < \epsilon <$ and let $S \in G^{M_0}$ be a long enough Schottky set for μ with cardinality greater than $10/\epsilon$. Let $m := \min\{\mu^{M_0}(s) : s \in S\}$, let $N > 20/\epsilon m$ and let ν be a probability measure defined as*

$$\nu := \frac{1}{1-0.5m}\left(\mu^{*M_0} - 0.5m \cdot 1_{\{\Pi(s):s\in S\}}\right).$$

Then there exists $K > 0$ such that for each n we have a measurable partition $\mathscr{P}_{n,N,\epsilon,\nu} = \{\mathcal{E}_\alpha\}_\alpha$ of the ambient space $(G^{\mathbb{Z}_{>0}}, \mu'^{\mathbb{Z}_{>0}})$ into (n, N, ϵ, ν)-pivotal equivalence classes that satisfies

$$\mathbb{P}_{\mu'^{\mathbb{Z}_{>0}}}\left(\omega : \frac{1}{2}\#\mathcal{P}^{(n,N,\epsilon,\nu)}(\omega) \leq (1-\epsilon)\frac{n}{2M_0N}\right) \leq Ke^{-n/K}.$$

Before proving Theorem F, we state two lemmata that are well-known to the experts.

{lem:lengthSpec}

Lemma 7.4. *Let g be a contracting isometry of a metric space (X, d). Then for each N there exists K such that*

$$\#\{i > 0 : a \le \|g^i\| \le a + K\} < N$$

for every $a \in \mathbb{R}_{>0}$.

Proof. Let g be a K'-contracting isometry. Since $|d(o, g^i o) - d(o, g^j o)| \le d(g^i o, g^j o) \le |i - j|\|g\|$, the set $\{\|g^{ki}\| : i \in \mathbb{Z}\}$ is $k\|g\|$-coarsely dense in $\{\|g^i\| : i \in \mathbb{Z}\}$. Hence, it suffices to prove the statement for a power g^k of g. We take k large enough such that $\|g^k\| \ge 10K'$. Now, for each i, the contracting property of the set $\{o, go, \ldots, g^{k(i+1)}o\}$ forces that $[o, g^{k(i+1)}o]$ is $2K'$-close to $g^{ki}o$. Hence we have

$$\|g^{k(i+1)}\| = d(o, g^{k(i+1)}o) \ge d(o, g^{ki}o) + d(g^{ki}o, g^{k(i+1)}o) - 2K' \ge d(o, g^{ki}o) + 8K'.$$

Now the conclusion follows. ☐

Given an isometry $g \in G$ of a metric space (X, d), we define its *elementary closure*

$$E(g) := \{h \in G : d_{\mathrm{Hauss}}(\{hg^n o\}_{n\in\mathbb{Z}}, \{g^n o\}_{n\in\mathbb{Z}}) < +\infty\}.$$

{lem:eltClosure}

Lemma 7.5 ([Yan19, Lemma 2.11])**.** *If $g \in G$ is a WPD contracting isometry of X, then $\langle g \rangle$ is a finite-index subgroup of $E(g)$.*

When the group action is proper, this is proven in [Yan19, Lemma 2.11]. For completeness, we explain the argument for WPD actions.

Proof. The proof of [Yan19, Lemma 2.11] still guarantees the following: there exists a constant $C > 0$ such that $d_{\mathrm{Hauss}}(\{hg^n o\}_{n\in\mathbb{Z}}, \{g^n o\}_{n\in\mathbb{Z}}) < C$ for each $h \in E(g)$. We also pick a large enough N, that will be specified later on. We now consider an index-2 subgroup of $E(g)$:

$$E^+(g) := \{h \in E(g) : d_{\mathrm{Hauss}}(\{hg^n o\}_{n>0}, \{g^n o\}_{n>0}) < +\infty\}.$$

In other words, $h \in E^+(g)$ preserves the direction of the orbit of g. Since $h^2 \in E^+(g)$ for every $h \in E(g)$, it suffices to prove that $\langle g \rangle$ is a finite-index subgroup of $E^+(g)$.

For $h \in E^+(g)$, pick $i, j \in \mathbb{Z}$ such that $d(ho, g^i o) < C$ and $d(hg^N o, g^j o) < 0$. Since h preserves the direction of $\{g^n o\}_{n\in\mathbb{Z}}$, we have $i < j$. Then $|d(o, g^{j-i}o) - d(o, g^N o)| = |d(g^i o, g^j o) - d(ho, hg^N o)| < 2C$, and by Lemma 7.4, we have $N - K < j - i < N + K$ for some constant K that only depends on C. Then there exists a constant K' that depends on K such that

$$|d(hg^N o, g^{i+N} o)| < K'$$

holds for some $\epsilon \in \{1, -1\}$. In other words, hg^{-i} belongs to $\mathrm{Stab}_{K'}(o, g^N o)$, which is a finite set when N is sufficiently large. Hence, $h = g^i a$ belongs to a finite extension of $\langle g \rangle$. ☐

We now prove Theorem F.

Proof. Let g be a WPD contracting isometry of X, and let μ be a probability measure on G whose suitable power has nonzero weight on g. Let $\{\mu_i\}_i$ be a sequence of probability measures that converges simply to μ with $H(\mu_i) \to H(\mu)$, and let $\epsilon > 0$. We first note that $\limsup_{i\to\infty} H(\mu_i) \le H(\mu)$: this is the upper-semicontinuity of the asymptotic entropy that holds on any group. See [AAV13, Proposition 3.3] and the proof of [GMM18, Theorem 2.9]. Hence, the nontrivial part is to prove

$$\{\texttt{eqn:liminfEntropy}\}\liminf_{i\to\infty} H(\mu_i) \ge H(\mu). \tag{7.1}$$

We divide into elementary and non-elementary cases. First, consider the case that $\operatorname{supp}\mu$ is contained in the elementary closure $E(g)$ of g. Since $E(g)$ is a virtually cyclic, the μ-random walk is a random walk on an amenable group and its symptotic entropy must vanish. In this case, Inequality 7.1 is immediate.

In the case that there exists $h \in \operatorname{supp}\mu \setminus E(g)$, g^m and $hg^m h^{-1}$ are independent contracting isometries and μ is non-elementary. In particular, Fact 6.1 provides a constant K such that, given $\epsilon, M > 0$, there exists a long enough $10K/\epsilon$-Schottky set S for μ with cardinality $10/\epsilon$ such that $\Gamma(s)$ contains a translate of $(g^i o)_{i=0}^M$ for each $s \in S$.

Recall the choice of constants D_0, E_0 in Definition 2.7 that are determined by the value of K_0. Fixing an arbitrary $0 < \epsilon < 0.01$, we set $K_0 = 10K/\epsilon$ and take large enough M such that $\mathrm{Stab}_{12E_0+1}(o, g^M o)$ is finite. We then take a long enough K_0-Schottky set S for μ with cardinality $10/\epsilon$ such that $\Gamma(s)$ contains a translate of $(g^i o)_{i=0}^M$ for each $s \in S$. Let M_0 be such that $S \subseteq G^{M_0}$. From now on, except in the last paragraph, we discuss everything on the probability space $(G^{\mathbb{Z}}, \mu^{\mathbb{Z}})$.

Now fix a large integer N. By Proposition 7.3, there exist a non-elementary probability measure ν, and for each n a measurable partition $\mathscr{P}_{n,N,\epsilon,\nu} = \{\mathcal{E}_\alpha\}_\alpha$ of $G^{\mathbb{Z}_{>0}}$ into (n, N, ϵ, ν)-pivotal equivalence classes such that $\frac{1}{2}\#\mathcal{P}^{(n,N,\epsilon,\nu)}(\omega) > (1-\epsilon)n/2M_0N$ for asymptotic probability 1.

We now condition on a (n, N, ϵ, ν)-pivotal equivalence class $\mathcal{E} \in \mathscr{P}_{n,N,\epsilon,\nu}$. For convenience, we use notation $T := \frac{1}{2}\#\mathcal{P}^{(n,N,\epsilon,\nu)}(\mathcal{E})$. Note that $T \le n/2M_0N$ always. Our first claim is

{claim:entropyContinui

Claim 7.6. *For each $i = 1, \ldots, T$, we have*

$$H(r_i) \ge (1-\epsilon)H(\mu^{*2M_0N}) - 2\epsilon.$$

Proof of Claim 7.6. Let f_i be the distribution of $\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{M_0}-1-2N}$, i.e., $f_i(g) = \mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{M_0}-1-2N}(g)$. Note that the entropy of f_i is at least $H(\mu^{*2M_0N})$, because

$$H(Y_1Y_2) \ge \mathbb{E}[H(Y_1Y_2|Y_2)] \ge H(Y_1)$$

holds for every pair of RVs Y_1 and Y_2 on G.

Note that

$$\begin{aligned}
H(r_i) &= \sum_{g \in G} \mathbb{P}(r_i = g) \log \mathbb{P}(r_i = g) \\
&\geq -\sum_{g \in G} \mathbb{P}(r_i = g) \big| \log \mathbb{P}(r_i = g) - \log f_i(g) \big| - \big| \mathbb{P}(r_i = g) - f_i(g) \big| \log f_i(g) + \sum_{g \in G} f_i(g) \log f_i(g) \\
&\geq -\sum_{g \in G} \mathbb{P}(r_i = g) \cdot \log(1-\epsilon) - \sum_{g \in G} \epsilon f_i(g) \log f_i(g) + \sum_{g \in G} f_i(g) \log f_i(g) \\
&\geq (1-\epsilon) H(\mu^{*2M_0 N}) - 2\epsilon.
\end{aligned}$$

□

Let $\mathscr{F}$ be the partition of $\mathcal{E}$ based on the values of $\lfloor \|Z_{j'(1)}\| \rfloor, \ldots, \lfloor \|Z_{j'(T)}\| \rfloor$.

m:entropyContinuityC2}

Claim 7.7. *We have*

$$H(\mathscr{F}) \leq \epsilon n.$$

Proof of Claim 7.7. Recall that $Z_{j(1)-M_0}^{-1} Z_{j'(1)} = \Pi(s_1) \cdot r_1 \cdot \Pi(s_1')$, where s_1, s_1' are chosen from S. Hence, we have

$$\Big| \|Z_{j(1)-M_0}^{-1} Z_{j'(1)}\| - \|r_1\| \Big| \leq 2 \max_{s \in S} \|\Pi(s)\|.$$

Also, r_1 is almost distributed according to $\mu^{*2M_0 N} * \nu^{*\frac{j'(k)-j(k)}{M_0} - 1 - 2N}$. By Proposition 7.1 we have

$$H\left(\lfloor \|r_1\| \rfloor\right) \leq 0.25\epsilon (j'(1) - j(1) - M_0 - M_0 N),$$

provided that N is large enough. Since $\lfloor \|r_1\| \rfloor - \lfloor \|Z_{j(1)-M_0}^{-1} Z_{j'(1)}\| \rfloor$ is chosen among finitely many integers (between $-2\max_{s \in S} \|\Pi(s)\| - 2$ and $2\max_{s \in S} \|\Pi(s) + 2$), we have

$$H\left(\lfloor \|Z_{j(1)-M_0}^{-1} Z_{j'(1)}\| \rfloor \right) \leq 0.5\epsilon (j'(1) - j(1) - M_0 - M_0 N).$$

Similar estimate holds for $Z_{j(k)-M_0}^{-1} Z_{j'(k)}$ for $k = 2, \ldots, T$.

We now extract the information $(\lfloor \|Z_{j'(1)}\| \rfloor, \ldots, \lfloor \|Z_{j'(T)}\| \rfloor)$ from the RV $(\lfloor \|Z_{j(1)-M_0}^{-1} Z_{j'(1)}\| \rfloor, \ldots, \lfloor \|Z_{j(T)-M_0}^{-1} Z_{j'(T)}\| \rfloor)$. Since the pivotal loci are well aligned, we have

$$\|Z_{j'(k)}\| = \|Z_{j'(k-1)}\| + \|Z_{j'(k-1)}^{-1} Z_{j(k)-M_0}\| + \|Z_{j(k)-M_0}^{-1} Z_{j'(k)}\| + \alpha_k$$

for some real number $\alpha_k \in [0, 6E_0]$. Here, the term $\|Z_{j'(k-1)}^{-1} Z_{j(k)-M_0}\|$ is fixed on the entire $\mathcal{E}$. Hence, they do not contribute to the entropy. If we discretize the situation, we have

$$\lfloor \|Z_{j'(k)}\| \rfloor = \lfloor \|Z_{j'(k-1)}\| \rfloor + \lfloor \|Z_{j'(k-1)}^{-1} Z_{j(k)-M_0}\| \rfloor + \lfloor \|Z_{j(k)-M_0}^{-1} Z_{j'(k)}\| \rfloor + \alpha_k'$$

for some integer $\alpha'_k \in \{-6, -5, \ldots, 6E_0 + 6\}$. Hence, we have

$$\begin{aligned} H\big(\|Z'_{j'(1)}\|, \ldots, \|Z_{j'(T)}\|\big) &\leq H\left(\left(\lfloor \|Z^{-1}_{j(k)-M_0} Z_{j'(k)}\|\rfloor\right)_{k=1}^{T}, (\alpha'_k)_{k=1}^{T}\right) \\ &\leq \sum_{k=1}^{T} H\left(\lfloor \|Z^{-1}_{j(k)-M_0} Z_{j'(k)}\|\rfloor\right) + \sum_{k=1}^{T} H(\alpha'_k) \\ &\leq 0.5\epsilon n + T\log(6E_0 + 12) \leq \epsilon n. \quad \square \end{aligned}$$

Recall that s_i, s'_i are chosen from a finite set S for each i. Let $\mathscr{S}$ be the partition based on the choices of $s_1, s'_1, \ldots, s_T, s'_T$. Since $H(s_i) \leq \log \#S \leq N$ for large enough N, we have:

{claim:entropyContinui

Claim 7.8. . $H(\mathscr{S}) \leq \epsilon n$.

We then have

$$\{\texttt{eq:conditionedEntropyR}\} \quad H\Big((r_1, \ldots, r_T) \,\Big|\, \mathscr{F} \vee \mathscr{S}\Big) \geq (1-\epsilon)TH(\mu^{*2M_0N}) - 3\epsilon(T+n). \tag{7.2}$$

Indeed, using the Bayes' rule we have

$$\begin{aligned} H\Big((r_i)_i \,\Big|\, \mathscr{F} \vee \mathscr{S}\Big) &= H\Big(\mathscr{F} \vee \mathscr{S} \,\Big|\, (r_i)_i\Big) - H\Big(\mathscr{F} \vee \mathscr{S}\Big) + H((r_i)_i) \\ &\geq \sum_{i=1}^{T} H(r_i) - H(\{\mathscr{F}) - H(\mathscr{S}) \end{aligned}$$

and Inequality 7.2 follows from Claim 7.6, 7.7, and 7.8.

Note that

$$Z_{j'(k)} = h_0 \Pi(s_1) r_1 \Pi(s_1)' h_1 \cdots h_{k-1} \Pi(s_k) r_k \quad (h_0 := Z_{j(1)-M_0}, h_i := Z^{-1}_{j'(i)} Z_{j'(i+1)-M_0} \text{ for } i \geq 1)$$

Since we are discussing everything inside $\mathcal{E}$, h_i's are always fixed. Now, we claim:

{claim:entropyContinui

Claim 7.9. *Conditioned on each equivalence class E of $\mathscr{F} \vee \mathscr{S}$ (i.e., allowing the change of r_i's only), we have*

$$H(Z_n \,|\, E) \geq H\big((r_1, \ldots, r_T) \,\big|\, E\big) - AT$$

for some fixed constant A that does not depend on N.

Proof of Claim 7.9. We will show that

$$\#\{\omega \in E : Z_n(\omega) = a\} < B^T$$

holds for some constant B for every $a \in G$.

Let $g \in G$. If $\{\omega \in E : Z_n(\omega) = g\}$ contains at most 1 element, we are done. If not, suppose that $Z_n(\omega) = Z_n(\omega') = a$ for some $\omega, \omega' \in E$, i.e., with the same values of $\lfloor \|Z_{j'(k)}\|\rfloor$'s and s_k, s'_k's.

For each k, let $\sigma_k \in G$ be such that $\Gamma(s'_k)$ contains $\Pi(s_k)\sigma_k(o, go, \ldots, g^M o)$. We now claim that

$$\sigma_k^{-1} Z_{j'(k)}(\omega)^{-1} Z_{j'(k)}(\omega')\sigma_k \in \mathrm{Stab}_{10E_0}(o, g^M o).$$

To observe this, recall that $(o, Z_{j'(k)}(\omega)\sigma_k \Gamma(s'_k), Z_n(\omega))$ is D_0-semi-aligned and $[o, Z_n(\omega)o]$ possesses a subsegment that E_0-fellow travel with $Z_{j'(k)}(\omega)\sigma_k\Gamma(s'_k)$. Hence, $[o, a]$ contains points p_1, p_2, p_3, from left to right, such that

$$d(p_1, Z_{j'(k)}(\omega)\sigma_k o) < E_0, \quad d(p_2, Z_{j'(k)}(\omega)\sigma_k g^M o) < E_0, \quad d(p_3, Z_{j'(k)}(\omega')o) < E_0.$$

From a similar alignment we obtain points q_1, q_2, q_3 on $[o, a(\omega)]$ from left to right such that

$$d(q_1, Z_{j'(k)}(\omega')\sigma_k o) < E_0, \quad d(q_2, Z_{j'(k)}(\omega')\sigma_k g^M o) < E_0, \quad d(q_3, Z_{j'(k)}(\omega')o) < E_0.$$

Finally, thanks to the assumption on ω and ω', $\|Z_{j'(k)}\|$ and $\|Z_{j'(k)}(\omega')\|$ differ by at most 1. Hence, we have

$$\begin{aligned} d(p_3, q_3) &= |d(o, p_3) - d(o, q_3)| \le 2E_0 + 1, \\ d(p_2, q_2) &= |d(o, p_2) - d(o, q_2)| \\ &\le |d(o, p_3) - d(p_2, p_3) - d(o, q_3) + d(q_2, q_3)| \\ &\le 2E_0 + 1 + 4E_0 \le 6E_0 + 1, \end{aligned}$$

and for a similar reason we have $d(p_1, q_1) \le 10E_0 + 1$. This in turn implies

$$\begin{aligned} d(Z_{j'(k)}(\omega)\sigma_k o, Z_{j'(k)}(\omega')\sigma_k o) < 8E_0 + 1, \\ d(Z_{j'(k)}(\omega)\sigma_k g^M o, Z_{j'(k)}(\omega')\sigma_k g^M o) < 12E_0 + 1. \end{aligned}$$

Hence, $\sigma_k^{-1} Z_{j'(k)}(\omega)^{-1} Z_{j'(k)}(\omega')\sigma_k$ belongs to $\mathrm{Stab}_{12E_0+1}(o, g^M o)$.

To summarize, fixing an $\omega_0 \in E$, every $\omega \in E$ such that $Z_n(\omega_0) = \mathbb{Z}_n(\omega)$ satisfies that

$$\sigma_k^{-1} Z_{j'(k)}(\omega_0)^{-1} Z_{j'(k)}(\omega)\sigma_k \in \mathrm{Stab}_{12E_0+1}(o, g^M o) \quad (k = 1, \ldots, T).$$

We have set M such that $B := \# \mathrm{Stab}_{12E_0+1}(o, g^M o)$ is finite, and moreover, B is independent of N.

It remains to show that the choices of $\sigma_k^{-1} Z_{j'(k)}(\omega_0)^{-1} Z_{j'(k)}(\omega)\sigma_k$ determine the values of $r_k(\omega)$'s. We prove this by induction: we show that the information $\{\sigma_i^{-1} Z_{j'(i)}(\omega_0)^{-1} Z_{j'(i)}(\omega)\sigma_i\}_{i=1}^k$ determines $r_1(\omega), \ldots, r_i(\omega)$ and $Z_{j(i+1)}(\omega)$. For $k = 1$, we have

$$\begin{aligned} r_1(\omega) &= Z_{j(1)}^{-1}(\omega) \cdot Z_{j'(1)}(\omega_0)\sigma_1 \cdot (\sigma_1^{-1} Z_{j'(1)}(\omega_0)^{-1} Z_{j'(1)}(\omega)\sigma_1) \cdot \sigma_1^{-1}\Pi(s'_1)^{-1}, \\ Z_{j'(1)}(\omega) &= Z_{j'(1)}(\omega_0)\sigma_1 \cdot (\sigma_1^{-1} Z_{j'(1)}(\omega_0)^{-1} Z_{j'(1)}(\omega)\sigma_1) \cdot \sigma_1^{-1}. \end{aligned}$$

Since s'_1, σ_1 and $Z_{j(1)}$ are fixed on E, the above equality determines $r_1(\omega)$ and $Z_{j'(1)}(\omega)$ whenever $\sigma_1^{-1} Z_{j'(1)}(\omega_0)^{-1} Z_{j'(1)}(\omega)\sigma_1$ is given. Moreover, since $Z_{j'(1)}(\omega)^{-1} Z_{j(2)}(\omega)$ is also known (fixed on E), we obtain $Z_{j(2)}(\omega)$ from this.

Now assume that the induction works till step $k-1$. We have

$$r_k(\omega) = Z_{j(k)}^{-1}(\omega) \cdot Z_{j'(k)}(\omega_0)\sigma_k \cdot (\sigma_k^{-1} Z_{j'(k)}(\omega_0)^{-1} Z_{j'(k)}(\omega)\sigma_k) \cdot \sigma_k^{-1}\Pi(s'_k)^{-1}.$$

Since s'_k, σ_k are fixed on E and $Z_{j(k)}(\omega)$ is obtained from the induction hypothesis, we can determine $r_k(\omega)$. Using this we can also determine $Z_{j(k+1)}(\omega) = Z_{j(k)}(\omega) r_k(\omega) \cdot \Pi(s_k) \cdot (Z_{j'(k)}^{-1}(\omega) Z_{j(k+1)}(\omega))$, finishing the induction. □

We now combine the results. By Inequality 7.2 and Claim 7.9, we have

$$H(Z_n|E) \geq (1-\epsilon)TH(\mu^{*2M_0N}) - 3\epsilon(T+n) - AT$$

for each $E \in \mathscr{F} \vee \mathscr{S}$, which is a partition of a pivotal equivalence class $\mathcal{E}$. Summing up these conditional entropies, we deduce

$$H_{\mu^n}(Z_n|\mathcal{E}) \geq \sum_{E \in \mathscr{F} \vee \mathscr{S}} H(Z_n|E)\,\mathbb{P}(E|\mathcal{E}) \geq (1-\epsilon)TH(\mu^{*2M_0N}) - 3\epsilon(T+n) - AT.$$

We now take large enough N such that:

(1) $40/N\epsilon$ is smaller than $\mu^{M_0}(s)$ for every $s \in S$;
(2) $T \leq n/2M_0N \leq \frac{\epsilon}{1+A}N$, and
(3) $H(\mu^{*2M_0N}) \geq (1-\epsilon) \cdot 2M_0Nh(\mu)$.

Then we have

$$\begin{aligned} H_{\mu^n}(Z_n) &\geq \sum_{\mathcal{E} \in \mathscr{P}^{(n,N,\epsilon,\nu)}} H_{\mu_i^n}(Z_n|\mathcal{E})\mu_i^n(\mathcal{E}) \\ &\geq (1-\epsilon)H(\mu^{*2M_0N}) \times \mathbb{E}_{\mu_i^n}[T] - 10\epsilon n. \end{aligned}$$

Since $\mathbb{P}_{\mu_i^{\mathbb{Z}_{>0}}}(T = \frac{1}{2}\#\mathcal{P}^{(n,N,\epsilon,\nu)} > (1-\epsilon)n/2M_0N) \to 1$ as $n \to +\infty$, we conclude that

$$h(\mu) = \lim_{n \to +\infty} \frac{1}{n} H_{\mu^n}(Z_n) \geq (1-\epsilon)^2 \frac{1}{2M_0N} H(\mu^{*2M_0N}) - 10\epsilon.$$

We now claim that fixing $\epsilon > 0$ and the corresponding long enough Schottky set S, all the arguments so far works for all but finitely many μ_i's with the same N. First recall that ν is the non-elementary measure for μ as in Proposition 7.3, i.e.,

$$\nu := \frac{1}{1 - 0.5\min\{\mu^{M_0}(s) : s \in S\}} \left(\mu^{*M_0} - 0.5\min\{\mu^{M_0}(s) : s \in S\} \cdot 1_{\{\Pi(s) : s \in S\}}\right).$$

We similarly define

$$\nu_i := \frac{1}{1 - 0.5\min\{\mu_i^{M_0}(s) : s \in S\}} \left(\mu^{*M_0} - 0.5\min\{\mu_i^{M_0}(s) : s \in S\} \cdot 1_{\{\Pi(s) : s \in S\}}\right).$$

Let ϵ_1, N_1 be as in Proposition 7.1. Since $\mu_i \to \mu$ and $H(\mu_i) \to H(\mu)$, we have the following for all large i:

(1) $\|\mu_i - \mu\|_0 < \epsilon_1$ and $|H(\mu_i) - H(\mu)| < \epsilon_1$;
(2) $\|\nu_i - \nu\|_0 < \epsilon_1$ and $|H(\nu_i) - H(\nu)| < \epsilon_1$;
(3) $\min\{\mu_i^{M_0}(s) : s \in S\} > 0.5\min\{\mu^{M_0}(s) : s \in S\}$.

Then for these large i's and for $N > N_1 + 40/\epsilon \min\{\mu^{M_0}(s) : s \in S\}$, all the arguments so far is valid for μ_i put in place of μ.

Hence, for each sufficiently large i we have

$$h(\mu_i) \geq (1-\epsilon)^2 \frac{1}{2M_0N} H(\mu_i^{*2M_0N}) - 10\epsilon.$$

Since $H(\mu_i^{*2M_0N})$ converges to $H(\mu^{*2M_0N})$, we have

$$\liminf_i h(\mu_i) \geq (1-\epsilon)^2 \frac{1}{2M_0N} H(\mu^{*2M_0N}) - 10\epsilon.$$

By taking $N \to \infty$, we then have

$$\liminf_i h(\mu_i) \geq (1-\epsilon)^2 h(\mu) - 10\epsilon.$$

By sending $\epsilon \to 0$, we conclude that $\limsup_i h(\mu_i) \geq h(\mu)$. $\square$

References

[AAV13] Gideon Amir, Omer Angel, and Bálint Virág. Amenability of linear-activity automaton groups. *J. Eur. Math. Soc. (JEMS)*, 15(3):705–730, 2013.

[AKKP19] Yael Algom-Kfir, Ilya Kapovich, and Catherine Pfaff. Stable strata of geodesics in outer space. *Int. Math. Res. Not. IMRN*, (14):4549–4578, 2019.

[Ave72] André Avez. Entropie des groupes de type fini. *C. R. Acad. Sci. Paris Sér. A-B*, 275:A1363–A1366, 1972.

[BC12] Jason Behrstock and Ruth Charney. Divergence and quasimorphisms of right-angled Artin groups. *Math. Ann.*, 352(2):339–356, 2012.

[BF09] Mladen Bestvina and Koji Fujiwara. A characterization of higher rank symmetric spaces via bounded cohomology. *Geom. Funct. Anal.*, 19(1):11–40, 2009.

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

[BHM08] Sébastien Blachère, Peter Haïssinsky, and Pierre Mathieu. Asymptotic entropy and Green speed for random walks on countable groups. *Ann. Probab.*, 36(3):1134–1152, 2008.

[BHM11] Sébastien Blachère, Peter Haïssinsky, and Pierre Mathieu. Harmonic measures versus quasiconformal measures for hyperbolic groups. *Ann. Sci. Éc. Norm. Supér. (4)*, 44(4):683–721, 2011.

[BMM10] Jeffrey Brock, Howard Masur, and Yair Minsky. Asymptotics of Weil-Petersson geodesic. I. Ending laminations, recurrence, and flows. *Geom. Funct. Anal.*, 19(5):1229–1257, 2010.

[CF10] Pierre-Emmanuel Caprace and Koji Fujiwara. Rank-one isometries of buildings and quasi-morphisms of Kac-Moody groups. *Geom. Funct. Anal.*, 19(5):1296–1319, 2010.

[CFFT22] Kunal Chawla, Behrang Forghani, Joshua Frisch, and Giulio Tiozzo. The poisson boundary of hyperbolic groups without moment condition. *arXiv preprint arXiv:2209.02114*, 2022.

[CFI16] Indira Chatterji, Talia Fernós, and Alessandra Iozzi. The median class and superrigidity of actions on CAT(0) cube complexes. *J. Topol.*, 9(2):349–400, 2016. With an appendix by Pierre-Emmanuel Caprace.

[Cho22a] Inhyeok Choi. Random walks and contracting elements I: Deviation inequality and limit laws. *arXiv preprint arXiv:2207.06597v2*, 2022.

[Cho22b] Inhyeok Choi. Random walks and contracting elements III: Outer space and outer automorphism group. *arXiv preprint arXiv:2212.12122*, 2022.

[CS11] Pierre-Emmanuel Caprace and Michah Sageev. Rank rigidity for CAT(0) cube complexes. *Geom. Funct. Anal.*, 21(4):851–891, 2011.

[Der80] Yves Derriennic. Quelques applications du théorème ergodique sous-additif. In *Conference on Random Walks (Kleebach, 1979) (French)*, volume 74 of *Astérisque*, pages 183–201, 4. Soc. Math. France, Paris, 1980.

[DG21] Matthieu Dussaule and Ilya Gekhtman. Stability phenomena for Martin boundaries of relatively hyperbolic groups. *Probab. Theory Related Fields*, 179(1-2):201–259, 2021.

[DT18] Spencer Dowdall and Samuel J. Taylor. Hyperbolic extensions of free groups. *Geom. Topol.*, 22(1):517–570, 2018.

[FM11] Stefano Francaviglia and Armando Martino. Metric properties of outer space. *Publ. Mat.*, 55(2):433–473, 2011.

[Gil07] Lorenz A. Gilch. Rate of escape of random walks on free products. *J. Aust. Math. Soc.*, 83(1):31–54, 2007.

[Gil11] Lorenz A. Gilch. Asymptotic entropy of random walks on free products. *Electron. J. Probab.*, 16:no. 3, 76–105, 2011.

[GMM18] Sébastien Gouëzel, Frédéric Mathéus, and François Maucourant. Entropy and drift in word hyperbolic groups. *Invent. Math.*, 211(3):1201–1255, 2018.

[Gou17] Sébastien Gouëzel. Analyticity of the entropy and the escape rate of random walks in hyperbolic groups. *Discrete Anal.*, (7), 2017.

[Gou22] Sébastien Gouëzel. Exponential bounds for random walks on hyperbolic spaces without moment conditions. *Tunis. J. Math.*, 4(4):635–671, 2022.

[Gui80] Y. Guivarc'h. Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire. In *Conference on Random Walks (Kleebach, 1979) (French)*, volume 74 of *Astérisque*, pages 47–98, 3. Soc. Math. France, Paris, 1980.

[Ham09] Ursula Hamenstädt. Rank-one isometries of proper CAT(0)-spaces. In *Discrete groups and geometric structures*, volume 501 of *Contemp. Math.*, pages 43–59. Amer. Math. Soc., Providence, RI, 2009.

[HMM18] Peter Haïssinsky, Pierre Mathieu, and Sebastian Müller. Renewal theory for random walks on surface groups. *Ergodic Theory Dynam. Systems*, 38(1):155–179, 2018.

[KÈ13] V. A. Kaĭmanovich and A. G. Èrshler. Continuity of asymptotic characteristics for random walks on hyperbolic groups. *Funktsional. Anal. i Prilozhen.*, 47(2):84–89, 2013.

[KL07] Anders Karlsson and François Ledrappier. Linear drift and Poisson boundary for random walks. *Pure Appl. Math. Q.*, 3(4, Special Issue: In honor of Grigory Margulis. Part 1):1027–1036, 2007.

[KMPT22a] Ilya Kapovich, Joseph Maher, Catherine Pfaff, and Samuel J. Taylor. Random outer automorphisms of free groups: attracting trees and their singularity structures. *Trans. Amer. Math. Soc.*, 375(1):525–557, 2022.

[KMPT22b] Ilya Kapovich, Joseph Maher, Catherine Pfaff, and Samuel J. Taylor. Random trees in the boundary of outer space. *Geom. Topol.*, 26(1):127–162, 2022.

[KV83] Vadim A. Kaimanovich and Anatoly M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

[Led01] François Ledrappier. Some asymptotic properties of random walks on free groups. In *Topics in probability and Lie groups: boundary theory*, volume 28 of *CRM Proc. Lecture Notes*, pages 117–152. Amer. Math. Soc., Providence, RI, 2001.

[Led13] François Ledrappier. Regularity of the entropy for random walks on hyperbolic groups. *Ann. Probab.*, 41(5):3582–3605, 2013.

[LL12] François Ledrappier and François Ledrappier. Analyticity of the entropy for some random walks. *Groups Geom. Dyn.*, 6(2):317–333, 2012.

[Mas80] Howard Masur. Uniquely ergodic quadratic differentials. *Comment. Math. Helv.*, 55(2):255–266, 1980.

[Mas21] Hidetoshi Masai. On continuity of drifts of the mapping class group. *Math. Res. Lett.*, 28(3):823–847, 2021.

[Mat15] P. Mathieu. Differentiating the entropy of random walks on hyperbolic groups. *Ann. Probab.*, 43(1):166–187, 2015.

[Min96] Yair N. Minsky. Quasi-projections in Teichmüller space. *J. Reine Angew. Math.*, 473:121–136, 1996.

[MP16] Lee Mosher and Catherine Pfaff. Lone axes in outer space. *Algebr. Geom. Topol.*, 16(6):3385–3418, 2016.

[MS20] Pierre Mathieu and Alessandro Sisto. Deviation inequalities for random walks. *Duke Math. J.*, 169(5):961–1036, 2020.

[Sam02] Andrea Sambusetti. Growth tightness of surface groups. *Expo. Math.*, 20(4):345–363, 2002.

[Vog15] Karen Vogtmann. On the geometry of outer space. *Bull. Amer. Math. Soc. (N.S.)*, 52(1):27–46, 2015.

[Yan19] Wen-yuan Yang. Statistically convex-cocompact actions of groups with contracting elements. *Int. Math. Res. Not. IMRN*, (23):7259–7323, 2019.

June E Huh Center for Mathematical Challenges, KIAS, 85 Hoegiro Dongdaemun-gu, Seoul, 02455, South Korea

Email address: inhyeokchoi48@gmail.com