

# 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  $\text{CAT}(0)$  spaces. We discuss continuity and differentiability of the escape rate of random walks on  $G$ . For relatively hyperbolic groups,  $\text{CAT}(-1)$  groups and  $\text{CAT}(0)$  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, hyperbolic group,  $\text{CAT}(0)$  space, mapping class group

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

## 1. INTRODUCTION

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

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

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  $\mu$  on  $G$  has strictly positive escape rate  $l(\mu) > 0$  [KV83]. Moreover, for a probability measure  $\mu$  on  $G$  with finite one-time entropy, its asymptotic entropy  $H(\mu)$  vanishes if and only if the Poisson boundary for  $(G, \mu)$  is trivial, i.e., there is no non-constant  $\mu$ -harmonic bounded function on  $G$  (see [Ave72], [Der80] and [KV83]; see also [KL07] for the role of  $L(\mu)$  in this problem).

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]) reads

$$h \leq lv,$$

and for word hyperbolic groups, the equality holds if and only if the limiting distributions for the  $\mu$ -random walk (the  $\mu$ -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  $\mu$ . 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 measures  $\mu$  supported in  $S$ . One ingredient of their argument is the continuity of  $h(\mu)$  and  $l(\mu)$  with respect to the underlying measure  $\mu$ .

In this paper, we focus on group actions that involve contracting isometries (see Convention 2.4). We say that a probability measure  $\mu$  on  $G$  is *non-elementary* if the semigroup generated by the support of  $\mu$  contains two independent contracting isometries of  $X$ . We say that a sequence of measures  $\{\mu_i\}_{i>0}$  on  $G$  *converges simply* to  $\mu$  if  $\mu_i(g) \rightarrow \mu(g)$  for each  $g \in G$ . We recall the continuity of the escape rate:

**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  $\mu$  be a non-elementary probability measure on  $G$ , and let  $\{\mu_i\}_{i>0}$  be a sequence of probability measures that converges simply to  $\mu$  with  $L(\mu_i) \rightarrow L(\mu)$ . 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 for the case of Teichmüller space. Gouëzel reproved this result for non-elementary isometry groups of Gromov hyperbolic spaces and his argument applies to group actions involving contracting isometries (see [Cho22a, Section 3.2, 3.3] for the necessary modification, where the alignment lemmata are presented).

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  $\eta$  on  $G$ , we define

$$\begin{aligned} \|\eta\|_0 &:= \sum_{g \in G} |\eta(g)| \quad (\text{total variation}), \\ \|\eta\|_1 &:= \sum_{g \in G} d(o, go) |\eta(g)| \quad (\text{first moment}), \\ \|\eta\|_{0,1} &:= \|\eta\|_0 + \|\eta\|_1 = \sum_{g \in G} (d(o, go) + 1) |\eta(g)|. \end{aligned}$$

Our main results are as follows.

**Theorem A.** *Let  $G$  be as in Convention 2.4 and let  $\mu$  be a non-elementary probability measure on  $G$  with finite first moment. Then there exists  $C, \epsilon > 0$  such that*

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

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

**Theorem B.** *Let  $G$  be as in Convention 2.4 and let  $\mu$  be a non-elementary probability measure on  $G$  with finite first moment, let  $\eta$  be a signed measure that is absolutely continuous with respect to  $\mu$  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  $\mu$  and  $\eta$  in  $\|\cdot\|_{0,1}$ -topology.*

Analyticity of the escape rate of random walks on free groups and free products was discussed independently by Ledrappier and Gilch (see [Led01], [Gil07], [Led12]). 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.

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

- *Teichmüller space (with the Teichmüller metric or the Weil-Petersson metric) and the mapping class group;*
- *Culler-Vogtmann Outer space  $CV_N$  and  $\text{Out}(F_N)$ ;*
- *a CAT(-1) space and a countable group of its isometries;*
- *a CAT(0) cube complex and a countable group of its isometries;*
- *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  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite first moment, let  $\eta$  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*

$$\sigma_1(\mu, \eta) := \mathbb{E}_\eta d(o, go) - \lim_{n \rightarrow \infty} \mathbb{E}_{\mu^{*n} \times \eta \times \mu^{*n}} [d(o, g_1 o) + d(o, g_2 o) + d(o, g_3 o) - d(o, g_1 g_2 g_3 o)].$$

*Moreover, this derivative is continuous with respect to  $\mu$  and  $\eta$  in  $\|\cdot\|_{0,1}$ -topology.*

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  $C^\infty$ -ness. For simplicity, we here present a result about  $C^2$ -regularity.

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

- *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  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite first moment and let  $\eta, \eta'$  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

$$\begin{aligned} \sigma_2(\mu, \eta, \eta') &:= \mathbb{E}_{\eta'} d(o, go) - \lim_{n \rightarrow \infty} \mathbb{E}_{\mu^{*n} \times \eta' * \mu^{*n}} [d(o, g_1 o) + d(o, g_2 o) + d(o, g_3 o) - d(o, g_1 g_2 g_3 o)] \\ &\quad - \sum_{i=1}^{\infty} \lim_{n \rightarrow \infty} \mathbb{E}_{\mu^{*n} \times \eta \times (\mu^{*(i-1)} * \eta * \mu^{*(n-i)})} [d(o, g_1 o) + d(o, g_2 o) + d(o, g_3 o) - d(o, g_1 g_2 g_3 o)] \\ &\quad - \sum_{i=1}^{\infty} \lim_{n \rightarrow \infty} \mathbb{E}_{(\mu^{*(n-i)} * \eta * \mu^{*(i-1)}) \times \eta \times \mu^{*n}} [d(o, g_1 o) + d(o, g_2 o) + d(o, g_3 o) - d(o, g_1 g_2 g_3 o)]. \end{aligned}$$

Moreover, this derivative is continuous with respect to  $\mu$ ,  $\eta$  and  $\eta'$  in  $\|\cdot\|_{0,1}$ -topology.

As mentioned, analyticity of the escape rate on Gromov hyperbolic groups is due to Gou  zel [Gou17]. Gou  zel'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  $\pi : \partial_B G \rightarrow \partial G$  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  $\pi$ , i.e.,  $\pi^{-1}(g^+)$  consists of a single horofunction. This holds on the standard Cayley graph of a surface group, for example, but is not true in general. For example, on a ‘‘ladder’’  $G = \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ , the two ends of the group correspond to two boundary points on  $\partial G$  and each of them has two preimages in  $\partial_B G$ .

Another difference between Gou  zel's method and ours is the condition on the underlying measure  $\mu$ . Gou  zel's method investigates spectral property of the Markov operator which requires  $\mu$  to be admissible (i.e., the support of  $\mu$  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).

We next turn to the continuity of the asymptotic entropy of  $\mu$ . 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.

**Theorem E.** *Let  $(X, G)$  be as in Convention 2.4 with a WPD contracting element  $g \in G$ , let  $\mu$  be a probability measure with finite time-one entropy satisfying  $\mu^{*n}(g^m) > 0$  for some  $n, m > 0$ , and let  $\{\mu_i\}_{i>0}$  be a sequence of probability measures that converges simply to  $\mu$  with  $H(\mu_i) \rightarrow H(\mu)$ . Then  $h(\mu_i)$  converges to  $h(\mu)$ .*

This continuity of the asymptotic entropy on hyperbolic groups was proven by Erschler and Kaimanovich [K  13] when the underlying measure  $\mu$  has finite first moment, and later by Gou  zel, Math  us and Maucourant when  $\mu$  has finite logarithmic moment [GMM18]. See [Mas21] for the case of Teichm  ller space. While preparing the current manuscript, the author was informed by Anna Erschler and Joshua Frisch of their independent proof of Theorem E.

For higher regularity (differentiability, analyticity, etc.) of the asymptotic entropy on hyperbolic groups, see [Gil11], [Led12], [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 E by combining Gou  zel’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 Teichm  ller space, CAT(0) spaces, CAT(0) cube complexes, CAT(-1) spaces, Culler-Vogtmann Outer space and (relatively) hyperbolic groups. In Section 3, we summarize and explain the results from pivoting technique. In section 4, we establish a single probabilistic estimate (Lemma 4.1) that leads to Theorem A and B. Once Lemma 4.1, 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 5, similarly, we single out a probabilistic estimate (Proposition 5.2) that leads to Theorem C. In Section 6, a slightly more delicate estimate (Proposition 6.2) is stated and Theorem D is deduced. 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 E.

**Acknowledgments.** The author thanks Kunal Chawla, Giulio Tiozzo and Joshua Frisch for explaining the usage of pivotal times for estimating asymptotic entropy. The author also thanks Ilya Gekhtman and Hidetoshi Masai for fruitful discussion, and the anonymous referee for suggesting many valuable comments that helped improved the exposition of this paper. 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), Mid-Career Researcher Program (RS-2023-00278510) through the National Research Foundation funded by the government of Korea, and by a KIAS Individual Grant (SG091901) via the June E Huh Center for Mathematical Challenges at Korea Institute for Advanced Study.

## 2. PRELIMINARIES

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

**2.1. Contracting isometries and BGIP isometries.** Inspired from the previous work by many authors, including [ACT15], [Sis18], [Yan19], [KMPT22a], we focus on contracting isometries of metric spaces. The following definition is meant to incorporate Culler-Vogtmann Outer space with an asymmetric metric.

**Definition 2.1** (Metric space). *An (asymmetric) metric space  $(X, d)$  is a set  $X$  equipped with a function  $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$  that satisfies the following:*

- (non-degeneracy) for each  $x, y \in X$ ,  $d(x, y) = 0$  if and only if  $x = y$ ;
- (triangle inequality) for each  $x, y, z \in X$ ,  $d(x, z) \leq d(x, y) + d(y, z)$ ;
- (local symmetry) for each  $x \in X$ , there exist  $\epsilon, K > 0$  such that  $d(y, z) \leq Kd(z, y)$  holds for  $y, z \in \{a \in X : \min(d(x, a), d(a, x)) < \epsilon\}$ .

*In this situation, we say that  $d$  is a metric on  $X$ .  $d$  is said to be symmetric if  $d(x, y) = d(y, x)$  holds for all  $x, y \in X$ . We define a symmetric metric called the symmetrization of  $d$  by*

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

*We endow  $(X, d)$  with the topology induced by  $d^{sym}$ .*

A *path* is a map from an interval or a set of consecutive integers to the ambient space  $X$ . We say that two paths  $\gamma$  and  $\eta$  are *K-fellow traveling* if there is an orientation-preserving reparametrization  $\gamma'$  of  $\gamma$  such that  $d^{sym}(\gamma'(\tau), \eta(\tau)) < K$  for every input  $\tau$ .

We define the Gromov product between  $x$  and  $y$  based at  $z$  by

$$(x, y)_z := \frac{1}{2}[d(x, z) + d(z, y) - d(x, y)].$$

When a path  $\gamma : I \rightarrow X$  is understood, we denote by  $\bar{\gamma}$  its reversal, i.e., the same path with orientation reversed. So the beginning point of  $\gamma$  becomes the ending point of  $\bar{\gamma}$  and vice versa.

**Definition 2.2** (Quasigeodesics). *A path  $\gamma : I \rightarrow X$  from an interval or a set of consecutive integers  $I$  is called a K-quasigeodesic if*

$$(2.1) \quad \frac{1}{K}|t - s| - K \leq d(\gamma(s), \gamma(t)) \leq K|t - s| + K$$

*holds for all  $s, t \in I$  such that  $s < t$ . If Inequality 2.1 holds for all  $s, t \in I$ , we say that  $\gamma$  is a K-bi-quasigeodesic. A path is called a geodesic if  $d(\gamma(s), \gamma(t)) = t - s$  holds for all  $s, t \in I$  such that  $s < t$ .*

*A metric space  $X$  is said to be geodesic if every ordered pair of points can be connected by a geodesic, i.e., for every  $x, y \in X$  there exists a geodesic  $\gamma : [a, b] \rightarrow X$  such that  $\gamma(a) = x$  and  $\gamma(b) = y$ . We denote  $\gamma$  by  $[x, y]$ .*

**Remark 2.3.** *In general asymmetric metric spaces, the reversal of a geodesic may not be a geodesic. For an illuminating example due to Coulbois and Wiest, see [FM11, Section 6].*

For a closed set  $A \subseteq X$ , we denote by  $\pi_A(\cdot)$  the closest point projection onto  $A$ . A K-quasigeodesic  $\gamma$  is called a *K-contracting axis* if

$$\text{diam}(\pi_\gamma(B)) < K$$

holds for every metric ball  $B$  disjoint from the  $K$ -neighborhood of  $\gamma$ .

In metric spaces with a symmetric metric, contracting property is known to be equivalent to the *bounded geodesic image property (BGIP)*: a K-bi-quasigeodesic  $\gamma$  is called a *K-BGIP axis* if

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

for any geodesic  $\eta$  that is disjoint from the  $K$ -neighborhood of  $\gamma$ . (Some authors call this equivalent notion as contracting property.) In asymmetric metric spaces, these two properties are not equivalent in general. We stick to the language of BGIP axes, but they really mean contracting axes when the underlying metric is symmetric.

Fix a basepoint  $o$  of  $X$ . An isometry  $g$  of  $X$  is said to have *K-BGIP* if its orbit  $\{g^i o\}_{i \in \mathbb{Z}}$  is a K-BGIP axis. Two BGIP isometries  $g, h \in \text{Isom}(X)$  are said to be *independent* if their orbits have bounded projections onto each other. A group  $G \leq \text{Isom}(X)$  is said to be *non-elementary* if it contains two independent BGIP isometries. **From now on, unless otherwise stated, we adopt:**

**Convention 2.4.** *We fix a space  $X$  with possibly asymmetric metric  $d$  and with a basepoint  $o$ .  $G$  is a non-elementary countable isometry group of  $X$ , i.e.,  $G$  contains two independent BGIP isometries. When  $X$  is assumed to be geodesic and  $x, y \in X$ ,  $[x, y]$  denotes an arbitrary geodesic connecting  $x$  to  $y$ .*

For genuine metric spaces with symmetric metrics, one can replace the above convention with:

**Convention 2.5.** *We fix a metric space  $(X, d)$  (where  $d$  is symmetric) 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  $X$ , i.e.,  $G$  contains two independent contracting isometries.*

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  $\|\cdot\|$  [AKB12, Theorem 24].

We say that a BGIP isometry  $g$  of  $X$  has *WPD (weak proper discontinuity)* if for each  $K \geq 0$  there exists  $N > 0$  such that

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

is finite [BF02, Section 3]. For example, when the  $G$ -action on  $X$  is properly discontinuous, every BGIP isometry of  $X$  in  $G$  has WPD.

**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  $G^n$  through the composition map, we obtain the convolution measure  $\mu_1 * \cdots * \mu_n$  on  $G$ . When  $\mu_1 = \cdots = \mu_n = \mu$ , we write by  $\mu^n$  and  $\mu^{*n}$  the  $n$ -product measure and the  $n$ -convolution measure of  $\mu$ , respectively.

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

**Notation 2.6.** 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  $(g_i)_{i \in \mathbb{Z}}$ , the random variable  $Z_n(g_1, \dots, g_n)$  models the  $n$ -th step position of the random walk starting from the identity. We can write the escape rate of  $\mu$  as

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

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

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

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

Let us now fix the metric space  $X$  on which  $G$  acts by isometries. A probability measure  $\mu$  on  $G$  is *non-elementary* if the semigroup  $\langle\langle \text{supp } \mu \rangle\rangle$  generated by the support of  $\mu$  contains two independent BGIP isometries. We will say that a non-elementary probability measure  $\mu$  is *strongly non-elementary* in the following cases:

- $G$  is the mapping class group and  $X$  is Teichmüller space with either metric. In this case, since  $\mu$  is assumed to be non-elementary,  $\langle\langle \text{supp } \mu \rangle\rangle$  contains a pseudo-Anosov mapping class.
- $G$  is the outer automorphism group  $\text{Out}(F_N)$  of a free group,  $X$  is the Culler-Vogtmann Outer space and  $\langle\langle \text{supp } \mu \rangle\rangle$  contains a principal fully irreducible outer automorphism.
- $G$  is a countable automorphism group of a CAT(0) cube complex and  $\langle\langle \text{supp } \mu \rangle\rangle$  contains an automorphism that skewers a pair of strongly separated hyperplanes (see Fact 2.27).
- $G$  is a countable isometry group of a CAT(-1) space or a relatively hyperbolic group and  $\langle\langle \text{supp } \mu \rangle\rangle$  contains a loxodromic element.
- $G = \langle a_1, \dots, a_g, b_1, \dots, b_g \mid \prod_{i=1}^g [a_i, b_i] \rangle$  is the surface group and  $\langle\langle \text{supp } \mu \rangle\rangle$  contains  $a_1^{2g} b_2^{2g}$ .
- $G = F_1 * F_2 * \cdots * F_N$  is a free product of nontrivial groups and  $\langle\langle \text{supp } \mu \rangle\rangle$  contains  $w_1 w_2$  for some  $w_i \in F_i \notin \{id\}$  for  $i = 1, 2$ .

Later, we will evaluate some RVs on  $G^{\mathbb{Z}}$  with respect to several different measures. We introduce a notation for this purpose.

**Notation 2.7.** Let  $\{\mu_i\}_{i \in \mathbb{Z}}$  be signed measures on  $G$ , all but finitely many of which is probability measures. Let  $F : G^{\mathbb{Z}} \rightarrow \mathbb{R}$  be measurable function. When  $g_i$  are independently distributed according to  $\mu_i$  ( $i \in \mathbb{Z}$ ), we denote the expectation of  $F(\mathbf{g})$  by

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

and 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.

**2.3. Alignment lemma.** We fix a geodesic metric space  $X$ . We summarize the content of Subsection 3.2 and 3.3 of [Cho22a]. We refer the readers to [Cho22a] for the proofs.

**Definition 2.8** ([Cho22a, Definition 3.5]). For  $i = 1, \dots, n$ , let  $\gamma_i$  be a path on  $X$  whose beginning and ending points are  $x_i$  and  $y_i$ , respectively. We say that  $(\gamma_1, \dots, \gamma_n)$  is  $C$ -aligned if

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

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

The following clear:

- (1) If  $(\gamma_1, \dots, \gamma_N)$  and  $(\gamma_N, \dots, \gamma_M)$  are each  $C$ -aligned, then so is  $(\gamma_1, \dots, \gamma_N, \dots, \gamma_M)$ .
- (2) If  $(\gamma_1, \dots, \gamma_n)$  is  $C$ -aligned, so is  $(\bar{\gamma}_N, \dots, \bar{\gamma}_1)$ .

**Lemma 2.9** ([Cho22a, Lemma 3.7], [Cho22c, Lemma 2.18]). For each  $C > 0$  and  $K > 1$ , there exists  $D = D(C, K) > \max(C, K)$  that satisfies the following.

Let  $\gamma, \gamma'$  be  $K$ -BGIP axes whose beginning points are  $x$  and  $x'$ , respectively. If  $(\gamma, x')$  and  $(x, \gamma')$  are  $C$ -aligned, then  $(\gamma, \gamma')$  is  $D$ -aligned.

The following is a combination of the results in [Yan19]; see [Cho22a] for another version.

**Proposition 2.10.** For each  $D > 0$  and  $K > 1$ , there exist  $E = E(D, K) > \max(D, K)$  and  $L = L(D, K)$  that satisfy the following.

- (1) Let  $x, y \in X$  and let  $\gamma_1, \dots, \gamma_n$  be  $K$ -BGIP axes whose domains are longer than  $L$  and such that  $(x, \gamma_1, \dots, \gamma_n, y)$  is  $D$ -aligned. Then the geodesic  $[x, y]$  has subsegments  $\eta_1, \dots, \eta_n$ , in order from left to right, that are longer than  $100E$  and such that  $\eta_i$  and  $\gamma_i$  are  $0.1E$ -fellow traveling. Moreover, the nearest point projection of  $x$  onto  $\gamma_1 \cup \dots \cup \gamma_n$  is equal to  $\pi_{\gamma_1}(x)$ .
- (2) Let  $\gamma_1, \dots, \gamma_n$  be  $K$ -BGIP axes such that:
  - (a) the beginning point of  $\gamma_{i+1}$  and the ending point of  $\gamma_i$  are the same for  $i = 1, \dots, n-1$ ;
  - (b) each  $\gamma_i$  has domain longer than  $L$ , and
  - (c)  $(\gamma_1, \dots, \gamma_n)$  is  $D$ -aligned.
Then the concatenation  $\gamma_1 \cup \dots \cup \gamma_n$  of  $\gamma_i$ 's is an  $E$ -BGIP axis.

*Proof.* Let us assume the hypothesis of Item (1). The first half of the conclusion is proved in [Yan19, Proposition 2.7], [Cho22a, Proposition 3.11] and [Cho22c, Proposition 2.20]. Now, note that  $[x, y]$  contains disjoint subsegments  $\eta_1, \dots, \eta_n$ , in order from left to right, such that  $\eta_i$  and  $\gamma_i$  are  $0.1E$ -fellow traveling. Moreover, each  $\eta_i$  is longer than  $100E$ . From this, for each  $i$  we deduce

$$d(x, \gamma_{i+1}) \geq d(x, \eta_{i+1}) - 0.1E \geq d(x, \eta_i) + 99E \geq d(x, \gamma_i) + 98E.$$

This implies the second half of the conclusion.

Item (2) is proved in [Yan19, Proposition 2.9].  $\square$

**Remark 2.11.** We will later deal with non-geodesic metrics, namely, the Green metrics on a relatively hyperbolic group  $G$ . Note that we have other metrics on  $G$  that is  $G$ -equivariantly quasi-isometric to  $d$ , namely, the word metrics on  $G$ . Since the word metrics are geodesic metrics, the alignment lemmata applies to the contracting axes in  $G$  with respect to them. For this reason, we always refer to the word metrics when defining the alignment of paths in relatively hyperbolic groups. We will revisit this topic in Section 2.10.



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

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

Recall our notation for reversal: we have  $\bar{\Gamma}(s) := (\Pi(s) o, \dots, o)$ .

**Definition 2.12** (cf. [Gou22, Definition 3.11], [Cho22a, Definition 3.14]). *For a given constant  $K_0 > 0$ , we define:*

- $D_0 = D(K_0, K_0)$  be as in Lemma 2.9,
- $E_0 = E(D_0, K_0)$ ,  $L_0 = L(D_0, K_0)$  be as in Proposition 2.10.

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

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

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

In addition, we say that  $S$  is large if it has cardinality at least 400. Once a long enough and large Schottky set  $S$  is understood, its element  $s$  is called a Schottky sequence and the translates of  $\Gamma(s)$  are called Schottky axes. When  $\mu$  is a probability measure on  $G$  such that  $S \subseteq (\text{supp } \mu)^n$ , then we say that  $S$  is a long enough and large Schottky set for  $\mu$ .

Using the notation for the reversal, Item (3) above can be seen as:

$$\#\{s \in S : (x, \Gamma(s)) \text{ and } (\bar{\Gamma}(s), x) \text{ are } K_0\text{-aligned}\} \geq \#S - 1.$$

Also, Item (4) is requiring that  $(\bar{\Gamma}(s), \Gamma(s'))$  is  $K_0$ -aligned.

**Proposition 2.13** ([Cho22a, Proposition 3.18]). *Let  $\mu$  be a non-elementary probability measure on  $G$ . Then for each  $N, L > 0$ , there exists  $K = K(N) > 0$  and  $n = n(N, L) > L$  such that there exists a  $K$ -Schottky set  $S \subseteq (\text{supp } \mu)^n$  of cardinality  $N$ .*

**Definition 2.14** ([Cho22a, Definition 2.25]). *Let  $S$  be a long enough Schottky set, let  $D > 0$ , let  $x, y \in X$ , let  $(\gamma_1, \dots, \gamma_N)$  be a sequence of Schottky axes, and let  $(\gamma_{i(1)}, \dots, \gamma_{i(n)})$  be its subsequence. If  $(x, \gamma_1, \dots, \gamma_N, y)$  is  $D$ -aligned, then we say that  $(x, \gamma_{i(1)}, \dots, \gamma_{i(n)}, y)$  is  $D$ -semi-aligned.*

By applying Proposition 2.10 to Schottky axes for a long enough Schottky set, we observe:

**Corollary 2.15.** *Let  $n > 0$  and let  $S \subseteq G^n$  be a long enough  $K_0$ -Schottky set, associated with the constants  $D_0, E_0$  as in Definition 2.12. If  $x, y \in X$  and Schottky axes  $\gamma_1, \dots, \gamma_N$  are such that*

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

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

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

We now record a technical lemma about Schottky sets. It says that we can incorporate a given BGIP isometry into the construction of Schottky sets.

**Proposition 2.16.** *Let  $(X, G)$  be as in Convention 2.4. Let  $A \subseteq G$  be a subset such that the semigroup  $\langle\langle A \rangle\rangle$  generated by  $A$  contains two independent BGIP elements  $g$  and  $h$ . Then for each  $n > 0$ , there exists  $m > 0$  and a large enough and long Schottky set  $S \subseteq A^m$  whose each element  $s \in S$  contains a subsequence of the form  $(\varphi_s, \varphi_s g, \dots, \varphi_s g^n)$  for some  $\varphi_s \in G$ .*

More precisely, let  $\alpha$  be a sequence in  $A$  such that  $\Pi(\alpha) = g$ . For each  $N > 0$  there exists  $K_0 = K_0(N) > 0$  and a finite set  $S'' \subseteq A^m$  of cardinality  $N$  such that, the set of the form

$$S := \{(s, \alpha^{(n)}, s) : s \in S''\}$$

is a  $K_0$ -Schottky set for any sufficiently large  $n$ .

*Proof.* By the assumption, the BGIP isometry  $g$  is realized as  $g = \Pi(\alpha)$  by some sequence  $\alpha$  in  $A$ . We denote its  $m$ -self concatenation by  $\alpha^{(m)}$ , i.e.,

$$\alpha^{(m)} := (\underbrace{\alpha, \dots, \alpha}_{m \text{ times}}).$$

Note that  $\Pi(\alpha^{(m)}) = g^m$  and  $\Gamma(\alpha^{(m)})$  is quasi-isometric (uniformly in  $m$ ) to its subpath  $(o, go, \dots, g^m o)$ . Because  $g$  is a BGIP isometry, there exists a uniform constant  $K_g$  such that  $\Gamma(\alpha^{(m)})$  is a  $K_g$ -BGIP axis for any  $m \in \mathbb{Z}_{>0}$ .

Let  $K_{temp}$  be  $K(N+1)$  as in Proposition 2.13 or  $K_g$ , whichever larger. Let:

- $D_{temp} = D(K_{temp}, K_{temp}) > K_{temp}$  be as in Lemma 2.9,
- $K_0 = E(D_{temp}, K_{temp}) > D_{temp}$ ,  $L_0 = L(D_{temp}, K_{temp})$  be as in Proposition 2.10,
- $D_0 = D(K_0, K_0)$  be as in Lemma 2.9,
- $E_0 = E(D_0, K_0)$ ,  $L_1 = L(D_0, K_0)$  be as in Proposition 2.10.

By Proposition 2.13, there exists a  $K_{temp}$ -Schottky set  $S' \subseteq A^m$  of cardinality  $N+1$ , with  $m > L_0 + L_1$ . Now for  $n > L_0 + L_1 + 2E_0 + 3$ ,  $\Gamma(\alpha^{(n)})$  is a  $K_g$ -BGIP axis whose domain is longer than  $L_0$ ,  $L_1$  and  $2E_0 + 2$ .

By the Schottky property of  $S'$ , we can take a subset  $S'' \subseteq S'$  of cardinality  $N$  such that

$$(\Pi(\alpha^{(n)})^{-1}o, \Gamma(s)) \text{ and } (\Gamma(s), \Pi(s)\Pi(\alpha^{(n)})o) \text{ are } K_{temp}\text{-aligned. } (\forall s \in S'').$$

Note that for every path  $\gamma$ ,  $((\text{beginning point of } \gamma), \gamma)$  and  $(\gamma, (\text{ending point of } \gamma))$  are always 0-aligned. This means that  $(o, \Gamma(s))$  and  $(\Gamma(s), \Pi(s)o)$  are 0-aligned for each  $s \in S''$ . Note also that  $\Gamma(\alpha^{(n)})$  and  $\Gamma(s)$  for  $s \in S'$  are  $K_{temp}$ -BGIP axes whose domain is longer than  $L_0$ . Lemma 2.9 implies that

$$(2.2) \quad (\Gamma(s), \Pi(s)\Gamma(\alpha^{(n)}), \Pi(s)\Pi(\alpha^{(n)})\Gamma(s)) \text{ is } D_{temp}\text{-aligned } (\forall s \in S'').$$

We now consider the set

$$S := \{(s, \alpha^{(n)}, s) : s \in S''\}.$$

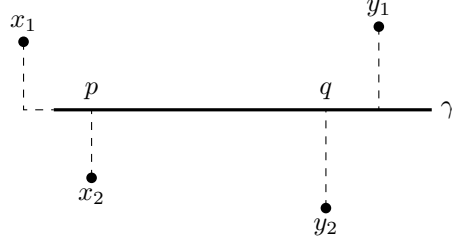
By Display 2.2 and Proposition 2.10,  $\Gamma(s, \alpha^{(n)}, s)$  is a  $K_0$ -BGIP axis for each  $s \in S''$ . Next, given  $x \in X$ , the Schottky property of  $S'$  tells us that

$$(2.3) \quad (x, \Gamma(s)) \text{ and } (\Gamma(s), \Pi(s)x) \text{ are } K_{temp}\text{-aligned}$$

for all  $s \in S'$  except at most one. Moreover, when the condition in Display 2.3 holds, then

$$\pi_{\Gamma(s, \alpha^{(n)}, s)}(x) = \pi_{\Gamma(s)}(x) \subseteq N_{K_{temp}}(o)$$

by Display 2.2, Proposition 2.10 and the fact that  $\Gamma(s)$  and  $\Gamma(\alpha^{(n)})$  are  $K_{temp}$ -BGIP axes whose domains are longer than  $L_0$ . Hence,  $(x, \Gamma(s, \alpha^{(n)}, s))$  is  $K_{temp}$ -aligned. For a similar reason, the condition in Display 2.3 implies that  $(\Gamma(s, \alpha^{(n)}, s), \Pi(s)\Pi(\alpha^{(n)})\Pi(s)x)$  is  $K_{temp}$ -aligned. Summing up,  $S$  satisfies the requirements for  $K_0$ -Schottky property. Furthermore,  $\#S = \#S'' = N$  holds, and each element of  $S$  has domain longer than  $L_1$ . Hence,  $S$  is a long enough and large Schottky set.  $\square$

FIGURE 1. 4 points and a geodesic in an  $\mathbb{R}$ -tree.

**2.4. Squeezing isometries.** We now explain a geometric ingredient for Theorem C and D.

**Definition 2.17.** Let  $(X, d)$  be a (possibly asymmetric) geodesic metric space. Let  $\gamma_1, \dots, \gamma_n$  be  $K$ -BGIP axes in  $X$ .

We say that the sequence  $(\gamma_1, \dots, \gamma_n)$  is  $(K, \epsilon)$ -squeezing if the following holds. Let  $x_1, x_2, y_1, y_2 \in X$  be points such that  $(x_i, \gamma_1)$  is  $K$ -aligned and  $(\gamma_n, y_i)$  is  $K$ -aligned for  $i = 1, 2$ . Then we have

$$|d(x_1, y_1) + d(x_2, y_2) - d(x_1, y_2) - d(x_2, y_1)| < \epsilon.$$

In particular, a BGIP axis  $\gamma$  is said to be  $(K, \epsilon)$ -squeezing if, for every quadruple of points  $x_1, x_2, y_1, y_2 \in X$  such that  $(x_i, \gamma, y_i)$  is  $K$ -aligned for  $i = 1, 2$ , we have

$$|d(x_1, y_1) + d(x_2, y_2) - d(x_1, y_2) - d(x_2, y_1)| < \epsilon.$$

Since we decided to use the word metric when discussing alignment in a relatively hyperbolic group, the definition reads as follows.

**Definition 2.18.** Let  $G$  be a relatively hyperbolic group equipped with a Green metric  $d$ . Let  $\gamma_1, \dots, \gamma_n$  be paths in  $G$  that are  $K$ -contracting with respect to the word metric.

We say that  $(\gamma_1, \dots, \gamma_n)$  is  $(K, \epsilon)$ -squeezing if the following holds. Let  $x_1, x_2, y_1, y_2 \in G$  be points such that  $(x_i, \gamma_1)$  is  $K$ -aligned and  $(\gamma_n, y_i)$  is  $K$ -aligned with respect to the word metric for  $i = 1, 2$ . Then we have

$$|d(x_1, y_1) + d(x_2, y_2) - d(x_1, y_2) - d(x_2, y_1)| < \epsilon.$$

Archetypal examples of squeezing axes are geodesic segments in an  $\mathbb{R}$ -tree. In Figure 1, 4 points are arranged so that  $x_1$  and  $x_2$  are on the left of  $\gamma$  while  $y_1$  and  $y_2$  are on the right of  $\gamma$ . Suppose that  $(x_i, \gamma, y_i)$  are  $K$ -aligned and that  $\gamma$  is longer than  $2K$ . Define  $p$  to be the rightmost one between  $\pi_\gamma(x_1)$  and  $\pi_\gamma(x_2)$ , and  $q$  to be the leftmost one between  $\pi_\gamma(y_1)$  and  $\pi_\gamma(y_2)$ . Then the assumption tells us that  $p$  is left to  $q$ , and we have

$$d(x_i, y_j) = d(x_i, p) + d(p, q) + d(q, y_j) \quad (i, j \in \{1, 2\}).$$

This implies that  $d(x_1, y_1) + d(x_2, y_2) - d(x_1, y_2) - d(x_2, y_1) = 0$ .

In general, when the  $K$ -alignment of  $(x_i, \gamma, y_j)$  forces  $[x_i, y_j]$  to pass nearby the midpoint of  $\gamma$ , then a similar computation is possible and  $\gamma$  is  $(K, \epsilon)$ -squeezing for some  $\epsilon$ . In view of Proposition 2.10, an aligned sequence of sufficiently long BGIP axes is squeezing. In the following subsections, we will see that the squeezing constant can be controlled explicitly in many spaces.

**2.5. Teichmüller space.** Let  $\Sigma$  be a connected, 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  $\text{Mod}(\Sigma)$  of  $\Sigma$  acts on  $X$  by isometries with respect to both metrics.

A mapping class in  $\text{Mod}(\Sigma)$  either is finite-order, preserves a multicurve or exhibits Anosov dynamics on  $\Sigma$  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,

Contraction Theorem], [Raf14, Theorem 8.1]; [BF09, Proposition 8.1]). Conversely, the orbit of a non-pseudo-Anosov mapping class lies in a quasi-flat and thus is not contracting. Since the mapping class group action is properly discontinuous, every pseudo-Anosov mapping class is WPD. In this context, a probability measure  $\mu$  on  $\text{Mod}(\Sigma)$  is non-elementary iff the semigroup generated by  $\text{supp } \mu$  contains two independent pseudo-Anosov mapping classes.

Each 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, Theorem 2] tells us that these fixed points are connected by a unique Teichmüller geodesic. Using this, we can observe:

**Fact 2.19.** *Let  $\varphi$  be a pseudo-Anosov mapping class and let  $\gamma : \mathbb{R} \rightarrow \mathcal{T}(\Sigma)$  be its invariant Teichmüller geodesic. Then for each  $\epsilon > 0$  there exists  $T > 0$  such that, if  $x, y \in \mathcal{T}(\Sigma)$  satisfy*

$$\pi_{\gamma}(x) \ni \gamma(t_1), \pi_{\gamma}(y) \ni \gamma(t_2)$$

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

This fact is well-known to the experts (cf. [EM11, Lemma 5.3]). In view of Proposition 2.13, Fact 2.19 has the following consequence:

**Proposition 2.20.** *Let  $(X, d)$  be the Teichmüller space with the Teichmüller metric, let  $G$  be the mapping class group and let  $o \in X$ . Let  $\mu$  be a non-elementary probability measure on  $G$  and let  $\epsilon > 0$ . Then there exists a long enough and large Schottky set  $S$ , associated with the constants  $D_0, E_0$  as in Definition 2.12, such that every Schottky axis is  $(D_0, \epsilon)$ -squeezing.*

The same conclusion for the Weil-Petersson geometry follows from the uniqueness of a recurrent Weil-Petersson geodesic connecting two ending laminations (Theorem 1.1, [BMM10]).

**2.6. CAT(0) spaces.** CAT(0) spaces are geodesic spaces where geodesic triangles are not fatter than the Euclidean triangles with the same side lengths. Typical examples complete Riemannian manifolds with non-positive sectional curvature including Euclidean spaces,  $n$ -dimensional hyperbolic spaces and their products. The Weil-Petersson metric renders Teichmüller space CAT(0): this is indeed the context of Proposition 2.21 below.

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:

**Proposition 2.21** ([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.

**Fact 2.22** ([BF09, Theorem 6.5]). *Let  $X$  be a proper CAT(0) space. Suppose that the action of a group  $\Gamma$  of isometries of  $X$  contains a WPD element (see [BF09, Definition 6.4]). Then  $\Gamma$  is non-elementary.*

**Fact 2.23** ([Ham09, Corollary 5.4]). *Let  $X$  be a proper CAT(0) space that admits a rank-1 isometry. Suppose that the limit set of a group  $\Gamma$  of isometries of  $X$  on the visual boundary has at least 3 points, and that  $\Gamma$  does not globally fix a point in  $\partial X$ . Then  $\Gamma$  is non-elementary.*

**Fact 2.24** ([CF10, Proposition 3.4]). *Let  $X$  be a proper CAT(0) space that admits a rank-1 isometry. Suppose that a group  $\Gamma$  of isometries of  $X$  does not globally fix a point in  $\partial X$  nor stabilize a geodesic line. Then  $\Gamma$  is non-elementary.*

**2.7. CAT(0) cube complexes.** CAT(0) cube complexes 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 some non-positively curved feature, namely, that its link is a flag complex. Originally proposed by Gromov, CAT(0) cube complexes 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 [Ago13], a major breakthrough in 3-manifold theory.

We refer the readers to [CS11] and [CFI16] for basic terminology about CAT(0) cube complexes. A CAT(0) cube complex contains plenty of halfspaces, which cut the space into two (closed) halfspaces. Each halfspace  $\mathfrak{h}$  is associated with a hyperplane that we denote by  $\hat{\mathfrak{h}}$ , and also with its reflected copy  $\overline{X} \setminus \mathfrak{h}$  that we denote by  $\mathfrak{h}^*$ . We say that two parallel hyperplanes of a CAT(0) cube complex are *strongly separated* if no hyperplane is transverse to both of them.

Caprace and Sageev showed the following rank rigidity phenomenon, which captures the action of an automorphism group  $G$  of a CAT(0) cube complex  $X$  by looking at  $X$  and the visual boundary  $\partial X$ .

**Fact 2.25.** *Let  $X$  be a CAT(0) cube complex, let  $x, y \in X$ , and let  $x \in \mathfrak{h}_1 \subsetneq \mathfrak{h}_2 \subsetneq \dots \subsetneq \mathfrak{h}_n \not\ni y$  be a nested sequence of distinct halfspaces between  $x$  and  $y$ . Then we have  $n \leq d(x, y)$ .*

**Proposition 2.26** ([CS11, cf. Theorem A]). *Let  $X$  be a finite-dimensional CAT(0) cube complex and  $G \leq \text{Aut}(X)$  be a group that does not globally fix a point nor stabilize a 1-dimensional flat in  $X \cup \partial X$ . Then  $X$  contains a geodesically convex  $G$ -invariant subcomplex  $Y$  such that either  $Y$  is the product of two unbounded subcomplexes or the action of  $G$  on  $Y$  is strongly non-elementary.*

*In particular, if  $G$  acts on an  $X$  essentially, without fixing a point in  $X \cup \partial X$ , and if  $X$  is irreducible, then  $G$  contains independent contracting isometries that skewers a pair of strongly separated hyperplanes.*

This result essentially follows from the Double Skewering Lemma of Caprace and Sageev in [CS11] that we describe below, with an additional information from [CFI16] and [Fer18].

**Fact 2.27** ([CS11, Proposition 3.2], [CFI16, Lemma 2.18, 2.24], [Fer18, Lemm 3.28]). *Let  $G$  be a group that essentially acts on an irreducible, finite-dimensional CAT(0) cube complex  $X$  without fixing a point. Then for any halfspace  $\mathfrak{h}$  of  $X$  associated with a hyperplane  $\hat{\mathfrak{h}}$ , there exists a rank-1 isometry  $g \in G$  such that  $\mathfrak{h}$  and  $g\hat{\mathfrak{h}}$  are strongly separated and such that  $\mathfrak{h} \subsetneq g\mathfrak{h}$ . (We say that  $g$  skewers  $\mathfrak{h}$  and  $g\hat{\mathfrak{h}}$ .) Moreover, there exists a geodesic segment  $\mathcal{I}$  connecting  $\mathfrak{h}$  to  $g\mathfrak{h}$  such that*

$$(2.4) \quad d(x, y) = d(x, p) + d(p, y)$$

*holds for every  $x \in \mathfrak{h}$ ,  $y \in (g\mathfrak{h})^*$  and  $p \in \mathcal{I}$ . (We call  $\mathcal{I}$  the bridge between  $\hat{\mathfrak{h}}$  and  $g\hat{\mathfrak{h}}$ .)*

Using this, we can observe that:

**Proposition 2.28.** *Let  $(X, d)$  be a CAT(0) cube complex with the combinatorial metric, let  $G$  be a countable automorphism group of  $X$ , and let  $o$  be a vertex of  $X$ . Let  $\mu$  be a strongly non-elementary probability measure on  $G$  and  $\epsilon > 0$ . Then there exists a long enough and large Schottky set  $S$ , associated with the constants  $D_0, E_0$  as in Definition 2.12, such that every Schottky axis is  $(E_0, 0)$ -squeezing with respect to  $d$ .*

*Proof.* Since  $\mu$  strongly non-elementary,  $\langle\langle \text{supp } \mu \rangle\rangle$  contains a skewer  $g$  for strongly separated hyperplanes, say  $\hat{\mathfrak{h}}$  and  $g\hat{\mathfrak{h}}$ , satisfying  $\mathfrak{h} \subsetneq g\mathfrak{h}$ . Let us pick  $x \in g\mathfrak{h} \setminus \mathfrak{h}$ . Since  $d(x, o)$  is finite, the number of mutually disjoint hyperplanes between  $x$  and  $o$  is finite. Using this, we can deduce that  $o \in g^m\mathfrak{h} \setminus g^n\mathfrak{h}$  for some  $m \neq n$ , and consequently, that  $o \in g^{k+1}\mathfrak{h} \setminus g^k\mathfrak{h}$  for some  $k$ . By replacing  $\mathfrak{h}$  with  $g^k\mathfrak{h}$ , we may assume that  $o \in g\mathfrak{h} \setminus \mathfrak{h}$ .

Keeping the choice of  $g$ , we pick  $K_0$  as in Proposition 2.16. This determines the constants  $D_0$  and  $E_0$ . Let  $n = 2E_0 + 2$ . By Proposition 2.16, there exists a long enough and large  $K_0$ -Schottky set  $S$  such that each  $s \in S$  contains a subsequence of the form  $(\varphi_s, \dots, \varphi_s g^n)$  for some  $\varphi_s \in G$ .

It remains to prove that  $\Gamma(s)$  is  $(E_0, 0)$ -squeezing for each  $s \in S$ . To show this, let  $x, y \in X$  be such that  $(x, \Gamma(s), y)$  is  $E_0$ -aligned. By Proposition 2.10,  $[x, y]$  contains a subsegment that is  $E_0$ -fellow traveling with  $\Gamma(s)$ . Hence, there exist  $p, q \in [x, y]$ ,  $p$  coming earlier than  $q$ , such that

$$d(p, \varphi_s o) < E_0, \quad d(q, \varphi_s g^n o) < E_0.$$

Since  $d(p, \varphi_s o) < E_0$  and  $\varphi_s o \in \varphi_s g \cdot \mathfrak{h}$ , Fact 2.25 tells us that  $p \in \Pi(s)g^{1+E_0}\mathfrak{h} \subseteq \Pi(s)g^{n-E_0-1}\mathfrak{h}$ . Similarly, from the ingredients  $d(q, \varphi_s g^n o) < E_0$ ,  $\varphi_s g^n o \notin \varphi_s g^n \mathfrak{h}$ , we deduce that  $q \notin \varphi_s g^{n-E_0}\mathfrak{h}$ . We then apply Fact 2.27 to conclude that

$$d(x, y) = d(x, p) + d(p, y)$$

for any point  $p$  in the bridge between  $\varphi_s g^{n-E_0-1}\hat{\mathfrak{h}}$  and  $\varphi_s g^{n-E_0}\hat{\mathfrak{h}}$ . Using this, given any  $x_1, x_2, y_1, y_2 \in X$  such that  $(x_i, \Gamma(s), y_i)$  is  $E_0$ -aligned, we observe

$$d(x_1, y_1) + d(x_2, y_2) = d(x_1, p) + d(x_2, p) + d(p, y_1) + d(p, y_2) = d(x_1, y_2) + d(x_2, y_1). \quad \square$$

An important family of examples comes from right-angled Artin groups (RAAGs). Recall that a RAAG  $\Gamma$  is associated with a simply connected CAT(0) cube complex  $\tilde{X}_\Gamma$ ;  $\Gamma$  acts properly and cocompactly on  $\tilde{X}_\Gamma$ , and the resulting quotient is called the *Salveti complex*  $X_\Gamma$  of  $\Gamma$ . It is proved in [BC12, Theorem 5.2] that if  $\Gamma$  is not a direct product, then the universal cover  $\tilde{X}_\Gamma$  of the Salvetti complex admits a rank-1 isometry.

**2.8. 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 2.29.** *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  $[x, y]$ .*

**2.9. Culler-Vogtmann Outer space.** We now turn to the Culler-Vogtmann Outer space equipped with the Lipschitz metric. We will only record the necessary facts; for precise definitions and proofs, we refer the readers to [Vog15], [FM11], [BF14] and [KMPT22a].

The Culler-Vogtmann Outer space  $CV_N$  parametrizes projectivized marked minimal metric graphs with fundamental group  $F_N$ . The outer automorphism group of  $F_N$ , denoted by  $\text{Out}(F_N)$ , naturally acts on  $CV_N$  by change of marking. Outer space has a natural simplicial structure that is preserved by the action of  $\text{Out}(F_N)$ . The difference of two marked graphs are measured by the Lipschitz constant of an 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  $\text{Out}(F_N)$ -orbit points of  $o$ , which are canonical  $d_{CV}$ -geodesics 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  $\gamma_1$  and  $\gamma_2$ , where  $\gamma_1$  is a rescaling segment on the simplex of  $x$  and  $\gamma_2$  is a folding path. (For the definition of rescaling/folding path, refer to [BF14].) Note that the orbit points  $G \cdot o$  are all  $\epsilon$ -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  $\phi \in \text{Out}(F_N)$  is *reducible* if there exists a free product decomposition  $F_N = C_1 * \cdots * C_k * C_{k+1}$ , with  $k \geq 1$  and  $C_1, \dots, C_{k+1} \neq \{e\}$ , such that  $\phi$  permutes the conjugacy classes of  $C_1, \dots, C_k$ . If not, we say that  $\phi$  is *irreducible*. We say that

$\phi$  is *fully irreducible* if no power of  $\phi$  is reducible. For us, an important point is that each fully irreducible automorphism  $\phi \in \text{Out}(F_N)$  has *bounded geodesic image property (BGIP)*, i.e., there exists  $K > 0$

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

holds for all geodesics  $\eta$  that does not enter the  $K$ -neighborhood of  $\{\phi^i o : i \in \mathbb{Z}\}$  ([KMPT22a, Theorem 7.8], cf. [Cho22c, Proposition 1.3]).

We now show a finer stability exhibited by certain outer automorphism. For each  $N \geq 3$ ,  $\text{Out}(F_N)$  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 2.30** ([KMPT22a, Lemma 5.9]). *Let  $\varphi \in \text{Out}(F_N)$  be a fully irreducible outer automorphism with lone axis  $\gamma : \mathbb{R} \rightarrow CV_N$  (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  $\epsilon$ -neighborhood of  $\gamma(t)$ .*

This implies the following:

**Fact 2.31.** *Let  $(X, d)$  be the Culler-Vogtmann Outer space of rank  $N$ , let  $G$  be the outer automorphism group  $\text{Out}(F_N)$  and let  $\mu$  be a strongly non-elementary probability measure on  $G$ . Then for each  $\epsilon > 0$ , there exist a long enough and large  $K_0$ -Schottky set  $S$  for  $\mu$ , associated with the constants  $D_0$  and  $E_0$ , such that each Schottky sequence is  $(E_0, \epsilon)$ -squeezing.*

**2.10. Relatively hyperbolic groups.** Relatively hyperbolic groups are generalizations of free products of groups and geometrically finite Fuchsian and Kleinian groups. For detailed theory, we refer the readers to [Bow12], [Yam04], [Osi06] and [GP13]. We recall 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. Another characterization of hyperbolic Cayley graphs is that geodesics in hyperbolic Cayley graphs are  $K$ -contracting for some uniform  $K$  (cf. [Pap95]).

More generally, let  $\{P_\alpha\}_\alpha$  be a (conjugacy invariant) 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  $P_\alpha$ '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*.

Let us now fix a word metric  $d_S$  for a finite generating set  $S$  of a relatively hyperbolic group  $G$ . Then each hyperbolic element of  $G$  is contracting with respect to  $d_S$  ([Yan14, Lemma 7.2], [GP16, Proposition 8.5]). Meanwhile, there is another natural metric on  $G$  studied by Blachère, Haïssinsky and Mathieu [BHM11]. Given a symmetric, finitely supported and admissible probability measure  $\mu$  on  $G$ , we define the Green function associated to  $\mu$  by

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

and the Green metric with respect to  $\mu$  by

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

**Remark 2.32** ([BHM11, Proposition 3.5]). *Green metrics are not geodesic in general but are roughly geodesic [Anc90]. 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  $\gamma : [0, L] \rightarrow G$  such that  $\gamma(0) = g$ ,  $\gamma(L) = h$  and*

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

holds for all  $t_1, t_2 \in [0, L]$ . For this metric, we can define that a  $K$ -quasigeodesic  $\gamma : I \rightarrow G$  is a  $K$ -contracting axis if  $\text{diam}(\pi_\gamma(\eta)) < K$  holds for all  $C$ -coarse geodesic  $\eta$  that does not enter the  $K$ -neighborhood of  $\gamma$ . And then we can discuss the alignment lemmata by using  $C$ -coarse geodesics in place of geodesics. Since we already have a theory of alignment on the relatively hyperbolic group  $G$  in terms of the word metric, we will not go deeper.

Let  $f$  be a hyperbolic element of  $G$  and let  $\epsilon > 0$ . Then the orbit  $\{f^n\}_{n \in \mathbb{Z}}$  of  $f$  cannot stay long in the  $\epsilon$ -neighborhood of each peripheral subgroup  $gP_\alpha$ . Indeed, if  $f^n, f^{n+1}, \dots, f^m \in N_\epsilon gP_\alpha$  for some  $n \ll m$ ,  $g \in G$  and a peripheral subgroup  $P_\alpha$ , we deduce that  $f^i a, f^j a \in gP_\alpha$  for some  $i \ll j$  and  $a \in N_\epsilon(id)$ , which means that  $f^{j-i} \in a^{-1}P_\alpha a$ , contradicting the fact that  $f$  is a hyperbolic element. Hence, there exists  $L > 0$  such that  $\{f^n o, f^{n+1} o, \dots, f^{n+L} o\}$  is not contained in  $N_\epsilon(gP_\alpha)$  for any  $n \in \mathbb{Z}$ ,  $g \in G$  and any peripheral  $P_\alpha$ . We record the strong Ancona inequality along a geodesic that does not go deep into any translates of peripheral subgroups.

**Proposition 2.33** ([DG21, Proposition 2.2, Theorem 3.14]). *Let  $G$  be a non-elementary relatively hyperbolic group, with a word metric  $d_S$  and a Green metric  $d_G^\mu$ . Let  $f$  be a hyperbolic element of  $G$  and let  $K > 0$ . Then there exists  $L > 0$  such that, if  $\gamma_1, \dots, \gamma_N$  are paths on  $G$  of the form*

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

*and if  $x, y, x', y' \in G$  satisfy that  $(x, \gamma_1, \dots, \gamma_N, y)$ ,  $(x', \gamma_1, \dots, \gamma_N, y')$  are  $K$ -aligned (with respect to  $d_S$ ), then we have*

$$(2.5) \quad |d_G^\mu(x, y) + d_G^\mu(x', y') - d_G^\mu(x, y') - d_G^\mu(x', y)| \leq e^{-N}.$$

Combining this with Proposition 2.16, we deduce:

**Corollary 2.34.** *Let  $G$  be a non-elementary relatively hyperbolic group equipped a word metric  $d$  and a Green metric  $d_G$ . Then there exist a long and large enough  $K_0$ -Schottky set  $S$  such that for each  $N$ , every  $D_0$ -semi-aligned sequence of Schottky axes  $(\gamma_1, \dots, \gamma_N)$  is  $(E_0, e^{-N})$ -squeezing.*

In some cases, we have an estimate analogous to Inequality 2.5 on the Cayley graph of  $G$  with the word metric. Two notable examples are surface groups and free products. First consider the Cayley graph of  $G = \langle a_1, \dots, a_g, b_1, \dots, b_g | \prod_{i=1}^g [a_i, b_i] \rangle$ , equipped with the word metric  $d$  for the generating set  $\{a_1, \dots, a_g, b_1, \dots, b_g\}$ . Note that:

**Lemma 2.35** (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  $\gamma$  from  $x$  to  $y$  containing  $w^n$  as a subword, and*
- (2)  *$\pi_\gamma(z)$  is left to the  $w^n$ -subword of  $\gamma$  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 2.35 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. From this, we can deduce:

**Corollary 2.36.** *Let  $w = a_i^k b_{i+1}^k$  and let  $x, y \in G$  be such that*

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

*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*

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

*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  $G = *_\alpha G_\alpha$ , with the word metric for the standard generating set  $S = \bigcup \{\text{generating sets of } G_\alpha\}$ . Indeed, if we take two distinct free factors  $G_\alpha, G_\beta$  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 2.35.

### 3. PIVOTING TECHNIQUE: SUMMARY OF RESULTS

In this section, we review the concept of pivoting technique pioneered by Gouëzel in [Gou22]. Most of the results we introduce here originate from [Gou22] and further adapted to the current setting in a series of paper [Cho22a], [Cho22b]. While these previous papers are concerned with the self-convolutions of a given probability measure, we need estimates for convolutions of possibly different probability measures later. The arguments in [Gou22] still works through with some adequate adaptation, and we provide proof here for completeness. Readers who are familiar with the techniques in [Gou22] can assume the following result and proceed to Section 4.

**Proposition 3.1.** *Let  $0 < \epsilon \leq 1$  and let  $S \subseteq G^{M_0}$  be a long and large enough  $K_0$ -Schottky set. Then there exists  $\kappa = \kappa(\epsilon, S) > 0$  such that the following holds.*

*Let  $\{\mu_i : i \in \mathbb{Z}_{>0}\}$  be probability measures such that for each  $s \in S$  and  $k \in \mathbb{Z}$ ,*

$$(\mu_{k+1} \times \cdots \times \mu_{k+M_0})(s) > \epsilon.$$

*Let  $g_i$  be independently distributed according to  $\mu_i$  for each  $i \in \mathbb{Z}$ , all independently.*

*Then for arbitrary  $M$  integers  $t(1), \dots, t(M)$  and for arbitrary  $g_{t(1)}, \dots, g_{t(M)} \in G$ , we have*

$$\mathbb{P} \left( \begin{array}{l} \exists \text{ Schottky axes } \gamma_1, \gamma_2, \dots, \gamma_{\lceil \kappa n \rceil} \subseteq \{o, Z_1 o, Z_2 o, \dots, Z_n o\} \text{ such that} \\ (Z_{-m} o, \gamma_1, \dots, \gamma_{\lceil \kappa n \rceil}, Z_{n+l} o) \text{ is } D_0\text{-semi-aligned for all } m, l \geq 0 \end{array} \middle| g_{t(1)}, g_{t(2)}, \dots, g_{t(M)} \right) \geq 1 - \frac{1}{\kappa} e^{-\kappa n - M}.$$

*In particular, we have*

$$\mathbb{P} \left( \begin{array}{l} \text{for each } m, l \geq 0 \text{ there exists } i \in \{0, \dots, M_0\} \text{ such that} \\ d(Z_i o, [Z_{-m} o, Z_{n+l} o]) \leq E_0 \end{array} \middle| g_{t(1)}, g_{t(2)}, \dots, g_{t(M)} \right) \geq 1 - \frac{1}{\kappa} e^{-\kappa n - M}.$$

We first reduce random walks (and more generally, products of independent increments) to a combinatorial model. Here is a toy case: suppose that a probability measure  $\mu$  on  $G$  puts weights  $\epsilon$  on  $a, b \in G$  each. Then  $\mu = \epsilon \cdot 1_{\{a,b\}} + (\mu - \epsilon 1_{\{a,b\}})$  holds. Hence, a random path  $(g_1, \dots, g_n)$  following the law  $\mu^n$  can be obtained as follows: first put flags on each index independently for probability  $\epsilon$ . Then we choose  $g_i$  according to the law  $1_{\{a,b\}}$  for flagged  $i$ , and according to the law  $\frac{1}{1-\epsilon} \cdot (\mu - \epsilon 1_{\{a,b\}})$  for unflagged  $i$ , all independently. The following proposition describes a situation analogous to the above, but with some indices reserved (which will account for conditional probabilities).

**Lemma 3.2.** *For each  $0 < \epsilon \leq 1$  and  $M_0 \in \mathbb{Z}_{>0}$ , there exists  $\kappa = \kappa(\epsilon, M_0) > 0$  such that the following holds.*

*Let  $n, M \in \mathbb{Z}_{>0}$ . Let  $S \subseteq G^{M_0}$  be a finite set of sequences in  $G$ . Let  $t(1), \dots, t(M)$  be  $M$  distinct integers. Let  $\mu_1, \mu_2, \dots$  be probability measures such that*

- (1)  $\mu_{t(l)}$  is the Dirac measure concentrated on an element  $g_{t(l)} \in G$  for  $l = 1, \dots, M$ , and
- (2)

$$(\mu_{M_0 k+1} \times \cdots \times \mu_{M_0 k+M_0})(s) > \epsilon$$

*whenever  $s \in S$  and  $\{M_0 k + 1, \dots, M_0 k + M_0\} \subseteq \{1, \dots, n\} \setminus \{t(1), \dots, t(M)\}$ .*

*Then there exists a probability space  $\Omega$ , a measurable partition  $\mathcal{Q}$  on  $\Omega$  and random variables*

$$\begin{aligned} \{g_1, g_2, \dots\} &\in G^{\mathbb{Z}_{>0}}, \\ \{\vartheta(1) < \vartheta(2) < \dots\} &\subseteq \mathbb{Z}_{\geq 0} \end{aligned}$$

*such that the following hold:*

(1) For each  $n$  we have

$$\mathbb{P}\left(\{\vartheta(1), \dots, \vartheta(\lfloor \epsilon^4 n / 8M_0 \rfloor - M)\} \subseteq \{0, 1, \dots, \lfloor n / 4M_0 \rfloor - 1\}\right) \geq 1 - \frac{1}{\kappa} e^{-\kappa n - M}.$$

(2)  $(g_1, g_2, \dots)$  is distributed according to  $\mu_1 \times \mu_2 \times \dots$ .

(3)  $n$  each equivalence class  $\mathcal{F} \in \mathcal{Q}$ , the RVs  $\vartheta(1)(\omega), \vartheta(2)(\omega), \dots$  are constant and the  $M_0$ -long sequences

$$(3.1) \quad \left\{ \begin{array}{l} (g_{4M_0 \vartheta(l)+1}, \dots, g_{4M_0 \vartheta(i)+M_0}), \\ (g_{4M_0 \vartheta(l)+M_0+1}, \dots, g_{4M_0 \vartheta(i)+2M_0}), \\ (g_{4M_0 \vartheta(l)+2M_0+1}, \dots, g_{4M_0 \vartheta(i)+3M_0}), \\ (g_{4M_0 \vartheta(l)+3M_0+1}, \dots, g_{4M_0 \vartheta(i)+4M_0}) \end{array} : l = 1, 2, \dots \right\}$$

are all independently and uniformly distributed on  $S$ . Furthermore, all the increments  $g_i$ 's not appearing in Display 3.1 are fixed (constant) on  $\mathcal{F}$ .

A version of this lemma was proven in [Cho22b]: see [Cho22b, Lemma 4.6]. We give proof for the reader's convenience.

*Proof of Lemma 3.2.* Let

$$\mathcal{I} := \{\lfloor t(l)/4M_0 \rfloor : l = 1, \dots, M\}, \quad \mathcal{J} := \{4M_0 i + 1, \dots, 4M_0 + 4M_0 : i \in \mathcal{I}\}.$$

We now consider the partition  $\mathcal{P}$  of the ambient probability space made by the values of  $\{g_i : i \in \mathcal{J}\}$ . On each  $\mathcal{E} \in \mathcal{P}$ ,  $\mu_i|_{\mathcal{E}}$  are Dirac atoms for  $i \in \mathcal{J}$ , and  $\mu_i|_{\mathcal{E}} = \mu_i$  otherwise. Our goal is to realize the RVs

$$\{g_1, \dots, g_n\} \in G, \quad \{\vartheta(1), \dots, \vartheta(\lfloor \epsilon^4 n / 8M_0 \rfloor - M)\} \subseteq \{0, \dots, \lfloor n / 4M_0 \rfloor - 1\} \setminus \mathcal{I},$$

and a partition  $\mathcal{Q}_{\mathcal{E}}$  on  $\mathcal{E}$  that satisfy the conditions (1), (2) and (3) with respect to  $\mathbb{P}(\cdot|_{\mathcal{E}}) \sim \mu_1|_{\mathcal{E}} \times \dots \times \mu_n|_{\mathcal{E}}$ . If this is possible, then  $\bigcup_{\mathcal{E} \in \mathcal{P}} \mathcal{Q}_{\mathcal{E}}$  becomes the desired partition.

Hence, from now on, we assume that  $\mu_i$ 's are Dirac atoms for  $i \in \mathcal{J}$ . Let us denote the uniform measure on  $S$  by  $\mu_S$ . From the assumption, we have the following decomposition for each  $k \in \mathbb{Z}_{>0} \setminus \mathcal{I}$ :

$$\mu_{4M_0 k+1} \times \dots \times \mu_{4M_0 k+M_0} = \epsilon^4 \cdot \mu_S^4 + (1 - \epsilon^4) \cdot \nu_k$$

holds for some probability measure  $\nu_k$ .

We consider Bernoulli RVs  $(\rho_i)_{i \geq 0}$  with average  $\epsilon^4$ ,  $(\eta_i)_{i \geq 0}$  with the law  $\mu_S^4$ , and  $v_k$  with the law  $\nu_k$  for  $k \in \mathbb{Z}_{\geq 0}$ . All RVs are set to be independent. Now let

$$(g_{4M_0 k+1}, \dots, g_{4M_0(k+1)}) = \begin{cases} v_k & \text{when } \rho_k = 0, \\ \eta_k & \text{when } \rho_k = 1 \end{cases} \quad (k \in \mathbb{Z}_{\geq 0} \setminus \mathcal{I}).$$

For  $k \in \mathcal{I}$ , we decide  $(g_{4M_0 k+1}, \dots, g_{4M_0(k+1)})$  using the Dirac measure  $\mu_{4M_0 k+1} \times \dots \times \mu_{4M_0(k+1)}$ . Then  $(g_1, g_2, \dots)$  is distributed according to  $\mu_1 \times \mu_2 \times \dots$ . We denote by  $\Omega$  the ambient probability space on which  $(\rho_i, v_i, \eta_i)_i, (g_i)_i$  are all measurable.

Recall that  $\rho_i$ 's are i.i.d. Bernoulli RVs with average  $\epsilon^4$ . Chernoff-Hoeffding's inequality implies that  $\mathbb{P}\left(\sum_i \rho_i \leq \frac{1}{2} \epsilon^4 (n/4M_0 - M)\right)$  decays exponentially:

$$\mathbb{P}\left(B_n := \left\{ \sum_{i \in \{0, 1, \dots, n-1\} \setminus \mathcal{I}} \rho_i < \frac{\epsilon^4 n}{8M_0} - M \right\}\right) \leq \frac{1}{\kappa} e^{-\kappa n}$$

holds for some  $\kappa = \kappa(\epsilon, M_0)$ . This gives the first item.

For  $\omega \in B$ , we collect those indices  $i$  at which  $\rho_i = 1$  and denote them by  $\vartheta(1), \vartheta(2), \dots$  in the increasing order. We now define  $\mathcal{Q}$  of  $B$  to be the partition determined by the values of

$$(3.2) \quad \{\rho_i, v_i : i \notin \mathcal{I}\} \cup \{g_i : i \in \mathcal{J}\}$$

and define

$$\begin{aligned}
 w_{i-1} &:= g_{4M_0[\vartheta(i-1)+1]+1} \cdots g_{4M_0 \vartheta(i)} & (i = 1, \dots, \lfloor \epsilon^4 n / 8M_0 \rfloor), \\
 w_{\lfloor \epsilon^4 n / 8M_0 \rfloor} &:= g_{4M_0 \vartheta(\lfloor \epsilon^4 n / 8M_0 \rfloor)+1} \cdots g_n, \\
 \alpha_i &:= (g_{4M_0 \vartheta(i)+1}, \dots, g_{4M_0 \vartheta(i)+M_0}), & (i = 1, \dots, \lfloor \epsilon^4 n / 8M_0 \rfloor), \\
 \beta_i &:= (g_{4M_0 \vartheta(i)+M_0+1}, \dots, g_{4M_0 \vartheta(i)+2M_0}), & (i = 1, \dots, \lfloor \epsilon^4 n / 8M_0 \rfloor), \\
 \gamma_i &:= (g_{4M_0 \vartheta(i)+2M_0+1}, \dots, g_{4M_0 \vartheta(i)+3M_0}), & (i = 1, \dots, \lfloor \epsilon^4 n / 8M_0 \rfloor), \\
 \delta_i &:= (g_{4M_0 \vartheta(i)+3M_0+1}, \dots, g_{4M_0 \vartheta(i)+4M_0}). & (i = 1, \dots, \lfloor \epsilon^4 n / 8M_0 \rfloor),
 \end{aligned}
 \tag{3.3}$$

On each equivalence class  $\mathcal{F}$  of  $\mathcal{Q}_n$ , the increments involved in  $w_i$ 's are constants (since these are determined by the RVs in Display 3.2) and  $(\alpha_i, \beta_i, \gamma_i, \delta_i)$  are i.i.d.s with distribution  $\mu_S^4$ . This ends the proof.  $\square$

**Remark 3.3.** Assuming that  $M$  is pre-determined (say  $M = 1$ ), by enlarging  $\kappa(\epsilon, S)$  if necessary, we even have

$$\mathbb{P}\left(\#\left(\{\vartheta(1), \vartheta(2), \dots\} \cap \{n, n+1, \dots, n+k\}\right) \geq 0.5\epsilon^4(n-k)\right) \geq 1 - \frac{1}{\kappa}e^{-\kappa k}$$

for any  $n, k \geq 0$  for some  $\kappa = \kappa(\epsilon, S)$ .

The above lemma associates each  $\omega \in \Omega$  with an infinite word

$$w_0 \Pi(\alpha_1) \Pi(\beta_1) \Pi(\gamma_1) \Pi(\delta_1) \cdots w_k \Pi(\alpha_k) \Pi(\beta_k) \Pi(\gamma_k) \Pi(\delta_k) \cdots$$

Here,  $w_i$ 's are fixed isometries, while  $\alpha_i, \beta_i, \gamma_i, \delta_i$ 's are independently and uniformly distributed on  $S$ . It might be the case that  $w_0$  is the inverse of the remaining subword and  $g_1 \cdots g_n$  becomes  $id$ . The following proposition, which is the heart of the pivoting technique, says that such a situation is unlikely.

**Proposition 3.4.** For each  $N_0 \geq 400$ , there exists  $\kappa = \kappa(N_0) > 0$  such that the following holds.

Fix a long enough Schottky set  $S$  with cardinality  $N_0$  and a sequence  $(w_i)_{i=0}^\infty$  in  $G$ . Suppose that  $\{\alpha_i, \beta_i, \gamma_i, \delta_i : i > 0\}$  are independently and uniformly distributed on  $S$ . Then there exists a measurable set

$$P(\mathbf{s} = \{\alpha_i, \beta_i, \gamma_i, \delta_i : i > 0\}) = \{j(1) < j(2) < \dots\} \subseteq \mathbb{Z}_{>0}$$

and a measurable partition  $\mathcal{P}$  of the probability space that satisfy the following:

(1) (alignment) For each  $\mathbf{s} \in S^{\mathbb{Z}_{>0}}$ , if we define

$$U_1 := w_0 \Pi(\alpha_1), \quad U_{i+1} := U_i \cdot \Pi(\beta_i) \cdot \Pi(\gamma_i) \Pi(\delta_i) w_i \Pi(\alpha_{i+1}) \quad (i = 1, \dots, n),$$

then the sequence

$$(o, U_{j(1)} \Gamma(\beta_{j(1)}), U_{j(2)} \Gamma(\beta_{j(2)}), \dots)$$

is  $D_0$ -semi-aligned.

(2) (high chance) For each  $n$ , we have

$$\mathbb{P}(\#(P(\mathbf{s}) \cap \{1, \dots, n\}) \leq (1 - 10/N_0)n) \leq e^{-\kappa n}.$$

(3) On each  $\mathcal{E} \in \mathcal{P}$ , the set  $P$  is constant, as well as

$$\{\alpha_i, \gamma_i, \delta_i : i > 0\}, \quad \{\beta_i : i \notin P\}.$$

Meanwhile, on  $\mathcal{E}$ ,  $(\beta_{j(1)}, \beta_{j(2)}, \dots)$  are independently and uniformly distributed on  $S$ .

This proposition was first proved in [Gou22, Section 4B] and adapted in [Cho22a, Section 5.1]. Since the statements there are slightly different from this one, let us sketch the proof.

*Proof of Proposition 3.4.* In [Cho22a, Section 5.1], we proved:

**Proposition 3.5.** *In the setting of Proposition 3.4, there exists  $K = K(N_0) > 0$ , and for each  $n$  there exists a measurable set  $P_n = \{j(1) < j(2) < \dots < j(\#P_n)\} \subseteq \{1, \dots, n\}$  and an equivalence relation  $\sim_n$  such that the following holds:*

(a) *We have*

$$(3.4) \quad \mathbf{s} \sim_n \mathbf{s}' \Leftrightarrow \begin{cases} \alpha_i = \alpha'_i, \delta_i = \delta'_i & \text{for } i \in \{1, \dots, n\}, \\ \beta_i = \beta'_i, \gamma_i = \gamma'_i & \text{for } i \in \{1, \dots, n\} \setminus P_n(\mathbf{s}). \end{cases}$$

*Furthermore,  $P_n(\mathbf{s}) = P_n(\mathbf{s}')$  if  $\mathbf{s} \sim_n \mathbf{s}'$ .*

(b)  *$(o, U_{j(1)}\Gamma(\beta_{j(1)}), \dots, U_{j(\#P_n)}\Gamma(\beta_{j(\#P_n)}), U_n\Pi(\beta_n)\Pi(\gamma_n)\Pi(\delta_n)w_no)$  is  $D_0$ -semi-aligned.*

(c)  *$\mathbb{P}(\#P_n(\mathbf{s}) \leq (1 - 10/N_0)n) \leq e^{-n/K}$ . More generally, we have*

$$\mathbb{P}(\#P_{n+k}(\mathbf{s}) \leq \#P_n(\mathbf{s}) + (1 - 10/N_0)k) \leq e^{-k/K} \quad (\forall n, k \geq 0).$$

(d) *for each  $n \in \mathbb{Z}_{>0}$ , either  $P_{n+1} = P_n \cup \{n\}$  or  $P_{n+1}$  is an initial section of  $P_n$ , i.e.,  $P_{n+1} = \{\text{the first } \#P_{n+1} \text{ elements of } P_n\}$ .*

Item (a) is according to the definition of pivoting in the paragraph before [Cho22a, Proposition 5.6]; in our setting  $\mathbf{v}$  is kept to be identity sequence, cf. the paragraph before [Cho22a, Section 5.2]. The condition in Display 3.4 is indeed an equivalence relation thanks to [Cho22a, Lemma 5.5]. Item (b) is due to [Cho22a, Proposition 5.1]. Item (c) is proven in [Cho22a, Corollary 5.8]. Item (d) is according to the definition of the set of pivotal times before [Cho22a, Proposition 5.1].

Let us define  $P := \liminf_n P_n = \limsup_n P_n$ . Here, the second equality holds for the following reason: if an integer  $k$  is removed from some  $P_n$ , it cannot appear again in  $P_{n+1}, P_{n+2}, \dots$ . Hence, if  $k$  appears infinitely often in  $P_1, P_2, \dots$ , then  $k \in P_k, P_{k+1}, \dots$ . In fact, we observe:

$$P \subseteq P_k \cup \{k+1, k+2, \dots\} \quad \text{for each } k = 1, 2, \dots$$

We now define a relation:

$$\mathbf{s} \sim \mathbf{s}' \Leftrightarrow \begin{cases} \alpha_i = \alpha'_i, \gamma_i = \gamma'_i, \delta_i = \delta'_i & \text{for } i = 1, 2, \dots, \\ \beta_i = \beta'_i & \text{for } i \in \{1, 2, \dots\} \setminus P(\mathbf{s}). \end{cases}$$

Then  $\sim$  is a finer relation than each  $\sim_n$ . Hence, if  $\mathbf{s} \sim \mathbf{s}'$  then  $P_n(\mathbf{s}) = P_n(\mathbf{s}')$  for each  $n$ , which implies  $P(\mathbf{s}) = P(\mathbf{s}')$ . It follows that  $\sim$  is an equivalence relation.

Furthermore, any initial section of  $P$  is an initial section of some (in fact, all but finitely many)  $P_k$ . This combined with Item (a) implies Item (1), the alignment property. Furthermore, for each  $M$  and  $N$  we have

$$\#P_k(\mathbf{s}) \geq N \text{ for all } k \geq M \Rightarrow P(\mathbf{s}) \text{ contains the } N\text{-initial section of } P_M(\mathbf{s}).$$

This observation and Item (c) imply that for every  $n$ ,

$$\begin{aligned} \mathbb{P}(P(\mathbf{s}) \cap P_n(\mathbf{s}) \subseteq \{1, \dots, n\} \text{ has at least } (1 - 10/N_0)n \text{ elements}) \\ \geq \mathbb{P}(\#P_k(\mathbf{s}) \geq (1 - 10/N_0)n \text{ for each } k \geq n) \\ \geq 1 - \sum_{k \geq n} e^{-k/K} \geq 1 - \frac{1}{1 - e^{-1/K}} e^{-n/K}. \quad \square \end{aligned}$$

Given RVs  $g_1, g_2, \dots$ , we introduced the notation  $Z_i := g_1 \cdots Z_i$ . For convenience we also write

$$(3.5) \quad \mathbf{Y}_i := (Z_{i-M_0}, Z_{i-M_0+1}, \dots, Z_i)$$

when  $M_0 > 0$  is understood. We now recall a notion:

**Definition 3.6** ([Cho22a, Definition 4.1]). *Let  $S$  be a long enough Schottky set in  $G^{M_0}$ , and let  $\Omega$  be a probability space with  $G$ -valued RVs  $\{g_i : i > 0\}$ .*

*A subset  $\mathcal{E}$  of  $\Omega$ , accompanied with the choice of a subset  $\mathcal{P}(\mathcal{E}) = \{j(1) < j(2) < \dots\} \subseteq M_0 \mathbb{Z}_{>0}$ , is called a pivotal equivalence class (for  $S$ ) if:*

(1) *for each  $i \notin \{j(k) - l : k \geq 1, l = 0, \dots, M_0 - 1\}$ ,  $g_i(\omega)$  is fixed on  $\mathcal{E}$ ;*

(2) for each  $\omega \in \mathcal{E}$  and  $k \geq 1$ , the following is a Schottky sequence:

$$s_k(\omega) := (g_{j(k)-M_0+1}(\omega), g_{j(k)-M_0+2}(\omega), \dots, g_{j(k)}(\omega)) \in S;$$

(3) for each  $\omega \in \mathcal{E}$ ,  $(o, \mathbf{Y}_{j(1)}(\omega), \mathbf{Y}_{j(2)}(\omega), \dots)$  is  $D_0$ -semi-aligned, and

(4) on  $\mathcal{E}$ ,  $\{s_1(\omega), s_2(\omega), \dots\}$  are i.i.d.s distributed according to the uniform measure on  $S$ .

We say that  $\mathcal{P}(\mathcal{E})$  is the set of pivotal times for  $\mathcal{E}$ .

When a pivotal equivalence class  $\mathcal{E} \subseteq \Omega$  is understood, with the set of pivotal times  $\mathcal{P}(\mathcal{E})$ , for each element  $\omega$  of  $\mathcal{E}$  we call  $\mathcal{P}(\mathcal{E})$  the set of pivotal times for  $\omega$  and write it as  $\mathcal{P}(\omega)$ .

Lemma 3.2 and Proposition 3.4 together imply the following:

**Proposition 3.7.** For each  $0 < \epsilon \leq 1$  and  $M_0 \in \mathbb{Z}_{>0}$ , there exists  $\kappa = \kappa(\epsilon, M_0) > 0$  such that the following holds.

Let  $n, M \in \mathbb{Z}_{>0}$ . Let  $S \subseteq G^{M_0}$  be large and long-enough  $K_0$ -Schottky set. Let  $t(1), \dots, t(M)$  be  $M$  distinct integers. Let  $\mu_1, \mu_2, \dots$  be probability measures such that

- (1)  $\mu_{t(l)}$  is the Dirac measure concentrated on an element  $g_{t(l)} \in G$  for  $l = 1, \dots, M$ , and
- (2) whenever  $s \in S$  and  $\{k+1, \dots, k+M_0\} \subseteq \{1, \dots, n\} \setminus \{t(1), \dots, t(M)\}$ , we have

$$(\mu_{k+1} \times \dots \times \mu_{k+M_0})(s) > \epsilon.$$

Then there exist a probability space  $\Omega$  equipped with RVs

$$(g_1, g_2, \dots) \text{ distributed according to } \mu_1 \times \mu_2 \times \dots,$$

and a measurable partition  $\mathcal{Q}$  of  $\Omega$  into pivotal equivalence classes for  $S$  such that

$$(3.6) \quad \mathbb{P}(\omega : \#(\mathcal{P}(\omega) \cap \{1, \dots, n\}) \leq \kappa n - M) \leq \frac{1}{\kappa} e^{-\kappa n - M}$$

for each  $n$ .

When we are concerned with the pivotal time construction for step  $n$ , we do not need to construct  $P$  out of  $P_k$ 's. By combining Proposition 3.5 with Lemma 3.2 (with Remark 3.3 in mind), we obtain the following:

**Proposition 3.8.** For each  $0 < \epsilon \leq 1$  and  $M_0 \in \mathbb{Z}_{>0}$ , there exists  $\kappa = \kappa(\epsilon, M_0) > 0$  such that the following holds.

Let  $n \in \mathbb{Z}_{>0}$ . Let  $S \subseteq G^{M_0}$  be large and long-enough  $K_0$ -Schottky set. Let  $\mu_1, \mu_2, \dots$  be probability measures such that for each  $s \in S$  and  $k \in \mathbb{Z}$ , we have

$$(\mu_{M_0 k+1} \times \dots \times \mu_{M_0 k+M_0})(s) > \epsilon.$$

Let us then change just one measure  $\mu_i$  into a Dirac measure.

Then there exist a probability space  $\Omega_n$  equipped with RVs

$$(g_1, g_2, \dots, g_n) \text{ distributed according to } \mu_1 \times \mu_2 \times \dots \times \mu_n$$

and a measurable partition  $\mathcal{Q}_n$  of  $\Omega_n$  into pivotal equivalence classes for  $S$  such that the following conditions are satisfied:

- (1) for each pivotal equivalence class  $\mathcal{E}$ , the set of pivotal times  $\mathcal{P}_n(\mathcal{E}) = \{j(1) < \dots < j(\#\mathcal{P}_n(\mathcal{E}))\}$  is a subset of  $\{1, \dots, n\}$ . Moreover, for each  $\omega \in \mathcal{E}$ ,

$$(o, \mathbf{Y}_{j(1)}(\omega), \dots, \mathbf{Y}_{j(\#\mathcal{P}_n(\mathcal{E}))}(\omega), Z_n o)$$

is  $D_0$ -semi-aligned;

- (2) for each  $1 \leq k < l \leq n$ , we have

$$\mathbb{P}(\omega : \#(\mathcal{P}_n(\omega) \cap \{k, \dots, l\}) \leq \kappa(k-l)) \leq \frac{1}{\kappa} e^{-\kappa(k-l)}.$$

Let us now recall two facts:

**Lemma 3.9** ([Cho22a, Corollary 4.5]). *Let  $K_0, N_0 > 0$  and let  $S$  be a long enough  $K_0$ -Schottky set with cardinality  $N_0$ . Let  $\mathcal{E}$  be a pivotal equivalence class for  $S$  with  $\mathcal{P}(\mathcal{E}) = \{j(1) < j(2) < \dots\}$  and let  $x \in X$ . Then for each  $k \geq 1$  we have*

$$\mathbb{P}\left((x, \mathbf{Y}_{j(k)}(\omega), \mathbf{Y}_{j(k+1)}(\omega), \dots) \text{ is } D_0\text{-semi-aligned} \middle| \mathcal{E}\right) \geq 1 - (1/N_0)^k.$$

Moreover, for any  $m \geq 1$ ,  $n \geq j(m)$  and  $k = 1, \dots, m$ , we have

$$\mathbb{P}\left((\mathbf{Y}_{j(1)}(\omega), \dots, \mathbf{Y}_{j(m-k+1)}(\omega), Z_n(\omega)o) \text{ is } D_0\text{-semi-aligned} \middle| \mathcal{E}\right) \geq 1 - (1/N_0)^k.$$

**Lemma 3.10** ([Cho22a, Corollary 4.6]). *Let  $K_0, N_0 > 0$  and let  $S$  and  $\check{S}$  be long enough  $K_0$ -Schottky sets with cardinality  $N_0$ . Let  $\mathcal{E}$  be a pivotal equivalence class for  $S$  with  $\mathcal{P}(\mathcal{E}) = \{j(1) < j(2) < \dots\}$ , and let  $\check{\mathcal{E}}$  be a pivotal equivalence class for  $\check{S}$  with  $\mathcal{P}(\check{\mathcal{E}}) = \{\check{j}(1) < \check{j}(2) < \dots\}$ . Then we have*

$$\mathbb{P}\left((\bar{\mathbf{Y}}_{j(k)}(\omega), \mathbf{Y}_{\check{j}(k)}(\check{\omega})) \text{ is } D_0\text{-semi-aligned} \middle| \mathcal{E}\right) \geq 1 - (2/N_0)^k. \quad (\forall k > 0)$$

We are now ready to prove Proposition 3.1. Let us recall the statement:

**Proposition 3.11.** *Let  $0 < \epsilon \leq 1$  and let  $S \subseteq G^{M_0}$  be a long and large enough  $K_0$ -Schottky set. Then there exists  $\kappa = \kappa(\epsilon, S) > 0$  such that the following holds.*

*Let  $\{\mu_i : i \in \mathbb{Z}_{>0}\}$  be probability measures such that*

$$(\mu_{M_0 k+1} \times \dots \times \mu_{M_0 k+M_0})(s) > \epsilon. \quad (\forall s \in S, \forall k \in \mathbb{Z})$$

*Let  $g_i$  be distributed according to  $\mu_i$  for each  $i \in \mathbb{Z}$ , all independently, and let  $Z_i, \mathbf{Y}_i$  be as in Notation 2.6 and Display 3.5.*

*Then for arbitrary  $M$  integers  $t(1), \dots, t(M)$  and for arbitrary  $g_{t(1)}, \dots, g_{t(M)} \in G$ , we have*

$$\mathbb{P}\left(\begin{array}{l} \exists i(1), i(2), \dots, i(\kappa n) \in \{M_0, \dots, n\} \text{ such that} \\ \mathbf{Y}_{j(1)}, \dots, \mathbf{Y}_{i(\kappa n)} \text{ are Schottky axes and} \\ (Z_{-m}o, \mathbf{Y}_{i(1)}, \dots, \mathbf{Y}_{i(\kappa n)}, Z_{n+l}o) \text{ is } D_0\text{-semi-aligned for all } m, l \geq 0 \end{array} \middle| g_{t(1)}, g_{t(2)}, \dots, g_{t(M)}\right) \geq 1 - \frac{1}{\kappa} e^{-\kappa n - M}.$$

*Proof.* A very similar proof was given in [Cho22a, Lemma 4.10] so we only sketch the idea. Given  $M$  integer  $t(1), \dots, t(M)$ , we replace  $\mu_{t(1)}, \dots, \mu_{t(M)}$  with Dirac measures concentrated on the given elements, respectively. This will not affect the qualification for Proposition 3.7.

We employ the notation

$$\check{g}_i := g_{-i+1}^{-1}, \check{\mu}_i(\cdot) := \mu_{-i+1}(\cdot^{-1}), \check{S} := \{(s_{M_0}^{-1}, \dots, s_1^{-1}) : (s_1, \dots, s_{M_0}) \in S\}.$$

We then construct the probability space  $\Omega$  for  $\mu_1 \times \mu_2 \times \dots$  and  $\check{\Omega}$  for  $\check{\mu}_1 \times \check{\mu}_2 \times \dots$ , equipped with partitions  $\mathcal{Q}$  and  $\check{\mathcal{Q}}$  into pivotal equivalence classes as in Proposition 3.7. Let  $\kappa_1$  be coefficient of the exponential decay rate coming from Proposition 3.7.

Then  $(\check{\Omega}, \Omega)$  comes with independent RVs  $g_i$ 's and  $\check{g}_i$ 's, each distributed according to  $\mu_i$ 's and  $\check{\mu}_i$ 's, respectively. We define

$$Z_i := g_1 \cdots g_i, \quad \check{Z}_i := \check{g}_1 \cdots \check{g}_i, \quad \mathbf{Y}_i(\omega) := (Z_{i-M_0}, \dots, Z_i), \quad \mathbf{Y}_i(\check{\omega}) := (\check{Z}_{i-M_0}, \dots, \check{Z}_i).$$

It is then immediate that  $(\dots, \check{Z}_2, \check{Z}_1, id, Z_1, Z_2, \dots)$  has the same law as  $(\dots, Z_{-2}, Z_{-1}, Z_0, Z_1, Z_2, \dots)$ .

Let

$$\mathcal{A}_n := \{\omega \in \Omega : \#(\mathcal{P}(\omega) \cap \{1, \dots, k\}) \geq \kappa k - M \text{ for all } k \geq n\},$$

$$\check{\mathcal{A}}_n := \{\check{\omega} \in \check{\Omega} : \#(\check{\mathcal{P}}(\check{\omega}) \cap \{1, \dots, k\}) \geq \kappa k - M \text{ for all } k \geq n\}.$$

These sets are unions of some pivotal equivalence classes and have probability at least  $1 - K_1 e^{-\kappa n - M}$  for some  $K_1$  depending only on  $\kappa$ .

Next, we pick pivotal equivalence classes  $\mathcal{E} \in \mathcal{Q}$  and  $\check{\mathcal{E}} \in \check{\mathcal{Q}}$  from  $\mathcal{A}_n$  and  $\check{\mathcal{A}}_n$ , respectively. That means, the sets of pivotal times  $\mathcal{P}(\mathcal{E}) = \{j(1) < j(2) < \dots\}$  and  $\check{\mathcal{P}}(\check{\mathcal{E}}) = \{\check{j}(1) < \check{j}(2) < \dots\}$  are determined. On  $\mathcal{E} \times \check{\mathcal{E}}$ , we estimate the conditional probability for the following event:

(1) for  $x \in \{o, \check{Z}_1 o, \dots, \check{Z}_n o\}$ , the following sequence is  $D_0$ -semi-aligned:

$$(x, \mathbf{Y}_{j(\lceil \kappa_1 n/3 \rceil)}, \mathbf{Y}_{j(\lceil \kappa_1 n/3 \rceil + 1)}, \dots);$$

(2) for each  $k \geq n$ , the following are  $D_0$ -semi-aligned:

$$(o, \mathbf{Y}_{j(1)}(\omega), \mathbf{Y}_{j(2)}(\omega), \dots, \mathbf{Y}_{j(\lceil 2\kappa_1 k/3 \rceil)}(\omega), Z_k o), \\ (o, \mathbf{Y}_{\check{j}(1)}(\check{\omega}), \mathbf{Y}_{\check{j}(2)}(\check{\omega}), \dots, \mathbf{Y}_{\check{j}(\lceil 2\kappa_1 k/3 \rceil)}(\check{\omega}), \check{Z}_k o);$$

(3)  $(\bar{\mathbf{Y}}_{\check{j}(i)}(\check{\omega}), \mathbf{Y}_{j(i)}(\omega))$  is  $D_0$ -aligned for some  $i \leq n/3K$ .

Note that for we have  $j(\lceil \kappa_1 k \rceil - M) \leq k$  and  $\check{j}(\lceil \kappa_1 k \rceil - M) \leq k$  for each  $k \geq n$ . Lemma 3.9 and 3.10 tells us the above event happens for probability at least

$$1 - n(1/N_0)^{\kappa_1 n/3 - M} - 2 \sum_{k \geq n} (1/N_0)^{\kappa_1 k/3 - M} - (2/N_0)^{\kappa_1 n/3} \geq 1 - K(1/100)^{\kappa_1 n/3 - M}$$

for some suitable  $K > 0$  not depending on  $n$ . Here, we used the fact that  $N_0 \geq 400$ .

Now suppose that  $(\omega, \check{\omega})$  belongs to the above event. For  $m \in \{0, \dots, n\}$  and  $l \geq 0$ , we observe that

$$(\check{Z}_m o, \mathbf{Y}_{j(\lceil \kappa_1 n/3 \rceil)}, \dots, \mathbf{Y}_{j(\lceil 2\kappa_1(n+l)/3 \rceil)}(\omega), Z_{n+l} o)$$

is  $D_0$ -semi-aligned. For  $m \geq n$  and  $l \geq 0$ , we observe that

$$(\check{Z}_m o, \bar{\mathbf{Y}}_{\check{j}(\lceil 2\kappa_1 m/3 \rceil)}(\check{\omega}), \dots, \bar{\mathbf{Y}}_{\check{j}(i+1)}(\check{\omega}), \bar{\mathbf{Y}}_{\check{j}(i)}(\check{\omega}), \mathbf{Y}_{j(i)}(\omega), \dots, \mathbf{Y}_{j(\lceil n/3K \rceil)}(\omega), \dots, \mathbf{Y}_{j(\lceil 2\kappa_1(n+l)/3 \rceil)}(\omega), Z_{n+l} o)$$

is  $D_0$ -semi-aligned. In either case,  $(\check{Z}_m o, \mathbf{Y}_{j(\lceil n/3K \rceil)}, \dots, \mathbf{Y}_{j(\lceil 2n/3K \rceil)}(\omega), Z_{n+l} o)$  is a subsequence of a  $D_0$ -semi-aligned sequence, so it is itself a  $D_0$ -semi-aligned. Hence, we can choose

$$\{i(1), \dots, i(\lfloor n/3K \rfloor)\} := \{j(\lceil n/3K \rceil), \dots, j(\lceil 2n/3K \rceil)\} \subseteq M_0 \mathbb{Z} \cap \{M_0 \leq i \leq n\}.$$

Hence, the desired event holds for  $\mathbb{P}(\cdot | \mathcal{E} \times \check{\mathcal{E}})$ -probability at least  $1 - K(1/100)^{\kappa_1 n/3 - M}$ . The proof ends by averaging these conditional probabilities across  $\mathcal{A}_n \times \check{\mathcal{A}}_n$  and multiplying  $\mathbb{P}(\mathcal{A}_n \times \check{\mathcal{A}}_n) \geq 1 - K_1 e^{-\kappa n - M}$ .  $\square$

#### 4. DIFFERENTIABILITY OF THE ESCAPE RATE

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

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

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

**Lemma 4.1.** *Let  $\mu$  be a non-elementary probability measure on  $G$ . Then there exists  $K, \epsilon > 0$  such that*

$$\mathbb{P}_{\mu_1 \times \cdots \times \mu_k} \left( (g_1, \dots, g_k) : \begin{array}{l} \text{there exists } 0 \leq l < k \text{ such that} \\ d(Z_l o, [g^{-1} o, Z_k o]) < K \end{array} \middle| g, g_j, g_k \right) \geq 1 - K e^{-k/K}$$

for all  $1 \leq j < k$ , 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.* Let  $\mu$  be a non-elementary probability measure. Let  $N_0 = 400$ , let  $K_0 = K(N_0)$  be as in Proposition 2.13, let  $D_0, E_0, L_0$  be the ones determined as in Definition 2.12, and let  $M_0 = n(N_0, 2L_0)$  be as in Proposition 2.13. Then Proposition 2.13 guarantees a  $K_0$ -Schottky set in  $G^{M_0}$  such that  $\#S = 400$  and  $\mu^{M_0}(s) > 0$  for each  $s \in S$ . This  $S$  is clearly large and long enough.

Let  $\epsilon = \frac{1}{2} \min_{s \in S} \mu^{M_0}(s)$ . Now pick any Schottky sequence  $s = (s_1, \dots, s_{M_0}) \in S$ . We claim that  $\mu(s_i) > 2\epsilon$  for each  $i$ . Indeed,  $\mu(g) \leq 1$  for each  $g \in G$  and

$$\mu(s_i) \geq \mu(s_i) \cdot \mu(s_1) \cdots \mu(s_{i-1}) \cdot \mu(s_{i+1}) \cdots \mu(s_{M_0}) = \mu^{M_0}(s) \geq 2\epsilon.$$

We now fix  $\kappa = \kappa(2^{-M_0}\epsilon, S) > 0$  using Proposition 3.1.

Now consider a sequence of arbitrary measures  $\mu_i$ 's such that  $\|\mu_i - \mu\|_0 < \epsilon$  for each  $i$ . In particular,  $\mu_i(g) \geq \mu(g) - \epsilon$  for each  $g \in G$  and for each  $i$ . Then for any  $s = (s_1, \dots, s_i) \in S$  and for any integers  $i(1), \dots, i(M_0)$  we have

$$\prod_l \mu_{i(l)}(s_i) \geq \prod_i 0.5\mu(s_i) \geq 2^{-M_0} \cdot \mu^{M_0}(s).$$

In particular,  $(\mu_{k+1} \times \dots \times \mu_{k+M_0})(s) > 2^{-M_0}\epsilon$  for each  $s \in S$  and for each  $k$ . Now Proposition 3.1 tells us that

$$\mathbb{P}_{(g_i)_{i \in \mathbb{Z}} \sim \prod_{i \in \mathbb{Z}} \mu_i} \left( \exists i \in \{0, \dots, k-1\} \text{ such that } d(Z_i o, [Z_{-1} o, Z_k o]) \leq E_0 \mid g_{-1}, g_j, g_k \right) \geq 1 - \frac{1}{\kappa} e^{-\kappa(k-1)-3},$$

regardless of  $j$ .  $\square$

**Corollary 4.2.** *Let  $\mu$  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  $R_{n,m}(\mathbf{g}) = R'(\mathbf{g}) + R''(\mathbf{g})$ .

Let  $K, \epsilon > 0$  be as in Lemma 4.1 and set

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

By the previous lemma, we have

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

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

$$\begin{aligned} R'(\mathbf{g}) &= \|g_{-m} \cdots g_{-1} g_0\| + \|g_1 \cdots g_n\| - \|g_{-m} \cdots g_{-1} \cdot g_0 \cdot g_1 \cdots g_n\| \\ &\leq \|g_{-m} \cdots g_{-1} g_0\| + \|g_1 \cdots g_n\| - (\|g_{-m} \cdots g_{-1} g_0 g_1 \cdots g_l\| + \|g_l \cdots g_n\| - K) \\ &\leq \|g_1 \cdots g_l\|^{sym} + 2K \leq \sum_{i=1}^j \|g_i\|^{sym} + 2K. \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathbb{E}[R'] &\leq \mathbb{E} \left[ \sum_{i=1}^{\min\{j: \mathbf{g} \in E_j\}} \|g_i\|^{sym} + 2K \right] \leq \sum_{i=1}^m \mathbb{E} \left[ \|g_i\|^{sym} 1_{E_{i-1}^c} \right] + 2K \\ &\leq 2K + \sum_{i=1}^m \mathbb{E} \left[ \|g_i\|^{sym} \mathbb{P}(E_{i-1}^c | g_i) \right] \\ &\leq 2K + \sum_{i=0}^{m-1} K e^{-i/K} \|\mu_i\|_1 \leq 2K + \frac{K}{1 - e^{-1/K}} (\|\mu\|_1 + \epsilon). \end{aligned}$$



Similarly, if we define

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

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

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

Based on the same computation, we conclude

$$\mathbb{E}[R_{n,m}] = \mathbb{E}[R'] + \mathbb{E}[R''] \leq 4K + \frac{2K}{1 - e^{-1/K}} (\|\mu\|_1 + \epsilon). \quad \square$$

We now prove Theorem A.

*Proof.* Let  $\mu$  be a non-elementary probability measure with finite first moment, and let  $K_1, \epsilon > 0$  be the constants for  $\mu$  as in Corollary 4.2. Now consider a measure  $\mu'$  with  $\|\mu' - \mu\|_{0,1} < \epsilon$ . Then we have

$$\begin{aligned} (4.1) \quad L(\mu'^{*n}) - L(\mu^{*n}) &= \sum_{i=1}^n (\mathbb{E}_{\mu^{*(i-1)} \times \mu' \times \mu'^{*n-i}} \|g_1 g_2 g_3\| - \mathbb{E}_{\mu^{*(i-1)} \times \mu \times \mu'^{*n-i}} \|g_1 g_2 g_3\|) \\ &= \sum_{i=1}^n (\mathbb{E}_{\mu'} \|g\| + \mathbb{E}_{\mu^{*(i-1)}} \|g\| + \mathbb{E}_{\mu'^{*n-i}} \|g\| - \mathbb{E}_{\mu^{i-1} \otimes \mu' \otimes \mu'^{n-i}} [R_{n-i, i-1}(\mathbf{g})]) \\ &\quad - \sum_{i=1}^n (\mathbb{E}_{\mu} \|g\| + \mathbb{E}_{\mu^{*(i-1)}} \|g\| + \mathbb{E}_{\mu'^{*n-i}} \|g\| - \mathbb{E}_{\mu^{i-1} \otimes \mu \otimes \mu'^{n-i}} [R_{n-i, i-1}(\mathbf{g})]) \\ &= n \mathbb{E}_{\mu' - \mu} \|g\| + \sum_{i=1}^n \sum_{g \in G} (\mu'(g) - \mu(g)) \mathbb{E}_{\mu^{i-1} \otimes 1_{g_0=g} \otimes \mu'^{n-i}} [R_{n-i, i-1}(\mathbf{g})]. \end{aligned}$$

Here,  $\mathbb{E}_{\mu' - \mu} \|g\|$  is bounded by  $\|\mu' - \mu\|_1$ . Moreover, Corollary 4.2 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  $nK_1 \|\mu' - \mu\|_0$ . 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.

**Lemma 4.3.** *Let  $\mu, \eta$  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  $\eta$  is balanced, i.e.,  $\sum_{g \in G} \eta(g) = 0$ .*

**Lemma 4.4.** *Let  $\mu, \mu', \eta, \eta'$  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*

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

*Proof of Theorem B.* Again, let  $K_1, \epsilon > 0$  be the constants for  $\mu$  as described in Corollary 4.2.

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  $\mu_t$  (so that  $\|\mu_t - \mu\|_{0,1} < \epsilon$ ), 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  $\mu_t$  with  $\mu_t^{*k}$  and  $\mu$  with  $\mu^{*k}$ . 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  $K(\mu_t^{*k}, \mu^{*k}) \leq K(\mu_t, \mu) \leq K_1$ . 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

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

By Lemma 4.4, we have

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

and

$$(4.2) \quad \left| \limsup_{t \rightarrow 0} \frac{1}{t} [l(\mu_t) - l(\mu)] - \frac{1}{k} \sum_{i=1}^k \mathbb{E}_{\mu^{*(i-1)} * \eta * \mu^{*(k-i)}} \|g\| \right| \leq \limsup_{t \rightarrow 0} \left| \frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu^{*k}\|_0}{k} \right|.$$

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

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

Again, Lemma 4.4 implies

$$(4.3) \quad \begin{aligned} \limsup_{t \rightarrow 0} \left| \frac{K_1}{t} \cdot \frac{\|\mu_t^{*k} - \mu^{*k}\|_0}{k} \right| &= \limsup_{t \rightarrow 0} \frac{K_1}{kt} \cdot \left( t \left\| \sum_{i=1}^k \mu^{*(i-1)} * \eta * \mu^{*(k-i)} \right\|_0 + o(t) \right) \\ &= \frac{K_1}{k} \cdot \left\| \sum_{i=1}^k \mu^{*(i-1)} * \eta * \mu^{*(k-i)} \right\|_0. \end{aligned}$$

Let  $f(g) := \eta(g)/\mu(g)$  for  $g \in \text{supp } \mu$ . Then  $f$  is a  $\mu$ -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 4.2 and 4.3: under the assumptions on  $\mu, \eta, \{\mu_t\}$ , we have

$$(4.4) \quad \left| \frac{d}{dt} l(\mu_t) \right|_{t=0} - \frac{1}{k} \sum_{i=1}^k \mathbb{E}_{\mu^{*(i-1)} * \eta * \mu^{*(k-i)}} \|g\| \leq \frac{K_1}{k} \cdot \left\| \sum_{i=1}^k \mu^{*(i-1)} * \eta * \mu^{*(k-i)} \right\|_0$$

for each  $k$ . Here, the constant  $K_1$  depends on  $\mu$ , and can be kept the same even if we replace  $\mu$  with  $\mu'$  such that  $\|\mu' - \mu\| < \epsilon/2$ .

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

$$(4.5) \quad \frac{1}{k} \left\| \sum_{i=1}^k \mu^{*(i-1)} * \eta * \mu^{*(k-i)} \right\|_0 < \frac{\delta}{3K_1}.$$

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

$$(4.6) \quad \begin{aligned} & \frac{1}{k} \left\| \sum_{i=1}^k \mu'^{*(i-1)} * \eta' * \mu'^{*(k-i)} \right\|_0 < \frac{\delta}{2K_1}, \\ & \left| \frac{1}{k} \sum_{i=1}^k \mathbb{E}_{\mu'^{*(i-1)} * \eta' * \mu'^{*(k-i)}} \|g\| - \frac{1}{k} \sum_{i=1}^k \mathbb{E}_{\mu^{*(i-1)} * \eta * \mu^{*(k-i)}} \|g\| \right| < \frac{\delta}{2}. \end{aligned}$$

Now let  $\{\mu'_t\}_t$  be another family of probability measures such that  $\|\mu'_t - \mu'\|_{0,1} \rightarrow 0$ ,  $\frac{1}{t}\|\mu'_t - \mu' - t\eta'\|_{0,1} \rightarrow 0$ . By plugging Inequality 4.5 and 4.6 into Inequality 4.4, we conclude that  $\left| \frac{d}{dt} l(\mu_t) - \frac{d}{dt} l(\mu'_t) \right| < 2\delta$  as desired.  $\square$

## 5. SQUEEZING ISOMETRIES AND ESCAPE RATE

As we did in Lemma 4.1, we describe another simple geometric situation. Given that this situation happens with high probability, Theorem C can be deduced from elementary arguments.

**Definition 5.1.** Let  $(X, G, o)$  be as in Convention 2.4 and let  $\epsilon > 0$ . We define a collection  $\mathcal{A}_{k;\epsilon} \subseteq G^{\mathbb{Z}}$  of step paths as follows. A step path  $\mathbf{g} \in G^{\mathbb{Z}}$  is belongs to  $\mathcal{A}_{k;\epsilon}$  if there exist  $E > 0$  and an  $(E, \epsilon)$ -squeezing sequence of paths  $(\gamma_1, \dots, \gamma_N), (\eta_1, \dots, \eta_M)$  on  $X$  such that the following sequences are all  $E$ -aligned for every  $m \geq k$  and  $n \geq 0$ :

$$(Z_{-n}o, \gamma_1), (\gamma_N, Z_m o), (Z_{-(m+1)}o, \eta_1), (\eta_M, Z_{n-1}o).$$

We now state the main probabilistic estimate.

**Proposition 5.2.** Let  $(X, G, o)$  be as in Theorem C and let  $\mu$  be a strongly non-elementary probability measure on  $G$ . Then there exists  $\epsilon > 0$  such that for each  $\epsilon' > 0$ , the following holds for some  $K = K(\mu, \epsilon, \epsilon')$ .

For every sequence of probability measures  $\{\mu_i\}_{i \in \mathbb{Z}}$  with  $\|\mu_i - \mu\|_0 < \epsilon$ , 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

$$\mathbb{P}_{\prod_i \mu_i} (\mathbf{g} \in \mathcal{A}_{k;\epsilon'} \mid g_0, g_j) \geq 1 - K e^{-k/K}.$$

*Proof.* The condition for  $(Z_{-n}o, \gamma_1)$ ,  $(\gamma_N, Z_m o)$  and the condition for  $(Z_{-(m+1)}o, \eta_1)$ ,  $(\eta_M, Z_{n-1}o)$  are symmetric, so we will prove that the first condition has high probability.

Let  $\mu$  be a strongly non-elementary probability measure on  $G$ . This implies that  $\text{supp } \mu$  generates two independent BGIP elements  $g$  and  $h$ . Here, make sure to choose  $g$  as the one specified in the definition of strongly non-elementary measure, namely:

- If  $G$  is the mapping class group, we choose  $g$  to be a pseudo-Anosov (in fact, it must be so *a fortiori*).
- If  $G = \text{Out}(F_N)$ , we take  $g$  to be a principal fully irreducible outer automorphism.
- If  $G$  is a countable group of isometries of a CAT(-1) space, then we take  $g$  to be an arbitrary loxodromic element.
- If  $G$  is a countable group of automorphisms of a CAT(0) cube complex, then we take  $g$  to be an automorphism that skewers a pair of strongly separated hyperplanes.
- If  $G$  is the surface group or a free product of nontrivial groups, then we take  $g$  to be the element that is 0-constricting.

Let  $N_1, k, l$  be such that

$$\min(\mu^{*N_1}(g^k), \mu^{*N_1}(h^l)) := \epsilon_{\min} > 0.$$

We then choose  $\epsilon = \frac{1}{2}\epsilon_{\min}$ . Then for any probability measures  $\mu_i$ 's with  $\|\mu_i - \mu\|_0 < \epsilon$  for each  $i$ , we have

$$(5.1) \quad \begin{aligned} (\mu_{k+1} * \cdots * \mu_{k+N_1})(g) &\geq 2^{-N_1} \mu^{*N_1}(g) \geq 2^{-N_1} \epsilon_{\min}, \\ (\mu_{k+1} * \cdots * \mu_{k+N_1})(h) &\geq 2^{-N_1} \mu^{*N_1}(h) \geq 2^{-N_1} \epsilon_{\min}. \end{aligned}$$

Now, let  $K_0 > 0$  be the constant given by Proposition 2.16 for  $g^k, h^l$  and  $A = \{g^k, h^l\}$ , which determines  $D_0, E_0$  as well. Now let  $n$  be long enough that  $\Gamma_g := (o, g^k o, \dots, g^{kn} o)$  is an  $(E_0, \epsilon')$ -squeezing axis. We now construct a long and large enough  $K_0$ -Schottky set  $S \subseteq G^{M_0}$  as in Proposition 2.16 for  $n$ . Namely,  $\Gamma(s)$  for each  $s \in S$  has a subpath that is a translate of  $(o, g^k o, \dots, g^{kn} o)$ . Furthermore, Inequality 5.1 tells us that  $(\mu_{k+1} * \cdots * \mu_{k+M_0})(s) > 2^{-M_0 N_1} \epsilon_{\min}$  for each  $s \in S$ .

Now let  $\kappa = \kappa(2^{-M_0 N_1} \epsilon_{\min}, S)$  for this specific  $S$ , as in Proposition 3.1. Then we have

$$\mathbb{P}_{(g_i)_{i \in \mathbb{Z}} \sim \prod_{i \in \mathbb{Z}} \mu_i} \left( \begin{array}{c} \exists \text{ a } (S)\text{-Schottky axis } \gamma \text{ such that } (Z_{-m}o, \gamma, Z_{n+l}o) \text{ is } \\ D_0\text{-semi-aligned for each } m, l \geq 0 \end{array} \middle| g_0, g_j \right) \geq 1 - \frac{1}{\kappa} e^{-\kappa(k-1)-2},$$

regardless of  $j$ . Furthermore, when  $(Z_{-m}o, \gamma, Z_{n+l}o)$  is  $D_0$ -semi-aligned for some  $S$ -Schottky axis  $\gamma$ , then Proposition 2.10 tells us that  $\gamma$  is  $0.1E_0$ -fellow traveling with the subpath  $\varphi\Gamma_g$ , for some  $\varphi$ . In particular, we have:

$$(Z_{-m}, \varphi\Gamma_g, Z_{n+l}o) \text{ is } E_0\text{-aligned for each } m, l \geq 0.$$

This ends the proof.  $\square$

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

**Lemma 5.3.** *Let  $\mathbf{g} \in A_{k, \epsilon'}$ . Then for each  $n, m, n', m' \geq k$ , we have*

$$|R_{n, m}(\mathbf{g}) - R_{n', m'}(\mathbf{g})| < 2\epsilon'.$$

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

Since  $\mathbf{g} \in A_{k, \epsilon'}$ , there exists an  $(E, \epsilon')$ -squeezing sequence of  $(\gamma_1, \dots, \gamma_N)$  for some  $E$  such that

$$(Z_{-(m+1)}o, \gamma_1), (o =: Z_0 o, \gamma_1), (\gamma_N, Z_n o), (\gamma_N, Z_{n'} o)$$

are all  $E$ -aligned. Since  $(\gamma_1, \dots, \gamma_N)$  is  $(E, \epsilon')$ -squeezing, we conclude that

$$|d(Z_{-(m+1)}o, Z_n o) + d(o, Z_{n'} o) - d(o, Z_n o) - d(Z_{-(m+1)}o, Z_{n'} o)| < \epsilon'.$$

By adding and subtracting  $d(Z_{-(m+1)}o, Z_{-1}o) + d(o, Z_{-1}o)$ , the above display can be rewritten as

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

Note that for each  $\epsilon'$ , we have

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

It follows that  $\cap_{n>0} \cup_{k>0} A_{k;1/n}$  has measure 1. Moreover, Lemma 5.3 says that  $\{R_{n,m}(\mathbf{g})\}_{n,m}$  is Cauchy for each  $\mathbf{g} \in \cap_{n>0} \cup_{k>0} A_{k;1/n}$ . Hence, we have:

**Corollary 5.4.** *In the setting of Proposition 5.2,*

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

*is well-defined almost everywhere.*

Note that in Corollary 5.4,  $R_{n,m}(\mathbf{g})$ 's are well-defined for each  $n, m$  and is finite everywhere. This does not guarantee that  $R_{n,m}(\mathbf{g})$  is integrable. Still,  $R(\mathbf{g}) = \lim_{n,m} R_{n,m}(\mathbf{g})$  is defined as a finite value almost everywhere.

If  $\mu$  is moreover assumed to have finite first moment and if  $\mu_i$ 's are suitably close to  $\mu$  in  $\|\cdot\|_{0,1}$ -norm, then we also have the  $L^1$ -convergence of  $R_{n,m}$ .

**Lemma 5.5.** *Let  $(X, G)$  be as in Theorem C, let  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite first moment and let  $\epsilon > 0$  be as in Proposition 5.2. Then for each  $\epsilon' > 0$  there exists  $N > 0$  such that, for every sequence of probability measures  $\{\mu_i\}_i$  with  $\|\mu_i - \mu\|_{0,1} < \epsilon$  for each  $i$ , we have*

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

*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

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

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

$$(5.2) \quad |R_{n',m'}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \leq 2 \sum_{i=n+1}^{n'} \|g_i\| + 2 \sum_{i=m+1}^{m'} \|g_{-i}\|.$$

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

$$(5.3) \quad |R(\mathbf{g}) - R_{n,m}(\mathbf{g})| \leq 10\eta + 2 \sum_{1 \leq |i| \leq N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{N-1;\eta}} + 2 \sum_{|i| > N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}.$$

To prove this, let  $k = \sup\{i : \mathbf{g} \notin A_{i-1;\eta}\}$ . Lemma 5.3 tells us that  $|R(\mathbf{g}) - R_{n',m'}(\mathbf{g})| \leq 10\eta$  for any  $n', m' \geq k$ . In particular, Inequality 5.3 holds when  $n, m \geq k$ .

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

$$\begin{aligned} |R(\mathbf{g}) - R_{n,m}(\mathbf{g})| &\leq |R(\mathbf{g}) - R_{k,k}(\mathbf{g})| + |R_{k,k}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \\ &\leq 10\eta + 2 \sum_{1 \leq |i| \leq k} \|g_i\| & (\because \text{Inequality 5.2}) \\ &\leq 10\eta + 2 \sum_{1 \leq |i| \leq N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{N-1;\eta}} + 2 \sum_{|i| > N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}. & (\because N < n < k) \end{aligned}$$

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

$$\begin{aligned}
|R(\mathbf{g}) - R_{n,m}(\mathbf{g})| &\leq |R(\mathbf{g}) - R_{k,m}(\mathbf{g})| + |R_{k,m}(\mathbf{g}) - R_{n,m}(\mathbf{g})| \\
&\leq 10\eta + 2 \sum_{i=n+1}^k \|g_i\| & (\because \text{Inequality 5.2}) \\
&\leq 10\eta + 2 \sum_{|i|>N} \|g_i\| \cdot 1_{\mathbf{g} \notin A_{|i|-1;\eta}}. & (\because N < n < k)
\end{aligned}$$

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

We now estimate the expectation of the RHS of Inequality 5.3. For each  $j \in \mathbb{Z}$  and  $k > 0$  we have

$$\begin{aligned}
\mathbb{E}_{\prod_{i=1}^{\infty} \mu_i \otimes 1_{g_0=g} \otimes \prod_{i=1}^{\infty} \mu_{-i}} [\|g_j\| \cdot 1_{\mathbf{g} \notin A_{k;\eta}}] &= \mathbb{E}_{\mu_j} [\|g_j\| \mathbb{P}(A_{k;\eta}^c | g_0, g_j)] \\
&\leq \mathbb{E}_{\mu_j} \|g_j\| \cdot K e^{-k/K} & (\because \text{Proposition 5.2}) \\
&\leq K(1 + \epsilon) \|\mu\|_1 \cdot e^{-k/K}.
\end{aligned}$$

Hence, we have

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

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

**Lemma 5.6.** *Let  $(X, G)$  be a metric space and a countable group of its isometries, let  $\mu_1, \mu_2, \mu_3$  be signed measures on  $G$  with finite  $\|\cdot\|_{0,1}$ -norm and let  $\{\mu_{i,j}\}_{i>0}$  be a sequence of signed measures converging to  $\mu_j$  in  $\|\cdot\|_{0,1}$ -topology, for  $j = 1, 2, 3$ . Then  $\mu_{i,1} * \mu_{i,2}$  also converges to  $\mu_1 * \mu_2$  in  $\|\cdot\|_{0,1}$ -topology, and*

$$\lim_{i \rightarrow +\infty} \mathbb{E}_{g_{-1} \sim \mu_{i,1}, g_0 \sim \mu_{i,2}, g_1 \sim \mu_{i,3}} R_{1,1}(\mathbf{g}) = \mathbb{E}_{g_{-1} \sim \mu_1, g_0 \sim \mu_2, g_1 \sim \mu_3} R_{1,1}(\mathbf{g}).$$

Recall that  $R_{1,1}(\mathbf{g}) := \|g_{-1}\| + \|g_0\| + \|g_1\| - \|g_{-1}g_0g_1\|$ .

*Proof.* For the first claim, observe that

$$\begin{aligned}
\sum_{g \in G} |(\mu_1 * \mu_2)(g) - (\mu_{i,1} * \mu_{i,2})(g)| \cdot d(o, go) &\leq \sum_{g \in G} \left| \sum_{h \in G} (\mu_1(h) \mu_2(h^{-1}g) - \mu_{i,1}(h) \mu_{i,2}(h^{-1}g)) \right| \cdot d(o, go) \\
&\leq \sum_{h' \in G} \sum_{h \in G} |\mu_1(h) \mu_2(h') - \mu_{i,1}(h) \mu_{i,2}(h')| \cdot d(o, hh'o) \\
&\leq \sum_{h \in G} \sum_{h' \in G} |\mu_1(h) - \mu_{i,1}(h)| |\mu_2(h')| (d(o, ho) + d(ho, hh'o)) \\
&\quad + \sum_{h \in G} \sum_{h' \in G} |\mu_{i,1}(h)| |\mu_2(h') - \mu_{i,2}(h')| (d(o, ho) + d(ho, hh'o)) \\
&\leq \|\mu_1 - \mu_{i,1}\|_1 \|\mu_2\|_0 + \|\mu_1 - \mu_{i,1}\|_0 \|\mu_2\|_1 \\
&\quad + \|\mu_{i,1}\|_1 \|\mu_2 - \mu_{i,2}\|_0 + \|\mu_{i,1}\|_0 \|\mu_2 - \mu_{i,2}\|_1.
\end{aligned}$$

The RHS of the above inequality converges to 0 as  $i$  tends to infinity. This gives the  $\|\cdot\|_1$ -convergence of  $\mu_{i,1} * \mu_{i,2}$ . By an analogous argument without the distance factor and with  $\|\cdot\|_0$ -norm instead of  $\|\cdot\|_1$ -norm, we also obtain the  $\|\cdot\|_0$ -convergence. The first claim now follows.

For the second claim, an analogous argument can be applied:

$$\begin{aligned}
& \left| \mathbb{E}_{g_{-1} \sim \mu_{i,1}, g_0 \sim \mu_{i,2}, g_1 \sim \mu_{i,3}} R_{1,1}(\mathbf{g}) - \mathbb{E}_{g_{-1} \sim \mu_1, g_0 \sim \mu_2, g_1 \sim \mu_3} R_{1,1}(\mathbf{g}) \right| \\
& \leq \sum_{g_{-1}, g_0, g_1 \in G} \left| \mu_1(g_{-1})\mu_2(g_0)\mu_3(g_1) - \mu_{i,1}(g_{-1})\mu_{i,2}(g_0)\mu_{i,3}(g_1) \right| \cdot R_{1,1}(\mathbf{g}) \\
& \leq \sum_{g_{-1}, g_0, g_1 \in G} \left| \mu_1(g_{-1})\mu_2(g_0)\mu_3(g_1) - \mu_{i,1}(g_{-1})\mu_{i,2}(g_0)\mu_{i,3}(g_1) \right| \cdot (\|g_{-1}\| + \|g_0\| + \|g_1\|) \\
& \leq \|\mu_1 - \mu_{i,1}\|_1 \|\mu_2\|_0 \|\mu_3\|_0 + \|\mu_1 - \mu_{i,1}\|_0 (\|\mu_2\|_1 \|\mu_3\|_0 + \|\mu_2\|_0 \|\mu_3\|_1) \\
& + \|\mu_{i,1}\|_0 \|\mu_{i,2} - \mu_2\|_1 \|\mu_3\|_0 + \|\mu_{i,2} - \mu_2\|_0 (\|\mu_{i,1}\|_1 \|\mu_3\|_0 + \|\mu_{i,1}\|_0 \|\mu_3\|_1) \\
& + \|\mu_{i,1}\|_0 \|\mu_{i,2}\|_0 \|\mu_{i,3} - \mu_3\|_1 + \|\mu_{i,3} - \mu_3\|_0 (\|\mu_{i,1}\|_1 \|\mu_{i,2}\|_0 + \|\mu_{i,1}\|_0 \|\mu_{i,2}\|_1).
\end{aligned}$$

The RHS converges to 0 as  $i$  tends to infinity. This ends the proof.  $\square$

We now prove Theorem C that we state again:

**Theorem 5.7.** *Let  $(X, G)$  be as in Theorem C, let  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite first moment, let  $\eta$  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  $\mu$  and  $\eta$ .*

*Proof of Theorem 5.7.* Let  $\epsilon > 0$  be as in Proposition 5.2. Without loss of generality we suppose  $\|\mu_t - \mu\|_{0,1} \leq \epsilon$  for  $t \in [-1, 1]$ .

Recall first that Corollary 4.2 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 4.1 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

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

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  $2K_1\mu(g) + \mu_t(g)$ , which is integrable:  $\sum_{g \in G} 2K_1(\mu(g) + \mu_t(g)) = 2K_1 < +\infty$ . Finally, for all  $u \in (0, 1)$  and for all  $g \in G$ , we have

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

by Lemma 5.5. Hence, by the dominated convergence theorem, we deduce that

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

We now divide this by  $t$  and claim

$$(5.4) \quad \lim_{t \rightarrow 0} \frac{1}{t} (l(\mu_t) - l(\mu)) = \mathbb{E}_\eta \|g\| - \mathbb{E}_{\prod_{i=1}^\infty \mu \otimes \eta \otimes \prod_{i=1}^\infty \mu} R(\mathbf{g}).$$

First, we have

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

Next, note that

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

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  $t \rightarrow 0$ . The second summation is over  $G$  and the summand is bounded by  $2K_1|\eta(g)|$  for any  $t$ , which is integrable. Hence, we can conclude by DCT once we show for each  $g \in G$  that

$$(5.5) \quad \lim_{t \rightarrow 0} \mathbb{E}_{\prod_{i=1}^\infty \mu_t \otimes 1_{g_0=g} \otimes \prod_{i=1}^\infty \mu} R(\mathbf{g}) = \mathbb{E}_{\prod_{i=1}^\infty \mu \otimes 1_{g_0=g} \otimes \prod_{i=1}^\infty \mu} R(\mathbf{g}).$$

Let  $\epsilon' > 0$  and let  $N$  be as in Lemma 5.5. Then we have

$$\begin{aligned}
&\left| \mathbb{E}_{\prod_{i=1}^\infty \mu_t \otimes 1_{g_0=g} \otimes \prod_{i=1}^\infty \mu} R(\mathbf{g}) - \mathbb{E}_{\prod_{i=1}^N \mu_t \otimes 1_{g_0=g} \otimes \prod_{i=1}^N \mu} R(\mathbf{g}) \right| < \epsilon' \quad (t \in [-1, 1]), \\
&\left| \mathbb{E}_{\prod_{i=1}^\infty \mu \otimes 1_{g_0=g} \otimes \prod_{i=1}^\infty \mu} R(\mathbf{g}) - \mathbb{E}_{\prod_{i=1}^N \mu \otimes 1_{g_0=g} \otimes \prod_{i=1}^N \mu} R(\mathbf{g}) \right| < \epsilon', \\
&\lim_{t \rightarrow 0} \mathbb{E}_{\prod_{i=1}^N \mu_t \otimes 1_{g_0=g} \otimes \prod_{i=1}^N \mu} R(\mathbf{g}) = \mathbb{E}_{\prod_{i=1}^N \mu \otimes 1_{g_0=g} \otimes \prod_{i=1}^N \mu} R(\mathbf{g}).
\end{aligned}$$

Here, the third line holds by Lemma 5.6. Put together, these imply that the two expectations in Equation 5.5 differ by at most  $2\epsilon'$  eventually as  $t \rightarrow 0$ . We now send  $\epsilon' \rightarrow 0$  to establish Equation 5.5.

We now discuss the continuity of  $\sigma_1(\mu, \eta)$  with respect to  $\mu$  and  $\eta$ . For this let  $\mu_n, \eta_n$  be probability measures and signed measures, respectively, such that  $\|\mu_n - \mu\|_{0,1}, \|\eta_n - \eta\|_{0,1} \rightarrow 0$ . Without loss of generality we assume  $\|\mu_n - \mu\|_{0,1} < \epsilon$  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 5.5 did not require the differentiability of  $\{\mu_t\}_{t \in [-1,1]}$ . Namely, the exactly same argument implies that

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

holds for each  $g \in G$ . Moreover, Corollary 4.2 gives the bound  $D_n(g), D(g) \leq K_1$  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  $\|\eta 1_{G \setminus A}\|_{0,1} < \epsilon$ . We then have

$$\begin{aligned} |\sigma_1(\mu_n, \eta_n) - \sigma_1(\mu, \eta)| &= \left| \mathbb{E}_{\eta_n} d(o, go) - \mathbb{E}_{\eta} d(o, go) - \left( \sum_{g \in G} \eta_n(g) D_n(g) - \sum_{g \in G} \eta(g) D(g) \right) \right| \\ &\leq \|\eta_n - \eta\|_1 + \sum_{g \in A} |\eta_n(g) D_n(g) - \eta(g) D(g)| \\ &\quad + \sum_{g \notin A} |\eta_n(g) - \eta(g)| |D_n(g)| + \sum_{g \notin A} |\eta(g)| |D_n(g) - D(g)|. \end{aligned}$$

Here, the first summation is over a finite set, and it 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  $2K_1 \cdot \|\eta 1_{G \setminus A}\|_0 < 2K_1 \epsilon$ . We now bring  $\epsilon$  down to zero and conclude.  $\square$

## 6. EXPONENTIALLY SQUEEZING GEODESICS

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

**Convention 6.1.**  $(X, G)$  is either:

- a CAT(-1) space and a countable group of its isometries;
- a CAT(0) cube complex and a countable group of its isometries;
- 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.

**Proposition 6.2.** Let  $(X, d)$  be as in Convention 6.1 and let  $\mu$  be a strongly non-elementary probability measure on  $G$ . Then there exist  $\epsilon, K > 0$  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  $\{j_1, \dots, j_N\} \subseteq \mathbb{Z} \setminus \{0\}$ , we have

$$\mathbb{P}_{\prod_i \mu_i} (\mathbf{g} \in A_{k; e^{-k/K}} \mid g_0, g_{j_1}, g_{j_2}, \dots, g_{j_N}) \geq 1 - K e^{-k/K - N}.$$

*Proof.* We have to find an  $(E_0, e^{-k/K})$ -squeezing sequence  $(\gamma_1, \dots, \gamma_N)$  for the random path with high probability. Again, the condition for  $(Z_{-n}o, \gamma_1)$ ,  $(\gamma_N, Z_m o)$  and the condition for  $(Z_{-(m+1)}o, \eta_1)$ ,  $(\eta_M, Z_{n-1}o)$  are symmetric, so we will deal with the first condition.

Let  $\mu$  be a strongly non-elementary probability measure on  $G$ . This implies that  $\text{supp } \mu$  generates two independent BGIP elements  $g$  and  $h$ . Here, make sure to choose  $g$  as the one specified in the definition of strongly non-elementary measure, namely:

- If  $G$  is a countable isometry group of a CAT(-1) space, we take  $g$  a loxodromic element.
- If  $G$  is a countable automorphism of a CAT(0) cube complex, then we take  $g$  to be an automorphism that skewers a pair of strongly separated hyperplanes.
- If  $G$  is the surface group or a free product of nontrivial groups, then we take  $g$  to be the element that is 0-squeezing.

Let  $N_1, k, l$  be such that

$$\min(\mu^{*N_1}(g^k), \mu^{*N_1}(h^l)) := \epsilon_{\min} > 0.$$

We then choose  $\epsilon = \frac{1}{2}\epsilon_{\min}$ . Then for any probability measures  $\mu_i$ 's with  $\|\mu_i - \mu\|_0 < \epsilon$  for each  $i$ , we have

$$(6.1) \quad \begin{aligned} (\mu_{k+1} * \cdots * \mu_{k+N_1})(g) &\geq 2^{-N_1} \mu^{*N_1}(g) \geq 2^{-N_1} \epsilon_{\min}, \\ (\mu_{k+1} * \cdots * \mu_{k+N_1})(h) &\geq 2^{-N_1} \mu^{*N_1}(h) \geq 2^{-N_1} \epsilon_{\min}. \end{aligned}$$

Now, let  $K_0 > 0$  be the constant given by Proposition 2.16 for  $g^k, h^l$  and  $A = \{g^k, h^l\}$ , which determines  $D_0, E_0$  as well. Now let  $n$  be long enough that either:

- $d(o, g^n o) > 10E_0$  if  $G$  is a countable isometry group of a CAT(-1) space; or
- $\Gamma(g, \dots, g) = (o, go, \dots, g^n o)$  is  $(E_0, 0)$ -squeezing in the remaining case.

Here, for CAT(-1) space we are using Fact 2.29.

We now construct a long and large enough  $K_0$ -Schottky set  $S \subseteq G^{M_0}$  as in Proposition 2.16 for  $n$ . Namely,  $\Gamma(s)$  for each  $s \in S$  has a subpath that is a translate of  $(o, g^k o, \dots, g^{kn} o)$ . In this case, any  $D_0$ -semi-aligned sequence of length  $N$  of Schottky axes is  $(E_0, E_0 e^{-E_0})$ -squeezing.

Note that Inequality 6.1 tells us that  $(\mu_{k+1} * \cdots * \mu_{k+M_0})(s) > 2^{-M_0 N_1} \epsilon_{\min}$  for each  $s \in S$ .

Now let  $\kappa = \kappa(2^{-M_0 N_1} \epsilon_{\min}, S)$  for this specific  $S$ , as in Proposition 3.1. Then we have

$$\mathbb{P}_{(g_i)_{i \in \mathbb{Z}} \sim \prod_{i \in \mathbb{Z}} \mu_i} \left( \begin{array}{c} \exists (S\text{-})\text{Schottky axes } \gamma_1, \dots, \gamma_{\lfloor \kappa k \rfloor} \text{ such that} \\ (Z_{-m} o, \gamma_1, \dots, \gamma_{\lfloor \kappa k \rfloor}, Z_{n+l} o) \text{ is} \\ D_0\text{-semi-aligned for each } m, l \geq 0 \end{array} \middle| \begin{array}{c} g_0, g_{j_1}, \dots, g_{j_N} \end{array} \right) \geq 1 - \frac{1}{\kappa} e^{-\kappa(k-1)-N},$$

regardless of the values of  $j_1, \dots, j_N$ . Recall that our choice of  $S$  is such that  $D_0$ -semi-aligned sequences of Schottky axes of length  $\lfloor \kappa k \rfloor$  is  $(E_0, K e^{-k/K})$ -squeezing for some  $K > 0$ . This ends the proof.  $\square$

By Lemma 5.3, we have  $|R_{n,m}(\mathbf{g}) - R_{n',m'}(\mathbf{g})| < 2K e^{-k/K}$  for all  $n, m, n', m' \geq k$  if  $\mathbf{g} \in A_{k; e^{-k/K}}$ . Proposition 6.2 tells us that for whatever finite set of  $\{j_1, \dots, j_N\}$  and values of  $g_0, g_{j_1}, \dots, g_{j_N}$ , if we distribute  $g_i$  independently according to  $\mu_i$  for  $i \notin \{0, j_1, \dots, j_N\}$  as in the assumption of Proposition 6.2,  $\{R_{n,m}(\mathbf{g})\}_{n,m}$  is a.e. Cauchy. This remains true if we then equip  $g_0, g_{j_1}, \dots, g_{j_N}$  with arbitrary finite measures or finite signed measure. Hence, we can discuss the RV  $R(\mathbf{g})$  in the setting that some finitely many of  $g_i$ 's are defined on a signed measure space.

**Lemma 6.3.** *Let  $(X, d)$  be as in Convention 6.1, let  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite moment and let  $\epsilon, K > 0$  be as in Proposition 6.2.*

*Let  $\{j_1, \dots, j_N\}$  be  $N$  integers, and let  $\mu_0, \mu_{j_1}, \dots, \mu_{j_N}$  be probability measures with finite  $\|\cdot\|_1$ -norm. For  $i \in \mathbb{Z} \setminus \{0, j_1, \dots, j_N\}$ , let  $\mu_i$  be a probability measures such that  $\|\mu_i - \mu\|_{0,1} < \epsilon$ . Then we have*

$$\mathbb{E}_{\prod_i \mu_i} |R(\mathbf{g}) - R_{n,m}(\mathbf{g})| < K_2 e^{-\min(n,m)/K+N} \cdot \left(1 + \sum_{l=1}^N \|\eta\|_1\right)$$

for every  $n, m$ , where  $K_2$  is a constant that is determined by the values of  $\epsilon, K$  and  $\|\mu\|_1$ .

*Proof.* For notational convenience, let  $u = \min(n, m)$ . As in the proof of Lemma 5.5, 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

$$(6.2) \quad |R(\mathbf{g}) - R_{n,m}(\mathbf{g})| \leq 10e^{-u/K} + 2 \sum_{1 \leq |i| \leq u} \|g_i\| \mathbf{1}_{\mathbf{g} \notin B_{u-1}} + 2 \sum_{|i| > u} \|g_i\| \cdot \mathbf{1}_{\mathbf{g} \notin B_{|i|-1}}.$$

The proof is almost identical to the one in the proof of Lemma 5.5 so we omit it.

We now estimate the expectation of the RHS of Inequality 6.2. By Tonelli, it suffices to compute the expectation of each summand. Moreover, because the RVs  $\|g_i\| \mathbf{1}_{\mathbf{g} \notin B_{u-1}}$  and  $\|g_i\| \cdot \mathbf{1}_{\mathbf{g} \notin B_{|i|-1}}$  are

integrable, one can condition on  $g_0, g_{j_1}, \dots, g_{j_N}$ , compute the conditional expectation and average them.

For each  $j \in \mathbb{Z} \setminus \{0, j_1, \dots, j_N\}$ , for each  $k > 0$  and for each prescribed choice  $g_0, g_{j_1}, \dots, g_{j_N}$ , we have

$$\begin{aligned} \mathbb{E} [\|g_j\| \cdot 1_{\mathbf{g} \notin B_k} \mid g_0, g_{j_1}, \dots, g_{j_N}] &\leq \mathbb{E} \left[ \|g_j\| \mathbb{P} \left( \bigcup_{i \geq k} A_{i; e^{-1/K}}^c \mid g_0, g_{j_1}, \dots, g_{j_N}, g_j \right) \right] \\ &\leq \mathbb{E}_{\mu_j} \|g_j\| \cdot \frac{K}{1 - e^{-1/K}} e^{-k/K+N} \quad (\because \text{Proposition 6.2}) \\ &\leq \frac{K}{1 - e^{-1/K}} e^{-k/K+N} (1 + \epsilon) \|\mu\|_1. \end{aligned}$$

The expectation of  $\mathbb{E} [\|g_j\| \cdot 1_{\mathbf{g} \notin B_k} \mid g_0, g_{j_1}, \dots, g_{j_N}]$  for  $g_0 \sim \mu_0, g_{j_1} \sim \mu_{j_1}, \dots, g_{j_N} \sim \mu_{j_N}$  is still bounded by the same bound  $\frac{K}{1 - e^{-1/K}} e^{-k/K+N} (1 + \epsilon) \|\mu\|_1$ .

if  $j \in \{j_1, \dots, j_m\}$  and  $k > 0$ , we have

$$\begin{aligned} \mathbb{E} [\|g_j\| \cdot 1_{\mathbf{g} \notin B_k} \mid g_0, g_{j_1}, \dots, g_{j_N}] &\leq \mathbb{E} \left[ \|g_j\| \mathbb{P} \left( \bigcup_{i \geq k} A_{i; e^{-1/K}}^c \mid g_0, g_{j_1}, \dots, g_{j_N}, g_j \right) \right] \\ &\leq \|g_j\| \cdot \frac{K}{1 - e^{-1/K}} e^{-k/K+N} \quad (\because \text{Proposition 6.2}) \\ &\leq \frac{K}{1 - e^{-1/K}} e^{-k/K+N} \|g_j\|. \end{aligned}$$

The expectation of this bound for  $g_0 \sim \mu_0, g_{j_1} \sim \mu_{j_1}, \dots, g_{j_N} \sim \mu_{j_N}$  is bounded by  $\frac{K}{1 - e^{-1/K}} e^{-k/K+N} \|\mu_j\|_1$ .

Hence, we have

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

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

In fact, the constant  $\epsilon, K$  we obtained in Proposition 6.2 only depends on how much weight  $\mu^{*n}$  puts on independent contracting isometries  $g, h \in G$ .

Also, we can modify  $\mu_0, \mu_{j_1}, \dots, \mu_{j_N}$  by constant scaling or by taking linear combination, which will render them finite measures or signed measures, respectively. After this modification, an estimate analogous to the one in Lemma 6.3 holds, namely:

$$|\mathbb{E}_{\prod_i \mu_i} R(\mathbf{g}) - R_{n,m}(\mathbf{g})| < K_2 e^{-\min(n,m)/K+N} \cdot \left(1 + \sum_{l=1}^N \|\eta\|_1\right) \cdot \|\mu_0\|_0 \cdot \prod_{l=1}^N \|\mu_{j_N}\|_0.$$

We will now see that this is in fact similar to  $\mathbb{E}_{\prod_i \mu_i} R(\mathbf{g})$  itself.

**Corollary 6.4.** *Let  $(X, d)$  be as in Convention 6.1, let  $\mu$  be a strongly non-elementary probability measure on  $G$  with finite moment and let  $\epsilon, K > 0$  be as in Proposition 6.2. Let  $K_2$  be as in Lemma 6.3. Then for  $K_3 = K_2 e^{1/K}$ , the following holds.*

*Let  $\{j_1, \dots, j_N\}$  be  $N$  integers, and let  $\mu_0, \mu_{j_1}, \dots, \mu_{j_N}$  be signed measures with finite  $\|\cdot\|_{0,1}$ -norm. Suppose further that  $\mu_{j_1}$  is balanced, i.e.,  $\sum_{g \in G} \mu_{j_1}(g) = 0$ . For  $i \in \mathbb{Z} \setminus \{0, j_1, \dots, j_N\}$ , let*

$\mu_i$  be a probability measures such that  $\|\mu_i - \mu\|_{0,1} < \epsilon$ . Then we have

$$\left| \mathbb{E}_{\prod_i \mu_i} R(\mathbf{g}) \right| < K_3 e^{-|j_1|/K+N} \cdot \left(1 + \sum_{l=1}^N \|\eta\|_1\right) \cdot \|\mu_0\|_0 \cdot \prod_{l=1}^N \|\mu_{j_N}\|_0.$$

*Proof.* Under the assumption, we claim that  $\mathbb{E}_{\prod_i} R_{|j_1|-1, |j_1|-1}(\mathbf{g}) = 0$ . To see this, we note that  $R_{|j_1|-1, |j_1|-1}(\mathbf{g})$  does not depend on the value of  $g_{j_1}$ . If we denote the value of  $\mathbb{E}_{g_i \sim \mu_i: |i| < |j_1|} R_{|j_1|-1, |j_1|-1}(\mathbf{g})$  by  $E$ , then we observe

$$\mathbb{E}_{\prod_i} R_{|j_1|-1, |j_1|-1}(\mathbf{g}) = \mathbb{E}_{g_{j_1} \sim \mu_{j_1}} \mathbb{E}_{g_i \sim \mu_i: |i| < |j_1|} R_{|j_1|-1, |j_1|-1}(\mathbf{g}) = \sum_g \mu_{j_1}(g) \cdot E = 0.$$

Given this, Lemma 6.3 for  $n = m = |j_1| - 1$  implies that

$$\left| \mathbb{E}_{\prod_i \mu_i} R(\mathbf{g}) \right| < K_2 e^{-(|j_1|-1)/K+N} \cdot \left(1 + \sum_{l=1}^N \|\eta\|_1\right) \cdot \|\mu_0\|_0 \cdot \prod_{l=1}^N \|\mu_{j_N}\|_0. \quad \square$$

Let us observe a simple lemma:

**Lemma 6.5.** *Let  $\{F_i(t)\}_{t \in [-\epsilon, \epsilon]}$  be a family of functions for each  $i \in \mathbb{Z}$  and let  $(c_i)_{i \in \mathbb{Z}}$  be a summable sequence of positive numbers. Suppose that  $F_t^{(i)}$  is differentiable at  $t = 0$  for each  $i$  and*

$$|F'_i(0)| \leq c_i \quad (\forall i).$$

*Suppose also that there exists a function  $f(t)$  with  $\lim_{t \rightarrow 0} f(t)/t = 0$  such that*

$$|F_i(t) - F_i(0) - tF'_i(0)| \leq c_i f(t) \quad (\forall i, \forall t).$$

*Then  $F(t) := \sum_{i \in \mathbb{Z}} F_i(t)$  is differentiable at  $t = 0$ , with  $F'(0) = \sum_{i \in \mathbb{Z}} F'_i(0)$ .*

*Proof.* For any  $N$  and  $t$ , we have

$$\left| \frac{F(t) - F(0)}{t} - \frac{\sum_{|i| \leq N} F_i(t) - \sum_{|i| \leq N} F_i(0)}{t} \right| \leq \sum_{|i| > N} \frac{|F_i(t) - F_i(0)|}{t} \leq \sum_{|i| \geq N} |F'_i(0)| + \frac{f(t)}{t} \sum_{|i| \geq N} c_i.$$

We send  $t \rightarrow 0$  to conclude that  $\sum_{|i| \leq N} F'_i(0)$ ,  $\limsup_{t \rightarrow 0} \frac{F(t) - F(0)}{t}$  and  $\liminf_{t \rightarrow 0} \frac{F(t) - F(0)}{t}$  differ by at most  $2 \cdot \sum_{|i| \geq N} |F'_i(0)|$ . We now send  $N$  to infinity to conclude.  $\square$

We now prove Theorem D.

*Proof.* Let  $\mu$  be a non-elementary probability measure, and let  $\epsilon, K > 0$  be as in Proposition 6.2 for  $\mu$ . Let us now consider a family of probability measure  $\{\mu_t^{(0)}\}_{t \in [-\epsilon, \epsilon]}$ , families of signed measures  $\{\mu_t^{(i)} : i = 1, 2, \dots\}_{t \in [-\epsilon, \epsilon]}$  and functions  $\{f_i(t) : i = 0, 1, \dots\}_{t \in [-\epsilon, \epsilon]}$  such that

$$\|\mu_t^{(i)} - \mu^{(i)} - t\mu^{(i+1)}\|_{0,1} \leq f_i(t). \quad (\forall t, \forall i).$$

For convenience, let us assume  $f_1 \leq f_2 \leq \dots$ ; we can just take the maximum of the first  $i$  functions if needed. Also, let  $M_1 \leq M_2 \leq \dots$  be such that

$$M_i \geq \|\mu^{(j)}(t)\|_{0,1}, \quad M_i \geq f_j(t)/t$$

for each  $j \leq i$  and each  $t$ .

Let us now consider the set

$$\mathcal{I} := \{(a_i)_{i \in \mathbb{Z}} : a_i \in \mathbb{Z}_{>0} \text{ for each } i \text{ and } a_i = 0 \text{ for all but finitely many entries}\}.$$

For  $\mathbf{a} = (a_i)_{i \in \mathbb{Z}}$ , we let  $\|\mathbf{a}\| = \max\{|i| : a_i \neq 0\}$ ,  $\sigma(\mathbf{a}) = \sum_i a_i$  and  $\text{supp } \mathbf{a} = \{i : a_i > 0\}$ .

For this proof, we will use the notation  $\mathbb{E}_{\mathbf{a}, t}$  and  $\mathbb{E}_{\mathbf{a}, t; A}$  for the expectation of RVs with respect to a specific product measure. Namely,  $\mathbb{E}_{\mathbf{a}, t} F(\mathbf{g})$  to be the expectation of the RV  $F(\mathbf{g})$  when  $g_i$ 's

are all independent and  $g_i \sim \mu_t^{(a_i)}$ . For example, we know that  $R(\mathbf{g})$  is well-defined almost surely with respect to any  $\mathbb{P}_{\mathbf{a},t}(\cdot)$  so we can discuss  $\mathbb{E}_{\mathbf{a},t} R(\mathbf{g})$ .

To differentiate  $\mathbb{E}_{\mathbf{a},t} R(\mathbf{g})$ , we need another notation. For each finite subset  $A$  of  $\mathbb{Z}$ , we define  $\mathbb{P}_{\mathbf{a},t;A}$  to denote that  $g_i$ 's are independently distributed according to

$$g_i \sim \begin{cases} \mu_t^{(a_i)} & a_i = 0, i \in A \\ \mu_0^{(a_i)} & a_i = 0, i \notin A. \end{cases}$$

Finally, if we first distribute  $g_i$ 's in the above rules except for some steps  $g_j$ , we write  $\mathbb{P}_{\mathbf{a},t;A;g_j \sim \text{distribution}}$ .

Here,  $\mathbb{E}_{\mathbf{a},t;A} R(\mathbf{g})$  is meant to be a finite approximation of  $\mathbb{E}_{\mathbf{a},t} R(\mathbf{g})$ . More precisely, we claim:

**Claim 6.6.** *For any  $\mathbf{a}$  and  $t \in [-\epsilon, \epsilon]$ , we have*

$$\mathbb{E}_{\mathbf{a},t} R(\mathbf{g}) = \lim_{N \rightarrow \infty} \mathbb{E}_{\mathbf{a},t;[-N,N]} R(\mathbf{g})$$

*Proof.* Take an arbitrary  $\epsilon' > 0$ . Then by Lemma 6.3, there exists  $n$  such that

$$\mathbb{E}_{\mathbf{a},t} |R(\mathbf{g}) - R_{n,n}(\mathbf{g})| < \epsilon', \quad \mathbb{E}_{\mathbf{a},t;[-N,N]} |R(\mathbf{g}) - R_{n,n}(\mathbf{g})| < \epsilon' \quad (\forall N).$$

We now take an arbitrary  $N > n$ . Then  $R_{n,n}(\mathbf{g})$  has the same distribution for both  $\mathbb{P}_{\mathbf{a},t}$  and  $\mathbb{P}_{\mathbf{a},t;[-N,N]}$ . This implies that

$$|\mathbb{E}_{\mathbf{a},t} R(\mathbf{g}) - \mathbb{E}_{\mathbf{a},t;[-N,N]} R(\mathbf{g})| < \epsilon' \quad (\forall N > n).$$

Since such a threshold  $n(\epsilon')$  exists for each  $\epsilon'$ , we conclude the claim.  $\square$

From this claim, we observe that

$$\begin{aligned} \mathbb{E}_{\mathbf{a},t} R(\mathbf{g}) &= \mathbb{E}_{\mathbf{a},t;\{0\}} R(\mathbf{g}) \\ &+ \sum_{N=0}^{\infty} \left( \mathbb{E}_{\mathbf{a},t;[-N,N];g_{N+1} \sim \mu_{N+1}^{(a_{N+1})}(t) - \mu_{N+1}^{(a_{N+1})}} R(\mathbf{g}) + \mathbb{E}_{\mathbf{a},t;[-N,N+1];g_{-(N+1)} \sim \mu_{-(N+1)}^{(a_{-(N+1)})}(t) - \mu_{-(N+1)}^{(a_{-(N+1)})}} R(\mathbf{g}) \right). \end{aligned}$$

Next, note that

$$\begin{aligned} \mathbb{E}_{\mathbf{a},t;[-N,N];g_{N+1} \sim \mu_{N+1}^{(a_{N+1})}(t) - \mu_{N+1}^{(a_{N+1})}} R(\mathbf{g}) &= \mathbb{E}_{\mathbf{a},0;g_{N+1} \sim \mu_{N+1}^{(a_{N+1})}(t) - \mu_{N+1}^{(a_{N+1})}} R(\mathbf{g}) \\ &+ \sum_{k=-N}^N \mathbb{E}_{\mathbf{a},t;[-N,k-1];g_{N+1} \sim \mu_{N+1}^{(a_{N+1})}(t) - \mu_{N+1}^{(a_{N+1})}, g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}} R(\mathbf{g}), \end{aligned}$$

and similarly  $\mathbb{E}_{\mathbf{a},t;[-N,N+1];g_{-(N+1)} \sim \mu_{-(N+1)}^{(a_{-(N+1)})}(t) - \mu_{-(N+1)}^{(a_{-(N+1)})}} R(\mathbf{g})$  is equal to  $\mathbb{E}_{\mathbf{a},0;g_{-(N+1)} \sim \mu_{-(N+1)}^{(a_{-(N+1)})}(t) - \mu_{-(N+1)}^{(a_{-(N+1)})}} R(\mathbf{g})$  plus  $2N+1$  expectations of the form  $\mathbb{E}_{\mathbf{a},t;A;g_{-N-1} \sim \mu_{-N-1}^{(a_{-N-1})}(t) - \mu_{-N-1}^{(a_{-N-1})}, g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}} R(\mathbf{g})$ .

In summary, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{a},t} R(\mathbf{g}) &= \mathbb{E}_{\mathbf{a},0} R(\mathbf{g}) + \sum_{k \in \mathbb{Z}} (S_k(t) := \mathbb{E}_{\mathbf{a},0;g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}} R(\mathbf{g})) \\ &+ \sum_{k \neq l} (S_{k,l}(t) := \mathbb{E}_{\mathbf{a},t;A_k;g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}, g_l \sim \mu_l^{(a_l)}(t) - \mu_l^{(a_l)}} R(\mathbf{g})) \end{aligned}$$

for some choices of finite sets  $A_{k,l}$  for each  $\{(k,l) \in \mathbb{Z}^2 : k \neq l\}$ . In fact, the above is the rearrangement of the implied summation, we have not yet observed that the summands are absolutely summable. To check it, first note that  $\mu_k^{(a_k)}(t) - \mu_k^{(a_k)}$  is *always* a balanced signed measure. Hence,

we can apply Corollary 6.4 to observe

$$\begin{aligned}
S_k(t) &= \mathbb{E}_{g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}} \mathbb{E}_{g_i \sim \mu^{(a_i)} : i \in \text{supp } \mathbf{a} \setminus \{k\}} \cdot \mathbb{E}_{\mu^{\mathbf{z}}} [R(\mathbf{g}) \mid g_i : i \in k \cup \text{supp } \mathbf{a}] \\
&\leq (1 + \|\mu_k^{(a_k)}(t) - \mu_k^{(a_k)}\|_1 + \sum_{i \in \text{supp } \mathbf{a} \setminus \{k\}} \|\mu^{(a_i)}\|_1) \cdot \|\mu_k^{(a_k)}(t) - \mu_k^{(a_k)}\|_0 \\
&\quad \cdot \prod_{i \in \text{supp } \mathbf{a} \setminus \{k\}} \|\mu^{(a_i)}\|_0 \cdot K_3 e^{-\|\mathbf{a}^{(k)}\|/K + \sigma(\mathbf{a}) + 1} \\
&\leq \left(1 + \|\mu_k^{(a_k)}(t)\|_1 + \|\mu_k^{(a_k)}\| + \sum_{i \in \text{supp } \mathbf{a} \setminus \{k\}} M_{a_i}\right) \cdot \prod_{i \in \text{supp } \mathbf{a} \setminus \{k\}} M_{a_i} \\
&\quad \cdot K_3 e^{-\|\mathbf{a}^{(k)}\|/K + \sigma(\mathbf{a}) + 1} \cdot (f_{a_k}(t)/t + \|\mu_k^{(a_k+1)}\|_{0,1}) \cdot t \\
&\leq (\sigma(\mathbf{a}) + 2) M_{\sigma(\mathbf{a})+1} \cdot M_{\sigma(\mathbf{a})}^{\sigma(\mathbf{a})} \cdot K_3 e^{-\|\mathbf{a}^{(k)}\|/K + \sigma(\mathbf{a}) + 1} \cdot 2M_{\sigma(\mathbf{a})} t := C_k t.
\end{aligned}$$

Here,  $\mathbf{a}^{(k)}$  means the sequence made from  $\mathbf{a}$  after increasing  $a_k$  by 1. A similar computation shows that  $\mathbb{E}_{\mathbf{a}, t; A_k; g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)}, g_l \sim \mu_l^{(a_l)}(t) - \mu_l^{(a_l)}} R(\mathbf{g})$  is bounded by

$$(6.3) \quad S_{k,l}(t) \leq (\sigma(\mathbf{a}) + 4) M_{\sigma(\mathbf{a})} \cdot M_{\sigma(\mathbf{a})}^{\sigma(\mathbf{a})} \cdot K_3 e^{-\|\mathbf{a}^{(k,l)}\|/K + \sigma(\mathbf{a}) + 2} (2M_{\sigma(\mathbf{a})})^2 t^2 := C_{k,l} t^2.$$

Note that  $C_k$  and  $C_{k,l}$  are exponentially decreasing in  $k$  and  $k, l$ , respectively. Hence, they are absolutely summable. Also, Inequality 6.3 tells us that  $S_{k,l}(t)' = 0$  for each  $k, l$ . For each  $S_k$ , we observe

$$\begin{aligned}
&|S_k(t) - t \mathbb{E}_{\mathbf{a}, 0; g_k \sim \mu_k^{(a_k+1)}} R(\mathbf{g})| \\
&= \left| \mathbb{E}_{g_k \sim \mu_k^{(a_k)}(t) - \mu_k^{(a_k)} - t\mu_k^{(a_k+1)}} \mathbb{E}_{g_i \sim \mu^{(a_i)} : a_i \neq 0} \cdot \mathbb{E}_{\mu^{\mathbf{z}}} [R(\mathbf{g}) \mid g_i : i \in k \cup \text{supp } \mathbf{a}] \right| \\
&\leq (1 + f_{a_k+1}(t) + \sum_{i \in \text{supp } \mathbf{a} \setminus \{k\}} M_{a_i}) \cdot \prod_{i \in \text{supp } \mathbf{a} \setminus \{k\}} M_{a_i} \\
&\quad \cdot K_3 e^{-\|\mathbf{a}^{(k)}\|/K + \sigma(\mathbf{a}) + 1} \cdot f_{a_k+1}(t) \\
&\leq f_{\sigma(\mathbf{a})+1}(t) \cdot C_k.
\end{aligned}$$

We now apply Lemma 6.5 to conclude that:

**Claim 6.7.** *For any  $\mathbf{a} \in \mathcal{I}$ , we have*

$$(6.4) \quad \frac{d}{dt} \Big|_{t=0} \mathbb{E}_{\mathbf{a}, t} R(\mathbf{g}) = \sum_{k \in \mathbb{Z}} (S'_k(t)) = \sum_{k \in \mathbb{Z}} \mathbb{E}_{\mathbf{a}^{(k)}, 0} R(\mathbf{g}),$$

Furthermore, we have uniform bound for all  $t \in [-\epsilon, \epsilon]$ :

$$(6.5) \quad \left| \mathbb{E}_{\mathbf{a}, t} R(\mathbf{g}) - \mathbb{E}_{\mathbf{a}, 0} R(\mathbf{g}) - t \sum_{k \in \mathbb{Z}} \mathbb{E}_{\mathbf{a}^{(k)}, 0} R(\mathbf{g}) \right| \leq \sum_k C_k f_{\sigma(\mathbf{a})+1}(t) + \sum_{k,l} C_{k,l} t^2,$$

where  $C_k \lesssim_{\|\mathbf{a}\|} e^{-\sigma(\mathbf{a})^{(k)}}$ , and  $C_{k,l} \lesssim_{\|\mathbf{a}\|} e^{-\sigma(\mathbf{a})^{(k,l)}}$ .

It is now time to sum up the discussion. Recall that  $\sigma_1(\mu, \eta) = \mathbb{E}_{\mu^{(1)}}(g) - \mathbb{E}_{\mathbf{a}} R(\mathbf{g})$  for  $\mathbf{a}$  such that  $a_i = 1$  only at  $i = 0$  and  $a_i = 0$  otherwise. Here, given the setting of the theorem, it is straightforward to see that  $\frac{d}{dt} \Big|_{t=0} \mathbb{E}_{\mu_t^{(k)}}(g) = \mathbb{E}_{\mu_t^{(k+1)}}(g)$ . So it remains to examine the differentiability of  $\mathbb{E}_{\mathbf{a}} R(\mathbf{g})$ . We observe that for each  $n = 0, 1, \dots$ ,

$$\frac{d}{dt} \Big|_{t=0} \sum_{k_1, \dots, k_n \in \mathbb{Z}} \mathbb{E}_{\mathbf{a}^{(k_1, \dots, k_n)}, t} R(\mathbf{g}) = \sum_{k_1, \dots, k_n, k_{n+1} \in \mathbb{Z}} \mathbb{E}_{\mathbf{a}^{(k_1, \dots, k_{n+1})}, t} R(\mathbf{g}).$$

First, the summand-wise derivative is given in Display 6.4, and the uniform linear approximation is given in Display 6.5. Now Lemma 6.5 implies the claim. This ends the proof.  $\square$

## 7. CONTINUITY OF THE ENTROPY

In this section, we prove Theorem E. Our strategy is as follows. Let  $\mu$  be a non-elementary probability measure  $\mu$  on  $G$  with finite (time-one) entropy, and suppose that  $\|\mu_i - \mu\|_0 \rightarrow 0$  and  $|H(\mu_i) - H(\mu)| \rightarrow 0$  holds. Let  $\epsilon > 0$  be arbitrary.

- (1) For a given non-elementary probability measure  $\mu$  on  $G$  with finite (time-one) entropy, we show that there exists  $n = N(\mu)$  such that  $\frac{1}{n}H(\mu^{*n})$  and  $h(\mu)$  differ by at most  $\epsilon$  for all  $n \geq N$ .
- (2) We also show that there exists a small  $\epsilon_{meas} > 0$  (depending on the choice of  $\mu$ ) such that, whenever

$$\|\mu_i - \mu\|_0 < \epsilon_{meas} \text{ and } |H(\mu_i) - H(\mu)| < \epsilon_{meas},$$

the same estimate holds for  $\mu_i$ , namely,  $\frac{1}{n}H(\mu_i^{*n})$  and  $h(\mu)$  differ by at most  $\epsilon$  for all  $n \geq N$ .

- (3) Finally, we observe that  $H(\mu_i^{*n}) \rightarrow H(\mu^{*n})$  for  $n = N$ .

This argument will show that  $\liminf_i h(\mu_i)$  and  $\limsup_i h(\mu_i)$  will differ from  $h(\mu)$  by at most  $2\epsilon$ . Since  $\epsilon$  is arbitrary, we conclude that  $\lim_i h(\mu_i) = h(\mu)$ .

The third step of the above is classical (cf. [AAV13, Lemma 3.5]). Hence, the key is to pick  $N(\mu)$  that works for not only  $\mu$  but all  $\mu_i$  that are sufficiently close to  $\mu$ .

**7.1. Schottky sets for WPD elements.** In this subsection, we construct a Schottky set  $S$  that exhibits WPD property.

**Definition 7.1.** Let  $(X, G)$  be as in Convention 2.4. Let  $S$  be a  $K_0$ -Schottky set associated with the auxiliary constants  $D_0, E_0$ . Let  $F$  be a finite subset of  $G$ . We say that  $S$  is  $F$ -WPD Schottky set if the following holds: if  $w \in G$  and  $s, s' \in S$  are such that the beginning and the ending points of  $\Gamma(s)$  and  $w\Gamma(s')$  are pairwise  $100E_0$ -close, then  $w \in F$  and  $\Pi(s)^{-1}w\Pi(s') \in F$ .

**Lemma 7.2.** Let  $(X, G)$  be as in Convention 2.4 and let  $g$  be a BGIP element in  $G$  with WPD property. Let  $A \subseteq G$  be a subset such that the semigroup  $\langle\langle A \rangle\rangle$  generated by  $A$  contains two independent BGIP elements and  $g$ . Then for each  $N$ , there exists  $m, K_0 > 0$ , a finite set  $F \subseteq G$  and a long enough  $K_0$ -Schottky set  $S \subseteq A^m$  with cardinality  $N$  and with  $F$ -WPD property.

*Proof.* Given  $N$ , Proposition 2.16 tells us that there exists  $K_0, m > 0$  and a finite set  $S'' \subseteq A^{m'}$  of cardinality  $N$  such that

$$S_k := \{(s, \alpha^{(k)}, s) : s \in S''\}$$

is a  $K_0$ -Schottky set for each  $k$ . We now take  $M = 2(100E_0 + \max_{s \in S''} \|s\|)$ . The WPD property of  $g$  tells us that

$$\text{Stab}_M(o, g^n o) = \{h \in G : d(o, ho) \leq E_0, d(g^n o, hg^n o) \leq E_0\}$$

is finite for all large enough  $n$ . We now claim that for a large enough  $n$ ,  $S_n$  is the desired long-enough  $K_0$ -Schottky set that has  $F$ -WPD property for

$$F := \{\Pi(s)t\Pi(u)^{-1} : s, u \in S', t \in \text{Stab}_M(o, g^n o)\}.$$

Indeed, for some  $s, u \in S$  and  $w \in G$ , suppose that the beginning and the ending points of

$$\Gamma(s, \alpha^{(n)}, s), w \cdot \Gamma(u, \alpha^{(n)}, u)$$

are each  $100E_0$ -close. In particular, we have

$$d(o, wo) < 100E_0, d(\Pi(s)o, w\Pi(u)o) < 100E_0 + 2\max_{t \in S''} \|t\| = M.$$

For a similar reason, we have  $\Pi(s)g^n o$  and  $w\Pi(u)g^n$  are  $M$ -close to each other. We now use the WPD property of  $g^n$  to conclude that  $\Pi(s)^{-1}w\Pi(u) \in \text{Stab}_M(o, g^n o)$ , or in other words,  $w \in F$ .  $\square$

**7.2. Sublinear growth of entropy of displacement.** In this section, we prove the following.

**Proposition 7.3.** *Let  $(G, X, o)$  be as in Convention 2.4. Let  $K_0, M_0 > 0$  and let  $S$  be a large and long enough  $K_0$ -Schottky set  $S \subseteq G^{M_0}$ . Suppose that:*

- (1)  $H(\mu) < \infty$ ,
- (2)  $\mu^{M_0}(s) > 0$  for each  $s \in S$ .

*Then for each  $\eta > 0$ , there exists  $\epsilon' = \epsilon'(\eta, \mu)$  and  $N = N(\eta, \mu)$  such that the following holds: for any probability measure  $\mu'$  on  $G$  with  $\|\mu - \mu'\|_0 < \epsilon'$  and  $|H(\mu) - H(\mu')| < \epsilon'$  and for any  $g \in G$ ,*

$$\begin{aligned} H_{g_1=g, (g_2, \dots, g_n) \sim \mu'^{n-1}}(\lfloor \|Z_n\| \rfloor = \lfloor d(o, g_1 \cdots g_n o) \rfloor) &\leq \eta n, \\ H_{(g_1, \dots, g_{n-1}) \sim \mu'^{n-1}, g_n=g}(\lfloor \|Z_n\| \rfloor = \lfloor d(o, g_1 \cdots g_n o) \rfloor) &\leq \eta n \end{aligned}$$

*holds for each  $n > N$ .*

This proposition is proved in [CFFT22] when the underlying space is Gromov hyperbolic, by means of pivoting technique. We explain Chawla-Forghani-Frisch-Tiozzo's argument here for the readers' convenience.

*Proof.* We will prove the first claim only; the argument for the second claim is symmetric.

Before we begin, we note that the Schottky set  $S$  and the constants  $K_0, E_0$  are fixed from the beginning. We now recall a lemma:

**Lemma 7.4** ([CFFT22, Lemma 2.4]). *For each  $M \in \mathbb{Z}_{>0}$  and  $\epsilon > 0$ , there exists  $\delta = \delta(M, \epsilon) > 0$  such that the following holds.*

*Let  $A$  be a countable set. Let  $Z : (\Omega, \mathbb{P}) \rightarrow A$  be a random variable and let  $U$  be a finite set of  $A$  of cardinality  $M$  such that  $\sum_{x \notin U} \varphi(\mathbb{P}(Z = x)) < \epsilon/2$ . Then for any measurable subset  $E \subseteq \Omega$  with  $\mathbb{P}(E) < \delta$ , the modified RV*

$$Y(\omega) := \begin{cases} Z(\omega) & \text{if } \omega \in E \\ * & \text{otherwise} \end{cases}$$

*has entropy smaller than  $\epsilon$ .*

Let  $\mu$  be a probability measure satisfying the assumption. Since  $H(\mu) = \sum_{g \in G} -\mu(g) \log \mu(g) < +\infty$ , there exists a finite set  $U \subseteq G$  such that  $\sum_{g \notin U} -\mu(g) \log \mu(g)$  is  $\eta/10$ -close to  $H(\mu)$ . Let  $\delta = \delta(\#U, \eta/3)$  be as in Lemma 7.4.

Let  $m = \min_{s \in S} \mu^{M_0}(s) > 0$ . Let  $\kappa = \kappa(\epsilon', M_0)$  as in Proposition 3.8 and let  $\alpha \in \mathbb{Z}_{>1}$  be a large enough integer such that

$$\frac{1}{\kappa} e^{-\kappa(\alpha-1)} < \delta/10.$$

We now take large enough finite subset  $F$  of  $G$  such that  $\mu(F) \geq 1 - \frac{1}{100(\alpha+1)} \delta$ .

We can take small enough  $\epsilon' < \eta/10$  such that if a probability measure  $\mu'$  satisfies

$$(7.1) \quad \|\mu' - \mu\|_0 < \epsilon' \text{ and } |H(\mu') - H(\mu)| < \epsilon'$$

then the following hold:

- (1)  $H(\mu') < H(\mu) + \eta/10$ ;
- (2)  $\sum_{g \in U} -\mu'(g) \log \mu'(g) > H(\mu) - \eta/5$ , and hence
- (3)  $\sum_{g \notin U} -\mu'(g) \log \mu'(g) < \eta/3$ ;
- (4)  $\mu'^{M_0}(s) > m/2$  for each  $s \in S$ , and
- (5)  $\mu'(F) > 1 - \frac{1}{50\alpha} \delta$ .

We will now take  $N(\eta, \mu)$  that works for  $\mu'$  satisfying Condition 7.1. But the actual value of  $N$  will only depend on  $L$  and  $\alpha$ .



Let  $n \in \mathbb{Z}_{>0}$  be given. We construct a probability space  $\Omega_n$  and its partition  $\mathcal{Q}_n$  into pivotal equivalence classes, aiming at RVs  $g_i$ 's with

$$(g_1, g_2, \dots, g_n) \sim 1_{g_1=g} \times \mu' \times \dots \times \mu'.$$

Then the set of pivotal times  $\mathcal{P}_n(\omega) = \{j(1) < \dots < j(\#\mathcal{P}_n)\} \subseteq \{1, \dots, n\}$  is a  $2^{\{1, \dots, n\}}$ -valued RV on  $\Omega_n$ . We then divide the index set  $\{1, \dots, n\}$  into  $\lceil n/\alpha \rceil$  intervals

$$I_1, I_2, \dots, I_{\lceil n/\alpha \rceil}$$

of lengths  $\alpha$  or  $\alpha - 1$ . Now, for each  $k = 1, \dots, \lceil n/\alpha \rceil$ , we declare  $k$  to be *good* if:

- (1) all elements in  $\{g_i : i \in I_{k-1} \cup I_k \cup I_{k+1}\}$  have norm less than  $L$ , i.e.,  $d(o, g_i o) < L$  for  $i \in I_{k-1} \cup I_k \cup I_{k+1}$  (by definition,  $k = 0$  and  $k = \lceil n/\alpha \rceil$  are not good);
- (2) there exists at least one pivotal time in  $I_{k-1}$  and at least one pivotal time  $I_{k+1}$ , i.e.,  $\mathcal{P}_n(\omega) \cap I_{k-1} \neq \emptyset \neq \mathcal{P}_n(\omega) \cap I_{k+1}$ .

For  $1 \leq k < \lceil n/\alpha \rceil$ ,  $\{\omega : k \text{ is not good}\}$  has probability less than  $\delta$ . To see this, first recall that  $\mathbb{P}(g_i : \|g_i\| > L)$  has probability at most  $\frac{1}{50\alpha}\delta$ . Since there are at most  $3\alpha$  in 3 consecutive intervals, the probability that  $\|g_i\| > L$  for at least one  $i \in I_{k-1} \cup I_k \cup I_{k+1}$  is still less than  $\delta/15$ . Next,  $I_{k-1}$  has no pivotal time for probability at most  $\frac{1}{\kappa}e^{-\kappa(\alpha-1)} < \delta/10$ , and similar estimate holds for  $I_{k+1}$ . Hence, the probability that both intervals do not have any pivotal time is less than  $\delta/5$ .

Now for each  $i = 1, \dots, n$ , we define

$$h_i := \begin{cases} g_i & \text{the index } k \text{ for which } i \in I_k \text{ is not good} \\ * & \text{the index } k \text{ for which } i \in I_k \text{ is not good.} \end{cases}$$

Since  $\mathbb{P}(\text{the index } k \text{ for which } i \in I_k \text{ is not good})$  is less than  $\delta$ , Lemma 7.4 tells us that  $H(h_i) < \eta/3$  for each  $i$ . Considering this, the entropy of the sequence  $\mathbf{h} = (h_1, h_2, \dots, h_n)$  is at most  $\eta n/3$ .

Now we consider the conditional entropy of  $\lfloor \|Z_n(\omega)\| \rfloor$  conditioned on the values of  $\mathbf{h}$ . It is clear that  $\mathbf{h} = (h_1, \dots, h_n)$  is an alternation of some sequences in  $G$  and some sequences of  $*$ . Let  $(h_{a(k)}, \dots, h_{b(k)})$  be the  $k$ -th sequence of  $G$  that appears consecutively in  $(h_1, \dots, h_n)$ , and we define

$$S(\mathbf{h}) := \left\lfloor \sum_k \|h_{a(k)} \cdots h_{b(k)}\| \right\rfloor = \left\lfloor \sum_k d(Z_{a(k)-1}o, Z_{b(k)}o) \right\rfloor.$$

We claim that:

$$S(\mathbf{h}) - \lfloor \|Z_n\| \rfloor \in \{l \in \mathbb{Z} : |l| \leq E_0 n + 2L\alpha n + 2\}.$$

If the entire  $\mathbf{h}$  is recorded with elements of  $G$  (not  $*$ ), then the claim is clear. If not, let  $T$  be the number of consecutive appearances of elements of  $G$  (so that there are  $T-1$  consecutive appearances of  $*$ ). The indices for the  $k$ -th consecutive appearance of  $*$ 's are  $(b(k) + 1, \dots, a(k+1) - 1)$ . This is a union of intervals with good index, say  $I_{c(k)}, I_{c(k)+1}, \dots, I_{d(k)}$ . Because  $c(k)$  is a good index,  $\|g_i\| < L$  holds for all  $i \in I_{c(k)-1}$ . (\*) Moreover,  $I_{c(k)-1}$  contains a pivotal time; take the largest one and label it by  $j''(k)$ . By (\*) and by the fact that each interval no longer than  $\alpha$ , we conclude:

$$(7.2) \quad d(Z_{b(k)}, Z_{j''(k)}) < L\alpha.$$

For a similar reason, there exists a pivotal time  $j'(k)$  in  $I_{d(k)+1}$  (we take the smallest one), and we conclude

$$(7.3) \quad d(Z_{a(k+1)-1}, Z_{j'(k)}) < L\alpha.$$

Finally, since  $c(k), c(k) + 1, \dots, d(k)$  are all good,  $\|g_i\| < L$  holds for all  $i \in I_{c(k)-1} \cup \dots \cup I_{d(k)+1}$ . This implies that

$$(7.4) \quad d(Z_{j''(k)}, Z_{j'(k)}) \leq L(j'(k) - j''(k)).$$

Now, note that  $j'(k) \leq j''(k+1)$ ; indeed, we have  $d(k) + 1 \leq c(k) - 1$  as the  $k$ -th and the  $k+1$ -th consecutive appearance of  $*$ 's are separated by some interval with bad index. This implies that

$$(o, \mathbf{Y}_{j''(1)}, \mathbf{Y}_{j'(1)}, \dots, \mathbf{Y}_{j''(k)}, \mathbf{Y}_{j'(k)}, \dots, Z_n o)$$

is a  $D_0$ -semi-aligned sequence if we remove the duplicates. By Proposition 2.10, we have

$$\left| d(o, Z_{j''(1)}) + \sum_{k=1}^{T-1} d(Z_{j''(k)}, Z_{j'(k)}) + \sum_{k=1}^{T-2} d(Z_{j'(k)}, Z_{j''(k+1)}) + d(Z_{j''(T-1)}, Z_n) - d(o, Z_n o) \right| \leq 2E_0 T.$$

By plugging in Inequality 7.2, 7.3 and 7.4, we conclude that

$$|\|Z_n\| - S(\mathbf{h})| \leq 2E_0 T + \sum_k L(j'(k) - j''(k)) \leq 2E_0 T + Ln.$$

Taking the integer part ends the proof.

Now, any random variable whose range is  $k$  has entropy at most  $\log k$ . Hence, we have

$$H(\lfloor \|Z_n\| \rfloor \mid \mathbf{h}) \leq \log n + \log(2E_0 + L) + 2.$$

This is smaller than  $\eta n/3$  for all large enough  $n$  (depending on  $E_0$  and  $L$ ). Since  $H(\mathbf{h}) \leq \eta n/3$ , we conclude

$$H(\lfloor \|Z_n\| \rfloor) \leq H(\lfloor \|Z_n\| \rfloor \mid \mathbf{h}) + H(\mathbf{h}) \leq \eta n. \quad \square$$

**Corollary 7.5.** *Let  $(G, X, o)$  be as in Convention 2.4 and let  $\mu$  be a non-elementary probability measure on  $\mu$  with finite entropy. Let  $\nu$  be a probability measure with finite entropy. Let  $\eta > 0$ , and let  $\epsilon' = \epsilon'(\eta, \mu) > 0$  and  $N = N(\eta, \mu)$  be as in Proposition 7.3.*

*Then for any  $g \in G$ , for any probability measure  $\mu'$  such that  $\|\mu' - \mu\|_0 < \epsilon'$  and  $|H(\mu) - H(\mu')| < \epsilon'$ , and for any  $n > N$ , we have*

$$H_{(g_1, \dots, g_{n-1}) \sim \mu'^{n-1}, g_n \sim \nu}(\lfloor \|Z_n\| \rfloor) \leq \eta n + H(\nu).$$

*Proof.* This is immediate from

$$H(\lfloor \|Z_n\| \rfloor) \leq H(\lfloor \|Z_n\| \rfloor \mid g_n) + H(g_n) \leq \eta n + H(\nu). \quad \square$$

**Lemma 7.6.** *Let  $A$  be a countable set, let  $\epsilon < 0.1$  and let  $Z, Y : (\Omega, \mathbb{P}) \rightarrow A$  be RVs such that*

$$(1 - \epsilon) \mathbb{P}(Z(\omega) = a) \leq \mathbb{P}(Y(\omega) = a) \leq (1 + \epsilon) \mathbb{P}(Z(\omega) = a) \quad (\forall a \in A).$$

*Then we have  $(1 - 2\epsilon)H(Z) - 2\epsilon \leq H(Y) \leq (1 + \epsilon)H(Z) + 2\epsilon$ .*

*Proof.* By the triangle inequality, we have

$$\begin{aligned} H(Y) &= \sum_{a \in A} -\mathbb{P}(Y = a) \log \mathbb{P}(Y = a) \\ &\leq \sum_{a \in A} \mathbb{P}(Y = a) |\log \mathbb{P}(Y = a) - \log \mathbb{P}(Z = a)| + \sum_{g \in G} -|\mathbb{P}(Y = a) - \mathbb{P}(Z = a)| \log \mathbb{P}(Z = a) \\ &\quad + \sum_{g \in G} -\mathbb{P}(Z = a) \log \mathbb{P}(Z = a) \\ &\leq \log(1 + \epsilon) \sum_{a \in A} \mathbb{P}(Y = A) + \sum_{a \in A} -\epsilon \mathbb{P}(Z = a) \log \mathbb{P}(Z = a) + H(Z) \\ &\leq 1.2\epsilon + (1 + \epsilon)H(Z). \end{aligned}$$

Also, the assumption tells us that  $\mathbb{P}(Z = a)/\mathbb{P}(Y = a)$  is bounded between  $1 - 1.2\epsilon$  and  $1 + 1.2\epsilon$  for any  $a \in A$ . The above argument then tells us that  $H(Z) \leq 1.44\epsilon + (1 + 1.44\epsilon)H(Y)$ , which implies  $H(Y) \geq (1 - 1.44\epsilon)H(Z) - 2\epsilon$ .  $\square$

**7.3. More precise model of pivoting.** 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 2.6. We also employ the notation  $\mathbf{Y}_i := (Z_i o, \dots, Z_{i+M_0} o)$ .

**Definition 7.7** ([Cho22a, Definition 6.1]). *Let  $\mu$  and  $\nu$  be non-elementary probability measures on  $G$  and  $(\Omega, \mathbb{P})$  be a probability space for  $\mu$ . Let  $0 < \epsilon < 1$ ,  $K_0, N > 0$  and let  $S \subseteq (\text{supp } \mu)^{M_0}$  be a long enough  $K_0$ -Schottky set.*

*A subset  $\mathcal{E}$  of  $\Omega$  is called an  $(n, N, \epsilon, \nu)$ -pivotal equivalence class for  $\mu$ , associated with the set of pivotal times*

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

*if the following hold:*

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

$$\begin{aligned} s_k(\omega) &:= (g_{j(k)-M_0+1}(\omega), g_{j(k)-M_0+2}(\omega), \dots, g_{j(k)}(\omega)), \\ s'_k(\omega) &:= (g_{j'(k)-M_0+1}(\omega), g_{j'(k)-M_0+2}(\omega), \dots, g_{j'(k)}(\omega)) \end{aligned}$$

*are Schottky sequences;*

(2) *for each  $\omega \in \mathcal{E}$ ,  $(o, \mathbf{Y}_{j(1)}, \mathbf{Y}_{j'(1)}, \dots, \mathbf{Y}_{j(\# \mathcal{P}/2)}, \mathbf{Y}_{j'(\# \mathcal{P}/2)}, Z_n o)$  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  $\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{2M_0}-N-0.5}$  in the sense that*

$$(1 - \epsilon)(\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{2M_0}-N-0.5})(g) \leq \mathbb{P}(r_k = g) \leq (1 + \epsilon)(\mu^{*2M_0N} * \nu^{*\frac{j'(k)-j(k)}{2M_0}-N-0.5})(g).$$

*for each  $g \in G$ .*

(4) *On  $\mathcal{E}$ , for  $i \notin \{j(k) - M_0 + 1 \leq l \leq j(k) : k = 1, \dots, \# \mathcal{P}/2\}$ ,  $g_i$  is fixed.*

**Proposition 7.8** ([Cho22a, Proposition 6.2]). *Let  $\mu$  be a non-elementary probability measure on  $G$ , let  $0 < \epsilon < 1$  and let  $S \in G^{M_0}$  be a long enough Schottky set for  $\mu$  with cardinality greater than  $10/\epsilon$ . Let  $m := \min\{\mu^{M_0}(s) : s \in S\}$ , let  $N > 20/\epsilon m$  and let  $\nu$  be a probability measure defined as*

$$\nu := \frac{1}{1 - 0.5m^2} (\mu^{*2M_0} - 0.5m^2 \cdot (\text{uniform measure on } \{\Pi(s)\Pi(s') : s, s' \in S\})).$$

*Then  $\nu$  is a non-elementary probability measure. Moreover, there exists  $K > 0$  depending only on  $S, m, N$  and  $\epsilon$  (but not on  $\mu$ ) such that, for each  $n$ , we have a probability space  $(\Omega, \mathbb{P})$  for  $\mu$  and its measurable partition  $\mathcal{P}_{n, N, \epsilon, \nu} = \{\mathcal{E}_\alpha\}_\alpha$  into  $(n, N, \epsilon, \nu)$ -pivotal equivalence classes, associated with the set of pivotal times  $\mathcal{P}^{(n, N, \epsilon, \nu)}$ , that satisfies*

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

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

**Lemma 7.9.** *Let  $g$  be a contracting isometry of a metric space  $(X, d)$ . Then for each  $N$  there exists  $K$  such that*

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

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

*Proof.* Let  $g$  be a  $K'$ -contracting isometry. Since  $|d(o, g^i o) - d(o, g^j o)| \leq d(g^i o, g^j o) \leq |i - j| \|g\|$ , 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  $\|g^k\| \geq 10K'$ . Now, for each

$i$ , the contracting property of the set  $\{o, go, \dots, g^{k(i+1)}o\}$  forces that  $[o, g^{k(i+1)}o]$  is  $2K'$ -close to  $g^{ki}o$ . Hence we have

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

Now the conclusion follows.  $\square$

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

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

**Lemma 7.10** ([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  $d_{\text{Hauss}}(\{hg^n o\}_{n \in \mathbb{Z}}, \{g^n o\}_{n \in \mathbb{Z}}) < C$  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  $E(g)$ :

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

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 7.9, 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  $\text{Stab}_{K'}(o, g^N o)$ , which is a finite set when  $N$  is sufficiently large. Hence,  $h = g^i a$  belongs to a finite extension of  $\langle g \rangle$ .  $\square$

We now prove Theorem E.

*Proof.* Let  $g$  be a WPD contracting isometry of  $X$ , and let  $\mu$  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  $\mu$  with  $H(\mu_i) \rightarrow H(\mu)$ , and let  $\epsilon > 0$ . We first note that  $\limsup_{i \rightarrow \infty} H(\mu_i) \leq H(\mu)$ : 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

$$(7.5) \quad \liminf_{i \rightarrow \infty} H(\mu_i) \geq H(\mu).$$

We divide into elementary and non-elementary cases. First, consider the case that  $\text{supp } \mu$  is contained in the elementary closure  $E(g)$  of  $g$ . Since  $E(g)$  is a virtually cyclic, the  $\mu$ -random walk is a random walk on an amenable group and its asymptotic entropy must vanish. In this case, Inequality 7.5 is immediate.

In the case that there exists  $h \in \text{supp } \mu \setminus E(g)$ ,  $g^m$  and  $hg^m h^{-1}$  are independent contracting isometries and  $\mu$  is non-elementary.

Let  $\epsilon > 0$  be arbitrary. We will construction  $\epsilon_{meas} > 0$  such that Item (2) of the strategy in the beginning of this section holds. For this, let  $\eta > 0$  be another small positive such that

$$((1 - 5\eta)h(\mu) - (4H(\mu) + 0.5)\eta) \cdot (1 - \eta) \geq (1 - \epsilon)h(\mu).$$

Using Lemma 7.2, we obtain a finite set  $F \subseteq G$  and  $K_0, M_0 > 0$  and a large enough and long  $K_0$ -Schottky set  $S \subseteq G^{M_0}$  of cardinality  $> 100/\eta$  such that  $\mu^{M_0}(s) > 0$  for each  $s \in S$ . Let  $\mathbf{m}$  be the minimum of  $\mu^{M_0}(s)$  for  $s \in S$ .

We now consider the condition

$$(7.6) \quad \|\mu' - \mu\|_0 < \epsilon_{meas} \text{ and } |H(\mu') - H(\mu)| < \epsilon_{meas}$$

for some  $\epsilon_{meas} > 0$ . There exists a small enough choice of  $\epsilon_{meas}$ , accompanied with a threshold  $N_{sub}$  such that for any  $g \in G$ , for any  $n > N_{sub}$  and for any  $\mu'$  satisfying Equation 7.6,

$$H_{(g_1, \dots, g_{n-1}) \sim \mu'^{n-1}, g_n = g} H(\lfloor d(o, g_1 \cdots g_n o) \rfloor) < \frac{\eta}{10} n.$$

Such  $\epsilon_{meas}$  and  $N_{sub}$  is given by Proposition 7.3. We now fix  $N > 40/\eta \mathbf{m} + N_{sub} + (\log \#S)/\eta M_0$  that is also large enough such that:

$$H(\mu^{*2M_0N}) \geq (1 - \eta) \cdot 2M_0 N h(\mu)$$

and such that Inequality 7.8 is satisfied. Choosing some smaller  $\epsilon_{meas}$  if necessary, we can also guarantee the following for the measures  $\mu'$  satisfying Equation 7.6:

$$0.5\mathbf{m} \leq \min \mu'^{M_0}(s) \leq 2\mathbf{m}, \quad H(\mu') < 2H(\mu), \quad H(\mu'^{*2M_0N}) > (1 - \eta)H(\mu^{*2M_0N}).$$

In the remaining, we will deal with an arbitrary probability measure  $\mu'$  satisfying Equation 7.6.

Let  $N > 40/\eta \mathbf{m} + N_{sub}$ . Let also  $m = \min_{s \in S} \mu'^{M_0}(s)$  and

$$\nu' := \frac{1}{1 - 0.5m^2} (\mu'^{*2M_0} - 0.5m^2 \cdot (\text{uniform measure on } \{\Pi(s)\Pi(s') : s, s' \in S\})).$$

For this choice, note that

$$H(\nu') \leq \frac{1}{1 - 0.5m^2} \cdot H(\mu'^{*2M_0}) \leq 2 \cdot 2M_0 H(\mu') \leq 8M_0 H(\mu).$$

By Proposition 7.8, for each  $n$  there exists a measurable partition  $\mathcal{P}_{n,N,\eta,\nu'} = \{\mathcal{E}_\alpha\}_\alpha$  of  $(G^n, \mu'^n)$  into  $(n, N, \epsilon, \nu)$ -pivotal equivalence classes such that  $\frac{1}{2} \# \mathcal{P}^{(n,N,\eta,\nu')}(w) > (1 - \eta)n/2M_0N$  holds for probability  $1 - Ke^{-n/K}$ . Here, note that  $K$  does not depend on the nature of  $\mu$  but only on our choice of  $S, \mathbf{m}, N$  and  $\eta$ . Hence, we can choose  $\mathbf{n}$  only depending on  $S, \mathbf{m}, N, \eta$  such that

$$(7.7) \quad \mathbb{P} \left( \frac{1}{2} \# \mathcal{P}^{(n,N,\eta,\nu')}(w) > (1 - \eta)n/2M_0N \right) \geq 1 - \eta$$

holds for all  $n > \mathbf{n}$ .

We now condition on a  $(n, N, \eta, \nu)$ -pivotal equivalence class  $\mathcal{E} \in \mathcal{P}_{n,N,\epsilon,\nu}$ . For convenience, we use notation  $T := \frac{1}{2} \# \mathcal{P}^{(n,N,\eta,\nu)}(\mathcal{E})$ , i.e.,

$$\mathcal{P}(\mathcal{E}) = \{j(1) < j'(1) < \dots < j(T) < j'(T)\}.$$

Note that  $T \leq n/2M_0N$  always hold, as

$$n > \sum_{k=1}^T (j'(k) - j(k)) > 2M_0N \cdot T.$$

Our first claim is

**Claim 7.11.** *For each  $i = 1, \dots, T$ , we have*

$$H(r_i | \mathcal{E}) \geq (1 - 2\eta)H(\mu'^{*2M_0N}) - 2\eta$$

*Proof of Claim 7.11.* Let  $f_i$  be the distribution of  $\mu'^{*2M_0N} * \nu'^{* \frac{j'(k) - j(k)}{2M_0} - N - 0.5}$ , i.e.,  $f_i(g) = \mu'^{*2M_0N} * \nu'^{* \frac{j'(k) - j(k)}{2M_0} - N - 0.5}(g)$ . Note that the entropy of  $f_i$  is at least  $H(\mu'^{*2M_0N})$ , because

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

holds for every pair of RVs  $Y_1$  and  $Y_2$  on  $G$ . Now, since  $r_i$  is almost distributed according to  $f_i$ , Lemma 7.6 tells us the  $H(r_i | \mathcal{E}) \geq (1 - 2\eta)H(\mu'^{*2M_0N}) - 2\eta$  as desired.  $\square$

Note that

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

Since we are discussing everything inside  $\mathcal{E}$ ,  $h_i$ 's are always fixed. Let  $\mathcal{F}$  be the partition of  $\mathcal{E}$  based on the values of

$$\{\lfloor \|Z_{j(k)-M_0}\| \rfloor, \lfloor \|Z_{j(k)}\| \rfloor, \lfloor \|Z_{j(k)'}\| \rfloor, \lfloor \|Z_{j(k)+M_0}\| \rfloor : k = 1, \dots, T\}.$$

**Claim 7.12.** *We have*

$$H(\mathcal{F}|\mathcal{E}) \leq \frac{\eta n}{4} + 4H(\mu) \cdot (n - 2M_0NT).$$

*Proof of Claim 7.12.* Let  $k \in \{1, \dots, T\}$ . Let  $Z$  be an RV distributed according to  $\mu'^{*2M_0N} * \nu'^{* \frac{j'(k)-j(k)}{2M_0} - N - 0.5}$ . Then the ratio of the distributions of  $\lfloor \|Z\| \rfloor$  and  $\lfloor \|r_k\| \rfloor$  is uniformly between  $(1 - \eta)$  and  $(1 + \eta)$ . By Lemma 7.6, we have

$$H(\lfloor \|r_k\| \rfloor | \mathcal{E}) \leq (1 + 2\eta)H(\lfloor \|Z\| \rfloor) + 2\eta.$$

Meanwhile, since  $2M_0N > N > N_{sub}$ , Proposition 7.3 and Corollary 7.5 tells us that

$$\begin{aligned} H(\lfloor \|Z\| \rfloor) &\leq \frac{\eta}{10} \cdot 2M_0N + H(\nu'^{* \frac{j'(k)-j(k)}{2M_0} - N - 0.5}) \\ &\leq \frac{\eta}{10} \cdot 2M_0N + 8M_0H(\mu) \cdot \left( \frac{j'(k) - j(k)}{2M_0} - N - 0.5 \right). \end{aligned}$$

Combined together, we have

$$H(\lfloor \|r_k\| \rfloor | \mathcal{E}) \leq \frac{\eta}{4} M_0N + 5M_0H(\mu) \cdot (j'(k) - j(k) - 2M_0N - M_0) + 2\eta.$$

Let us now extract the information

$$\{\lfloor \|Z_{j(k)-M_0}\| \rfloor, \lfloor \|Z_{j(k)}\| \rfloor, \lfloor \|Z_{j(k)'}\| \rfloor, \lfloor \|Z_{j(k)+M_0}\| \rfloor : k = 1, \dots, T\}.$$

from the RVs  $\lfloor \|r_k\| \rfloor, s_k, s'_k$ 's. Proposition 2.10 asserts that  $[o, Z_n o]$  passes through the  $E_0$ -neighborhoods of  $Z_{j(1)-M_0}, Z_{j(1)}, Z_{j'(1)}, Z_{j'(1)+M_0}, \dots, Z_{j(T)-M_0}, Z_{j(T)}, Z_{j'(T)}, Z_{j'(T)+M_0}$  in order. Hence, setting  $j'(-1) := -M_0$  for convenience, the following quantities:

$$\begin{aligned} &\|Z_{j'(k-1)+M_0}\| + \|Z_{j'(k-1)+M_0}^{-1} Z_{j(k)-M_0}\| - \|Z_{j(k)-M_0}\|, \\ &\|Z_{j(k)-M_0}\| + \|s_k\| - \|Z_{j(k)}\|, \\ &\|Z_{j(k)}\| + \|r_k\| - \|Z_{j(k)'}\|, \\ &\|Z_{j'(k)}\| + \|s'_k\| - \|Z_{j'(k)+M_0}\| \end{aligned} \quad (k = 1, \dots, T)$$

are all real numbers between 0 and  $6E_0$ . Hence, if we define  $\alpha_k, \beta_k, \gamma_k, \delta_k$  so that

$$\begin{aligned} \lfloor \|Z_{j(k)-M_0}\| \rfloor &= \lfloor \|Z_{j'(k-1)+M_0}\| \rfloor + \lfloor \|Z_{j'(k-1)+M_0}^{-1} Z_{j(k)-M_0}\| \rfloor + \alpha_k, \\ \lfloor \|Z_{j(k)}\| \rfloor &= \lfloor \|Z_{j(k)-M_0}\| \rfloor + \lfloor \|s_k\| \rfloor + \beta_k, \\ \lfloor \|Z_{j'(k)}\| \rfloor &= \lfloor \|Z_{j(k)}\| \rfloor + \lfloor \|r_k\| \rfloor + \gamma_k, \\ \lfloor \|Z_{j'(k)+M_0}\| \rfloor &= \lfloor \|Z_{j'(k)}\| \rfloor + \lfloor \|s'_k\| \rfloor + \delta_k \end{aligned}$$

then  $\alpha_k, \beta_k, \gamma_k, \delta_k$ 's are integers between 3 and  $-(6E_0 + 3)$ . Here, note that  $Z_{j'(k-1)+M_0}^{-1} Z_{j(k)-M_0}$ 's are fixed information across the pivotal equivalence class  $\mathcal{E}$ . Hence, as soon as we have information

$(\lfloor \|r_k\| \rfloor, s_k, s'_k, \alpha_k, \beta_k, \delta_k, \delta_k)_{k=1}^T$ , we determine all information for  $\mathcal{F}$ . We conclude that

$$\begin{aligned}
H(\mathcal{F} \mid \mathcal{E}) &\leq H\left(\lfloor \|r_k\| \rfloor, s_k, s'_k, \alpha_k, \beta_k, \delta_k, \delta_k)_{k=1}^T \mid \mathcal{E}\right) \\
&\leq \sum_{k=1}^T H(\lfloor \|r_k\| \rfloor) + \sum_{k=1}^T H(s_k) + \sum_{k=1}^T H(s'_k) \\
&\quad + \sum_{k=1}^T H(\alpha_k) + \sum_{k=1}^T H(\beta_k) + \sum_{k=1}^T H(\gamma_k) + \sum_{k=1}^T H(\delta_k) \\
&\leq T \cdot \frac{\eta M_0 N}{4} + 4H(\mu) \cdot \#\{1, \dots, n\} \setminus \{j(k) + 2M_0N + 1, \dots, j'(k) - M_0 : k = 1, \dots, T\} \\
&\quad + T(2 \log \#S + 4 \log(6E_0 + 6)) \\
&\leq \frac{\eta n}{8} + 4H(\mu) \cdot (n - 2M_0N \cdot T) + \frac{\eta n}{8}.
\end{aligned}$$

In the final step, we used the fact that  $\{j(k) + 1, \dots, j'(k) - M_0\}$ 's for  $k = 1, \dots, T$  are disjoint intervals containing at least  $2M_0N$  integers, and that  $N$  is large enough so that

$$(7.8) \quad \frac{2 \log \#S + 4 \log(6E_0 + 6)}{M_0N} \leq 0.1\eta. \quad \square$$

Now, we claim:

**Claim 7.13.** *Conditioned on each equivalence class  $E$  of  $\mathcal{F}$  (that partitions  $\mathcal{E}$ ), we have*

$$H(Z_n \mid E) \geq H((r_1, \dots, r_T) \mid E) - 2T \log \#F.$$

*Proof of Claim 7.13.* Fix an equivalence class  $E$ , an element  $g \in G$  and an element  $\omega \in E$  such that  $Z_n(\omega) = g$ . We claim that if  $\omega' \in E$  satisfies  $Z_n(\omega') = g$ , then  $r_i(\omega') \in F \cdot r_i(\omega) \cdot F$  for each  $i = 1, \dots, T$ . To see this, pick an arbitrary  $k \in \{1, \dots, T\}$ . Recall that

$$(o, \mathbf{Y}_{j(k)}(\omega), \mathbf{Y}_{j(k)}(\omega), go), \quad (o, \mathbf{Y}_{j(k)}(\omega'), \mathbf{Y}_{j(k)}(\omega'), go)$$

are each  $D_0$ -semi-aligned. Hence, there exists subsegments  $\eta_k, \eta'_k$  of  $[o, go]$  such that  $\eta_k$  is  $0.1E_0$ -fellow traveling with  $\mathbf{Y}_{j(k)}(\omega)$  and  $\eta_k$   $0.1E_0$ -fellow traveling with  $\mathbf{Y}_{j(k)}(\omega')$ . Let us denote the beginning point  $\eta_k$  and  $\eta'_k$  by  $p$  and  $p'$ , respectively. Since chose  $\omega, \omega'$  from the same equivalence class  $E$  of  $\mathcal{F}$ ,  $\|Z_{j(k)-M_0}(\omega)\|$  and  $\|Z_{j(k)-M_0}(\omega')\|$  have the same integer part and they differ by at most  $1 < E_0$ . This implies that  $d(o, p)$  and  $d(o, p')$  differ by at most  $0.1E_0 + E_0 + 0.1E_0 < 2E_0$ . Meanwhile,  $p$  and  $p'$  are on the same geodesic starting from  $o$ . It follows that  $d(p, p') < 2E_0$ . Hence,  $Z_{j(k)-M_0}(\omega)$  and  $Z_{j(k)-M_0}(\omega')$  are  $(2E_0 + 0.1E_0 + 0.1E_0)$ -close. For the same reason, the endpoints of  $\eta_k$  and  $\eta'_k$  are  $2E_0$ -close. This in turn implies that  $Z_{j(k)}(\omega)$  and  $Z_{j(k)}(\omega')$  are  $3E_0$ -close.

We are now in the situation that the beginning and the ending points of

$$\Gamma(s_k(\omega)), \quad (Z_{j(k)-M_0}(\omega))^{-1} Z_{j(k)-M_0}(\omega') \Gamma(s_k(\omega'))$$

are pairwise  $3E_0$ -close. Since  $S$  is a  $F$ -WPD  $K_0$ -Schottky set, we conclude that

$$(Z_{j(k)-M_0}(\omega))^{-1} Z_{j(k)-M_0}(\omega') \in F, \quad (Z_{j(k)}(\omega))^{-1} Z_{j(k)}(\omega') \in F. \quad (k = 1, \dots, T)$$

For a similar reason, we have

$$(Z_{j'(k)}(\omega))^{-1} Z_{j'(k)}(\omega') \in F, \quad (Z_{j(k)+M_0}(\omega))^{-1} Z_{j(k)+M_0}(\omega') \in F. \quad (k = 1, \dots, T)$$

We then conclude that

$$\begin{aligned}
r_k(\omega') &= (Z_{j(k)+M_0}(\omega'))^{-1} Z_{j'(k)}(\omega') \\
&= (Z_{j(k)+M_0}(\omega'))^{-1} Z_{j(k)+M_0}(\omega) \cdot (Z_{j(k)+M_0}(\omega))^{-1} Z_{j'(k)}(\omega) \cdot (Z_{j'(k)}(\omega))^{-1} Z_{j'(k)}(\omega') \\
&\in F \cdot (Z_{j(k)+M_0}(\omega))^{-1} Z_{j'(k)}(\omega) \cdot F = F \cdot r_k(\omega) \cdot F.
\end{aligned}$$

This claim leads to the fact that there are at most  $(\#F)^2$  choices of  $r_i$  for each  $i$  when conditioned with the value of  $Z_n$ . The entropy of such choices is at most  $T \cdot \log(\#F)^2$ . Now recall Bayes' rule:

$$H(Z_n \mid (r_i)_{i=1}^T, E) = H((r_i)_{i=1}^T \mid Z_n, E) - H((r_i)_{i=1}^T \mid E) + H(Z_n \mid E).$$

By rearranging, we obtain

$$\begin{aligned} H(Z_n \mid E) &= H(Z_n \mid (r_i)_{i=1}^T, E) + H((r_i)_{i=1}^T \mid E) - H((r_i)_{i=1}^T \mid Z_n, E) \\ &\geq H((r_i)_{i=1}^T \mid E) - 2T \log \#F. \quad \square \end{aligned}$$

Let us now put the claims together. Let  $\mathcal{E}$  be a pivotal equivalence class with  $\frac{1}{2} \# \mathcal{P}^{(n, N, \eta, \nu')}(\mathcal{E}) = T$ . Recall the partition  $\mathcal{F}$  made on  $\mathcal{E}$ . We have

$$\begin{aligned} H(Z_n \mid \mathcal{E}) &\geq H(Z_n \mid \mathcal{F}) \\ &\geq H((r_1, \dots, r_T) \mid \mathcal{F}) - 2T \log \#F \\ &\geq H((r_1, \dots, r_T) \mid \mathcal{E}) - H(\mathcal{F} \mid \mathcal{E}) - 2T \log \#F \\ &= \sum_{i=1}^T H(r_i \mid \mathcal{E}) - 2T \log \#F - \frac{\eta n}{4} - 4H(\mu)(n - 2M_0NT) \\ &\geq T(1 - 2\eta)H(\mu'^{2M_0N}) - 2T\eta - \frac{n}{M_0N} \log \#F - \frac{\eta n}{4} - 4H(\mu)(n - 2M_0NT) \\ &\geq T(1 - 2\eta)(1 - 2\eta)2M_0N \cdot h(\mu) - \frac{\eta n}{2} - 4H(\mu)(n - 2M_0NT) - 2\eta. \end{aligned}$$

Here we used the fact that  $N > 5(\log \#F)/M_0\eta$ .

Now consider the collection

$$\mathcal{Q} := \left\{ \mathcal{E} \in \mathcal{P}_{n, N, \eta, \nu'} : T = \frac{1}{2} \# \mathcal{P}^{(n, N, \eta, \nu')}(\mathcal{E}) > (1 - \eta) \frac{n}{2M_0N} \right\}$$

When  $\mathcal{E}$  belongs to  $\mathcal{Q}$ , the above estimate is greater than

$$(1 - 5\eta)nh(\mu) - (4H(\mu) + 0.5)\eta n.$$

It remains to sum up the conditional entropies on each  $\mathcal{E} \in \mathcal{P}_{n, N, \eta, \nu'}$ . Equation 7.7 guarantees that equivalence classes in  $\mathcal{Q}$  take up probability at least  $1 - \eta$ . This tells us that

$$\begin{aligned} H(Z_n) &\geq H(Z_n \mid \mathcal{P}_{n, N, \eta, \nu'}) \\ &= \sum_{\mathcal{E} \in \mathcal{P}_{n, N, \eta, \nu'}} H(Z_n \mid \mathcal{E}) \cdot \mathbb{P}(\mathcal{E}) \\ &\geq \sum_{\mathcal{E} \in \mathcal{Q}} H(Z_n \mid \mathcal{E}) \cdot \mathbb{P}(\mathcal{E}) \\ &\geq \sum_{\mathcal{E} \in \mathcal{Q}} ((1 - 5\eta)nh(\mu) - (4H(\mu) + 0.5)\eta n) \cdot \mathbb{P}(\mathcal{E}) \\ &\geq ((1 - 5\eta)nh(\mu) - (4H(\mu) + 0.5)\eta n) \cdot (1 - \eta). \end{aligned}$$

This is as we wanted. □

## 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.
- [ACT15] Goulmira N. Arzhantseva, Christopher H. Cashen, and Jing Tao. Growth tight actions. *Pacific J. Math.*, 278(1):1–49, 2015.
- [Ago13] Ian Agol. The virtual Haken conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning.
- [AKB12] Yael Algom-Kfir and Mladen Bestvina. Asymmetry of outer space. *Geom. Dedicata*, 156:81–92, 2012.



- [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.
- [Anc90] A. Ancona. Théorie du potentiel sur les graphes et les variétés. In *École d’été de Probabilités de Saint-Flour XVIII—1988*, volume 1427 of *Lecture Notes in Math.*, pages 1–112. Springer, Berlin, 1990.
- [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.
- [BF02] Mladen Bestvina and Koji Fujiwara. Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.*, 6:69–89, 2002.
- [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.
- [Bow12] B. H. Bowditch. Relatively hyperbolic groups. *Internat. J. Algebra Comput.*, 22(3):1250016, 66, 2012.
- [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 II: translation length and quasi-isometric embedding. *arXiv preprint arXiv:2212.12119*, 2022.
- [Cho22c] 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.
- [EM11] Alex Eskin and Maryam Mirzakhani. Counting closed geodesics in moduli space. *J. Mod. Dyn.*, 5(1):71–105, 2011.
- [Fer18] Talia Fernós. The Furstenberg-Poisson boundary and CAT(0) cube complexes. *Ergodic Theory Dynam. Systems*, 38(6):2180–2223, 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.

- [GP13] Victor Gerasimov and Leonid Potyagailo. Quasi-isometric maps and Floyd boundaries of relatively hyperbolic groups. *J. Eur. Math. Soc. (JEMS)*, 15(6):2115–2137, 2013.
- [GP16] Victor Gerasimov and Leonid Potyagailo. Quasiconvexity in relatively hyperbolic groups. *J. Reine Angew. Math.*, 710:95–135, 2016.
- [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.
- [Led12] François Ledrappier. Analyticity of the entropy for some random walks. *Groups Geom. Dyn.*, 6(2):317–333, 2012.
- [Led13] François Ledrappier. Regularity of the entropy for random walks on hyperbolic groups. *Ann. Probab.*, 41(5):3582–3605, 2013.
- [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.
- [Osi06] Denis V. Osin. Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems. *Mem. Amer. Math. Soc.*, 179(843):vi+100, 2006.
- [Pap95] P. Papasoglu. Strongly geodesically automatic groups are hyperbolic. *Invent. Math.*, 121(2):323–334, 1995.
- [Raf14] Kasra Rafi. Hyperbolicity in Teichmüller space. *Geom. Topol.*, 18(5):3025–3053, 2014.
- [Sam02] Andrea Sambusetti. Growth tightness of surface groups. *Expo. Math.*, 20(4):345–363, 2002.
- [Sis18] Alessandro Sisto. Contracting elements and random walks. *J. Reine Angew. Math.*, 742:79–114, 2018.
- [Vog15] Karen Vogtmann. On the geometry of outer space. *Bull. Amer. Math. Soc. (N.S.)*, 52(1):27–46, 2015.
- [Yam04] Asli Yaman. A topological characterisation of relatively hyperbolic groups. *J. Reine Angew. Math.*, 566:41–89, 2004.
- [Yan14] Wen-yuan Yang. Growth tightness for groups with contracting elements. *Math. Proc. Cambridge Philos. Soc.*, 157(2):297–319, 2014.
- [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,  
 CORNELL UNIVERSITY, 310 MALOTT HALL, ITHACA, NY, 14850, UNITED STATES,  
*Email address:* inhyeokchoi48@gmail.com