

# Counting and boundary limit theorems for representations of Gromov-hyperbolic groups

Stephen Cantrell<sup>1</sup> | Cagri Sert<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Chicago, Chicago, Illinois, USA

<sup>2</sup>Institut für Mathematik, Universität Zürich, Zurich, Winterthurerstrasse, Switzerland

## Correspondence

Stephen Cantrell, Department of Mathematics, University of Chicago, Chicago, IL 60637, USA.

Email: [scantrell@uchicago.edu](mailto:scantrell@uchicago.edu)

## Funding information

SNF, Grant/Award Number: Ambizione 193481

## Abstract

Given a Gromov-hyperbolic group  $G$  endowed with a finite symmetric generating set, we study the statistics of counting measures on the spheres of the associated Cayley graph under linear representations of  $G$ . More generally, we obtain a weak law of large numbers for subadditive functions, echoing the classical Fekete lemma. For strongly irreducible and proximal representations, we prove a counting central limit theorem with a Berry–Esseen type error rate and exponential large deviation estimates. Moreover, in the same setting, we show convergence of interpolated normalized matrix norms along geodesic rays to Brownian motion and a functional law of iterated logarithm, paralleling the analogous results in the theory of random matrix products. Our counting large deviation estimates address a question of Kaimanovich–Kapovich–Schupp. In most cases, our counting limit theorems will be obtained from stronger almost sure limit laws for Patterson–Sullivan measures on the boundary of the group.

## MSC 2020

20F67 (primary), 37D40, 60F05, 60F10 (secondary)

Dedicated to the memory of Émile Le Page.

© 2023 The Authors. *Proceedings of the London Mathematical Society* is copyright © London Mathematical Society. This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Contents

1. INTRODUCTION . . . . .	2
2. HYPERBOLIC GROUPS AND AUTOMATIC STRUCTURES . . . . .	12
3. LAW OF LARGE NUMBERS . . . . .	15
4. MARKOVIAN RANDOM MATRIX PRODUCTS . . . . .	21
5. LIMIT THEOREMS FOR THE STRONGLY MARKOV STRUCTURE . . . . .	37
6. LARGE DEVIATION THEOREMS . . . . .	43
7. WIENER PROCESS AND THE LAW OF THE ITERATED LOGARITHM ON THE BOUNDARY . . . . .	47
8. COUNTING CLT AND ERROR TERMS . . . . .	51
9. ON A QUESTION OF KAIMANOVICH–KAPOVICH–SCHUPP . . . . .	58
REFERENCES. . . . .	62

## 1 | INTRODUCTION

Let  $\mu$  be a probability measure on  $G = \mathrm{GL}_d(\mathbb{R})$  and  $(X_i)_{i \in \mathbb{N}}$  be a sequence of independent  $G$ -valued random variables with distribution  $\mu$ . Let  $Y_n$  denote the  $n$ th-step of the random product  $X_n \dots X_1$ . The theory of random matrix products is concerned with studying the asymptotic behavior of  $Y_n$ , for example, by investigating limit theorems (law of large numbers, central limit theorem [CLT], large deviations, etc.) for numerical quantities associated to matrices such as the operator norm  $\|Y_n\|$  or spectral radius. The most intricate part of the theory is when the probability measure  $\mu$  is finitely or countably supported say inside a countable group  $\Gamma < G$ . In that case, one has to deal with the possible singular behavior of the countable subgroup  $\Gamma$  inside the ambient group  $\mathrm{GL}_d(\mathbb{R})$ . After pioneering works of Furstenberg, Kesten [31, 32], and several others, significant progress was made by Le Page [50] in early 1980s; however many open questions still persist.

The theory of random matrix products provides a way to express asymptotic behavior of large elements of  $\Gamma$  in  $\mathrm{GL}_d(\mathbb{R})$ . Indeed, for a finitely supported probability measure  $\mu$  as above, the probabilistic description of the asymptotic behavior of  $Y_n$  is a problem of *symbolic counting*, that is, counting with certain multiplicities. A related but different way to study the asymptotic behavior of elements of  $\Gamma$ , perhaps more directly related to group  $\Gamma$  itself rather than its symbolic representation, would be to study statistics of asymptotics of *actual elements* of  $\Gamma$ . However, due to disparate algebraico-combinatorial structure of different countable groups  $\Gamma < \mathrm{GL}_d(\mathbb{R})$  such a general description is notoriously harder to obtain. Accordingly, such counting asymptotics results are much less developed compared to the theory of random matrix products.

In this article, we will be interested in describing *counting asymptotics* and *boundary limit laws* for representations of Gromov-hyperbolic groups. These include virtually free groups, cocompact isometry groups of negatively curved geodesic spaces, groups with small cancellation property, and so on. From another perspective, in some probabilistic models (e.g., random groups [41]), most finitely presented groups are Gromov-hyperbolic. We shall prove four main limit theorems.

- *Law of large numbers for subadditive functions*: This is of more general nature compared to the following results, it holds for any real-valued subadditive function on  $\Gamma$ .

The next results concern matrix representations of Gromov-hyperbolic groups, they hold under the standard (strong) irreducibility and proximality assumptions of random matrix products theory:

- *Exponential large deviation estimates for counting*: This one refines the aforementioned law of large numbers in the setting of matrix representations and addresses a question raised by Kaimanovich–Kapovich–Schupp [46, Problem 9.3]. Apart from representations, we also prove counting large deviation estimates for isometric actions on Gromov-hyperbolic spaces.
- *Counting CLT with Berry–Esseen type error term*.
- *Convergence of normalized interpolations along geodesic rays under a Patterson–Sullivan measure to the standard Brownian motion*: This one is of a different nature, it pertains to a measure on the boundary rather than counting. In fact, the first three points above also have corresponding boundary analogues that, beyond interest in themselves, serve as a tools to prove them.

Somewhat ironically, the key mechanism that will allow us to obtain these deterministic counting asymptotics is the inherent dynamical or probabilistic structure of the Gromov-hyperbolic groups. Indeed, as realized by Cannon [19] and Gromov [40], the geodesics on such a group can be coded by a finite state automaton. This makes it possible to approach the deterministic data of these groups by (a collection of) well-behaved stochastic processes, namely Markov chains. For example, for the last three results mentioned above, it enables us to employ probabilistic results of *Markovian random matrix products* (mainly due to Bougerol [10–12] and Guivarc’h [43]; we also develop some of them further) to the deterministic counting results. This transfer, however, requires handling some difficulties that we manage to do by, among others, elaborating on techniques developed by Calegari–Fujiwara [18] (generally) and Gekhtman–Taylor–Tiozzo [34] (for the CLT). The deterministic nature of our results, in particular the fact that we do not induce randomness using an external source (like a subshift of finite type [18, 59]) is of particular interest. We shall comment more on each of our results and on the past works below, let us now continue by stating our theorems and remarks more precisely.

Let  $\Gamma$  be a finitely generated group and  $S$  a generating set for  $\Gamma$ , all considered generating sets will be assumed to be finite and symmetric. The choice of  $S$  makes  $\Gamma$  into a metric space by considering the associated length function on  $\Gamma$ , namely  $|g|_S = \min\{n \in \mathbb{N} \mid s_1 \dots s_n = g, s_i \in S\}$  and for  $g, h \in \Gamma$  setting the (left) metric to be  $d_S(g, h) = |g^{-1}h|_S$ . Recall that for  $\Delta \geq 0$ , by a  $\Delta$ -hyperbolic metric space  $(M, d)$ , we understand a metric space such that for every  $x, y, z, o \in M$ , we have  $(x, y)_o \geq (x, z)_o \wedge (z, y)_o - \Delta$ , where  $(\cdot, \cdot)_o$  is the Gromov product given by  $(x, y)_o = \frac{1}{2}(d(x, o) + d(y, o) - d(x, y))$ . The group  $\Gamma$  is said to be *Gromov-hyperbolic* if there exists a real constant  $\Delta \geq 0$  and a generating set  $S$  such that the associated metric space is  $\Delta$ -hyperbolic. Given a generating set  $S \subseteq \Gamma$ , we write  $S_n$  for the sphere of radius  $n$  for the associated metric, namely  $S_n := \{g \in \Gamma : |g|_S = n\}$ . Finally, a Gromov-hyperbolic group  $\Gamma$  is said to be *nonelementary* if it is not virtually cyclic, that is, does not contain a cyclic subgroup of finite index.

## 1.1 | Convergence of subadditive spherical averages

A real-valued function  $\varphi$  on a group  $\Gamma$  is called *subadditive*, if for every  $g, h \in \Gamma$ , we have  $\varphi(gh) \leq \varphi(g) + \varphi(h)$ . The following is our first result.

**Theorem 1.1** (Weak law of large numbers for subadditive spherical averages). *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group endowed with a generating set  $S$  and  $\varphi : \Gamma \rightarrow \mathbb{R}$  is subadditive*

function on  $\Gamma$ . Then, there exists  $\Lambda \geq 0$  such that for any  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{\#S_n} \# \left\{ g \in S_n : \left| \frac{\varphi(g)}{n} - \Lambda \right| > \epsilon \right\} = 0.$$

In particular,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|g|_S=n} \frac{1}{\#S_n} \varphi(g) = \Lambda.$$

This result is due to Kaimanovich–Kapovich–Schupp [46] when  $\Gamma$  is a free group and  $S$  is a free generating set. We note that the first statement above is precisely a weak law of large numbers, whereas the second one corresponds to convergence in expectation (for a strong law, see Theorem 1.10). Notice also the curious analogy with the classical Fekete lemma that matches this convergence in expectation when  $\Gamma = \mathbb{N}$  and  $S = \{1\}$ .

Unlike our other results below where we will specialize to linear representations, the generality of subadditive functions in the previous result goes far beyond; we now briefly comment on various settings where such functions arise.

*Remark 1.2* (Examples of subadditive functions). Two large classes of subadditive functions contain the following.

- (1) (Seminorms on groups) Let  $H$  be any group endowed with a seminorm  $|\cdot|$  and  $\rho : \Gamma \rightarrow H$  a homomorphism (cf. [46]). The function  $\varphi(\gamma) := |\rho(\gamma)|$  is clearly a subadditive function on  $\Gamma$  and this construction encompasses many examples.
  - (1a) Already in the case  $H = \Gamma$ ,  $\rho = \text{id}$  and  $|\cdot|$  any length function on  $\Gamma$ , the previous theorem applied to  $\varphi(\cdot) = |\cdot|$  yields an asymptotic ratio  $\Lambda$  between  $|\cdot|_S$  and  $|\cdot|$ . Note that  $\Lambda > 0$  if, for example,  $|\cdot|$  comes from a finite generating set. More generally, let  $(X, d)$  be a metric space,  $o \in X$  and  $\Gamma \curvearrowright X$  by isometries. Then  $\varphi(\gamma) = d(\gamma \cdot o, o)$  is a subadditive function.
  - (1b) Let  $H = \text{GL}_d(\mathbb{R})$ ,  $\|\cdot\|$  an operator norm on the algebra  $\text{Mat}_d(\mathbb{R})$  of matrices and  $\rho : \Gamma \rightarrow H$  a representation. Then,  $\varphi(\gamma) := \log \|\rho(\gamma)\|$  is an example of a subadditive function.
- (2) (Quasi-morphisms) Another class of examples comes from the observation that Theorem 1.1 remains valid for any function  $\varphi'$  on  $\Gamma$  such that  $|\varphi - \varphi'|$  is bounded. In view of this, the previous result applies to any quasi-morphism (see [46, Theorem A]). For those, it is not hard to see that  $\Lambda = 0$ .

*Remark 1.3* (Possible extensions). Using different methods that rely on the topological flow introduced by Mineyev [55] and studied by Tanaka in [65], it might be possible to prove that Theorem 1.1 holds when we count with respect to some other hyperbolic metrics that are not necessarily word metrics (see [23, Theorem 3.12]). We have decided not to present the proof of this result as it is not clear how to obtain more refined counting limit laws below in this more general setting.

We note that Theorem 1.1 and its almost sure version (Theorem 1.10) generalize several previous works. For example, [33, Theorem 7.3] and [65, Theorem 7.4] follow from the particular case where  $\varphi$  is a displacement function associated to an isometric group action with additional requirements. It also generalizes (without error term) [21, Theorem 1.1]. See also Subsection 1.3.1.

It would be interesting to characterize when the constant  $\Lambda$  appearing in Theorem 1.1 is strictly positive. For a subadditive function  $\varphi$  coming from a seminorm (Remark 1.2(1)) one can typically say more, see Proposition 3.3. We will also see a characterization below in the case of strongly irreducible representations.

The rest of our counting results (except Theorem 1.7) concern finite-dimensional representations  $\Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  of Gromov-hyperbolic groups and we now specialize to this setting.

## 1.2 | Counting limit theorems for representations

Recall that a representation  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  is said to be *strongly irreducible* if there does not exist a finite collection of proper nontrivial subspaces of  $\mathbb{R}^d$  whose union is invariant under the action of  $\rho(\Gamma)$ . It is said to be *proximal* if there exists a sequence of elements  $g_n \in \rho(\Gamma)$  such that  $\frac{g_n}{\|g_n\|}$  converges to a rank-one linear transformation.

### 1.2.1 | Positivity of average growth rate

In what follows, whenever a representation  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  of a Gromov-hyperbolic group  $\Gamma$  (equipped with a generating set) is understood,  $\Lambda$  denotes the average growth rate given by applying Theorem 1.1 to  $\varphi(g) = \log \|\rho(g)\|$ . Clearly,  $\Lambda$  does not depend on the choice of the operator norm. The following result gives a characterization of when  $\Lambda$  is positive.

**Proposition 1.4.** *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group,  $S$  a generating set for  $\Gamma$  and  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  a strongly irreducible representation. Then the constant  $\Lambda \geq 0$  is strictly positive if and only if  $\rho(\Gamma)$  is not relatively compact in  $\mathrm{PGL}_d(\mathbb{R})$ .*

This result is ultimately a consequence of Furstenberg's result [31] on positivity of the top Lyapunov exponent for independent and identically distributed random products. However, for this statement, we additionally (need to) exploit the symmetry of the generating set because positivity of top Lyapunov exponent may fail for random products in  $\mathrm{GL}_d(\mathbb{R})$ .

Combined with Theorem 1.1, this result already implies that if such a  $\Gamma$  is Zariski-dense in a real semisimple linear Lie group  $G$ , the word metric  $d_S$  and any left- $G$ -invariant Riemannian metric  $d_G$  are Lipschitz equivalent when restricted to a large (i.e., full asymptotic density in the spheres  $S_n$ ) subset of  $\Gamma$ . We discuss this more in the large deviation part in Subsection 1.2.2 and in detail in Subsection 9.2.

### 1.2.2 | Exponential large deviation estimates

Establishing the next result was one of the earlier motivations of our work. In [46], Kaimanovich–Kapovich–Schupp asked whether exponential large deviation estimates can be obtained for free groups equipped with certain generating sets. The following therefore provides a class of such examples with considerably less restrictions on the underlying group and generating set (see Section 9).

**Theorem 1.5.** *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group,  $S$  a generating set and  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  a strongly irreducible and proximal representation. Then for every  $\epsilon > 0$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{1}{\#S_n} \# \left\{ g \in S_n : \left| \frac{\log \|\rho(g)\|}{n} - \Lambda \right| > \epsilon \right\} \right) < 0.$$

Here  $\Lambda > 0$  is the constant obtained from applying Theorem 1.1.

This result is analogous to a result of Le Page [50] (see [13, Theorem 6.2]) for independent and identically distributed random matrix products. A multidimensional version of this result (which is a consequence thereof) shows the uniqueness of maximum of the growth indicator function considered in [62]. See Subsection 9.1 where we discuss these in detail.

Theorem 1.5 will be deduced from an almost-sure version of it (with respect to geodesic rays following Patterson–Sullivan measure class on boundary) that we will discuss below (Theorem 1.11).

As we will explain in Subsection 9.2, it follows from Theorem 1.5 and positivity of  $\Lambda$  that, when  $\Gamma$  is a Zariski-dense subgroup of a real linear semisimple Lie group  $G$ , the word-metric  $d_S$  on  $\Gamma$  and left- $G$ -invariant Riemannian metric  $d_G$  on  $G$  coming from a Killing form are Lipschitz equivalent when restricted to an  $S$ -exponentially generic subset of  $\Gamma^\dagger$ .

**Theorem 1.6.** *Let  $\Gamma$  be a Zariski-dense, nonelementary Gromov-hyperbolic subgroup of a real semisimple linear Lie group  $G$ . Then, for every finite symmetric generating set  $S$  of  $\Gamma$  and constant  $\epsilon > 0$ , there exists a subset  $T_\epsilon$  of  $\Gamma$  with the property that*

$$\frac{\#(S_n \setminus (T_\epsilon \cap S_n))}{\#S_n} = O(e^{-\alpha n})$$

for some  $\alpha > 0$ , and there exists a constant  $\Lambda = \Lambda(S) > 0$  such that for every  $n \in \mathbb{N}$  and  $g \in S_n \cap T_\epsilon$ , we have

$$n(\Lambda - \epsilon) \leq d_G(g, \mathrm{id}) \leq n(\Lambda + \epsilon).$$

This is restated and proved in Subsection 9.2, see Corollary 9.3. It may be tempting to try to prove this result (or the previous one) using a random walk approach. However, to do this, one would need to construct a probability measure  $\mu$  on  $S$  for which we have the equality  $h_\mu = \ell_\mu \log \lambda$  in the fundamental inequality  $h_\mu \leq \ell_\mu \log \lambda$  of Guivarc’h (here,  $h_\mu$  is the asymptotic (Avez) entropy of  $\mu$ ,  $\ell_\mu$  is its drift and  $\lambda$  exponential growth rate of  $S$ -spheres in  $\Gamma$ , see, e.g., [39]). The reason for this is that a probability measure  $\mu$  with  $h_\mu < \ell_\mu \log \lambda$  will only see an exponentially small part of the spheres of  $S$ . On the other hand, as shown in [39, Theorem 1.3], the equality case  $h_\mu = \ell_\mu \log \lambda$  is very rigid and forces the ambient group to be virtually free.

The following result establishes exponential counting large deviation estimates in another setting, that of isometric actions on Gromov-hyperbolic spaces. This setting has recently attracted much attention both from probabilistic [3, 4, 7, 14, 38, 54] and counting [18, 21–23, 33, 34, 68]

<sup>†</sup> Notice that in general even if  $\Gamma$  is a (nonuniform) lattice in  $G$ , one cannot hope to have this Lipschitz equivalence on whole of  $\Gamma$ . On the other hand, by an important result of Lubotzky–Mozes–Raghunathan, such a global Lipschitz equivalence holds for higher rank irreducible lattices [52, 53] that are of course not Gromov-hyperbolic.

perspectives. To state our result, recall that the action of a group  $\Gamma$  on a Gromov-hyperbolic space  $H$  by isometries is said to be *nonelementary* if there exists  $\gamma_1, \gamma_2 \in \Gamma$  acting as loxodromic elements (see Subsection 4.5) with disjoint pairs of fixed points on the Gromov boundary of  $H$ .

**Theorem 1.7.** *Let  $\Gamma$  be a Gromov-hyperbolic group,  $S$  a generating set of  $\Gamma$  and  $(H, d)$  a geodesic Gromov-hyperbolic space and  $o \in H$  a basepoint. Suppose that  $\Gamma$  acts on  $H$  by isometries and that the action is nonelementary. Then, there exists a constant  $\Lambda > 0$  such that for every  $\epsilon > 0$ , we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{1}{\#S_n} \# \left\{ g \in S_n : \left| \frac{d(g \cdot o, o)}{n} - \Lambda \right| > \epsilon \right\} \right) < 0.$$

This result is analogous to the main result of the recent work [14] (see also [38]) in the setting of independent and identically distributed random walks on Gromov-hyperbolic spaces. As for Theorem 1.5, we will deduce Theorem 1.7 from a corresponding boundary limit theorem (Theorem 6.4) for Patterson–Sullivan measures. To prove the latter, we crucially make use of the large deviation estimates that we develop from the work of Benoist–Quint [6], for cocycles over random products of group elements in Markovian dependence (these tools also serve us in the Berry–Esseen estimate as explained above). We defer the statement of Theorem 6.4 to Section 6.

*Remark 1.8.* Recently, Cantrell and Tanaka [23, Theorem 4.23] proved a global large deviation principle that implies Theorem 1.7 when  $\Gamma$  acts on  $H = \Gamma$  by multiplication and  $d$  is a left-invariant hyperbolic metric that is quasi-isometric to a word metric.

### 1.2.3 | Central limit theorem with Berry–Esseen type error term

Equipped with a law of large numbers, we now state the first refined limit theorem for counting statistics in representations:

**Theorem 1.9.** *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group,  $S$  a generating set and  $\rho : \Gamma \rightarrow \text{GL}_d(\mathbb{R})$  a strongly irreducible and proximal representation. Fix an operator norm  $\| \cdot \|$  on  $\text{Mat}_d(\mathbb{R})$ . Then, there exists a constant  $C > 0$  such that for every  $t \in \mathbb{R}$*

$$\left| \frac{1}{\#S_n} \# \left\{ g \in S_n : \frac{\log \|\rho(g)\| - n\Lambda}{\sqrt{n}} \leq t \right\} - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^t e^{-s^2/2\sigma^2} ds \right| \leq \frac{C \log n}{\sqrt{n}},$$

where  $\Lambda > 0$  is as in Theorem 1.1 and  $\sigma^2 > 0$  are strictly positive constants.

In the sequel, whenever a strongly irreducible and proximal representation is fixed,  $\sigma^2$  will denote the variance in the above CLT.

The proof of this result requires several ingredients. We first prove a Berry–Esseen CLT for the norm of Markovian random matrix products (Theorem 4.8) based on the analogous result of Bougerol [11] for the norm cocycle. To do this, we use an idea due to Xiao–Grama–Liu from their recent work [67]. The core of the argument is based on large deviation estimates from Benoist–Quint [8] that we develop (Theorem 4.3) for the Markovian setting by elaborating on other work due to Benoist–Quint [6] that concerns large deviation estimates for cocycles. Equipped with



these results as well as techniques developed by Calegari–Fujiwara [18], we employ a quantitative version of an argument from recent work of Gekhtman–Taylor–Tiozzo [34] to carry out our proof.

### 1.3 | Boundary limit theorems for representations

As previously mentioned, limit theorems with respect to Patterson–Sullivan measures on the boundary will play a key role in our work: on the one hand, we will prove new results for them (such as Theorem 1.12 on convergence to the Brownian motion), on the other hand, they will be used to prove the counting law of large numbers (Theorem 1.1) and large deviation theorems (Theorems 1.5 and 1.7). More specifically, we will describe the growth rate of subadditive functions (and the log-norm function for linear representations) along Patterson–Sullivan typical geodesic rays. We achieve this by comparing Markov measures on a Cannon coding with Patterson–Sullivan measures on the boundary of our considered group. Along with techniques from ergodic theory and geometric group theory, this will allow us to translate results concerning Markovian random products to asymptotic behavior along Patterson–Sullivan typical geodesic rays.

In the statements of our boundary limit theorems (and throughout this work), we will consider the boundary  $\partial\Gamma$  of  $\Gamma$  equipped with generating set  $S$  to be the collection of  $|\cdot|_S$  geodesic rays up to the usual bounded distance equivalence. For  $\xi \in \partial\Gamma$ , we use the notation  $\xi_n \rightarrow \xi$  to indicate that  $(\xi_n)_{n \in \mathbb{N}}$  is a geodesic ray in the class of  $\xi$  (see Section 2).

#### 1.3.1 | Law of large numbers for the Patterson–Sullivan measure class

Here is the strong law underlying Theorem 1.1:

**Theorem 1.10** (Strong law of large numbers). *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group endowed with a generating set  $S$  and  $\varphi : \Gamma \rightarrow \mathbb{R}$  is subadditive function on  $\Gamma$ . Let  $\nu$  be a probability measure on  $\partial\Gamma$  in the Patterson–Sullivan measure class. Then, there exists a constant  $\Lambda \geq 0$  such that*

$$\lim_{n \rightarrow \infty} \frac{\varphi(\xi_n)}{n} = \Lambda$$

for  $\nu$ -almost every  $\xi \in \partial\Gamma$  and every representative  $\xi_n \rightarrow \xi$ .

This result generalizes [46, Theorem A]. The reason we call it a strong law is that, roughly speaking, the Patterson–Sullivan measures can be viewed as the law of a process for which the uniform counting measures correspond to finite time distributions. This is also the spirit of the deduction of Theorem 1.1 from the previous result.

#### 1.3.2 | Large deviations for Patterson–Sullivan measures

The quantitative analogue of Theorem 1.10 for linear representations is the following result.



**Theorem 1.11.** *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group,  $S$  a generating set and  $\rho : \Gamma \rightarrow \text{GL}_d(\mathbb{R})$  a strongly irreducible and proximal representation. Let  $\nu$  be a Patterson–Sullivan measure on  $\partial\Gamma$  for the  $S$  word metric and  $\Lambda \geq 0$  be the constant from Theorem 1.10. Then, for any  $\epsilon > 0$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \nu \left( \xi \in \partial\Gamma : \text{for all } \xi_m \rightarrow \xi \text{ with } \xi_0 = \text{id}, \left| \frac{\log \|\rho(\xi_n)\|}{n} - \Lambda \right| > \epsilon \right) < 0.$$

Here, when we say the  $\nu$  is a Patterson–Sullivan measure, we mean that it is constructed as a weak limit as in (1.2) or (2.1). We note that any two measures obtained in this way are mutually absolutely continuous and their densities are bounded away from 0 (and infinity).

The proof makes use of Bougerol’s results [11] that are translated to the group theoretic setting using techniques due to Calegari–Fujiwara [18] and extensions of these techniques due to Cantrell [22]. The scheme of proof, somewhat common to the next Theorem 1.12, is expounded in Subsection 1.4.

### 1.3.3 | Convergence to the Wiener measure and law of iterated logarithm

We now turn to our last result that is an invariance principle and functional law of iterated logarithm (LIL) with respect to Patterson–Sullivan measures. We first need some notation. Suppose  $\Gamma$  is a Gromov-hyperbolic group endowed with a generating set  $S$  and that  $\rho : \Gamma \rightarrow \text{GL}_d(\mathbb{R})$  is a strongly irreducible, proximal representation. Let  $C([0, 1])$  denote the continuous real valued functions on  $[0, 1]$  equipped with the Borel  $\sigma$ -algebra for the topology of uniform convergence. We define a sequence of random variables  $(S_n)_{n \in \mathbb{N}}$  on  $\partial\Gamma$  taking values in  $C([0, 1])$  as follows. For each  $\xi \in \partial\Gamma$ , integer  $n \geq 1$  and  $t \in [0, 1]$ , we define  $S_n \xi(t)$  to be

$$\min_{\xi_m \rightarrow \xi} \frac{1}{(n\sigma^2)^{1/2}} (\log \|\rho(\xi_{\lfloor nt \rfloor})\| - nt\Lambda + (nt - \lfloor nt \rfloor)(\log \|\rho(\xi_{\lfloor nt \rfloor + 1})\| - \log \|\rho(\xi_{\lfloor nt \rfloor})\|)), \tag{1.1}$$

where  $\Lambda$  and  $\sigma^2 > 0$  are the mean and variance from Theorem 1.9. The reason we consider the minimum over the set of representatives is only practical, it allows to define the random variables  $S_n$  on  $\partial\Gamma$ ; replacing min with max will not alter the asymptotic behavior (and hence the results to follow) because any two representatives of a boundary point  $\xi \in \partial\Gamma$  stay at bounded distance. We denote by  $\mathcal{W}$  the Wiener measure on  $C([0, 1])$ . Recall that this is the distribution of the standard Brownian motion  $B(\cdot) \in C([0, 1])$  which is characterized [48] by  $B(0) \stackrel{a.s.}{=} 0$ , and for every  $p \in \mathbb{N}$  and reals  $0 = t_0 < t_1 < \dots < t_p$ , the real-valued random variables  $B(t_1), B(t_2) - B(t_1), \dots, B(t_p) - B(t_{p-1})$  are independent and distributed with the Gaussian distribution, respectively,  $\mathcal{N}(0, t_i - t_{i-1})$ .

We will prove the convergence to Wiener measure with respect to the Patterson–Sullivan measure obtained as the weak limit

$$\nu = \lim_{n \rightarrow \infty} \frac{\sum_{|g|_S \leq n} \lambda^{-|g|_S} \delta_g}{\sum_{|g|_S \leq n} \lambda^{-|g|_S}}. \tag{1.2}$$

Here  $\lambda \in (1, \infty)$  denotes the exponential growth rate of the cardinality of  $S_n = \{g \in \Gamma : |g|_S = n\}$ . The fact that the limit (1.2) exists will be explained in Subsection 2.1. We prove the following result.

**Theorem 1.12.** *Let  $\Gamma$  be a nonelementary Gromov-hyperbolic group,  $S$  a generating set and  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  a strongly irreducible and proximal representation. Let  $\nu$  be the Patterson–Sullivan measure defined in (1.2). Then,*

- (1) *under  $\nu$ , the sequence  $(S_n)_{n \in \mathbb{N}}$  of  $C([0, 1])$ -valued random variables converges in distribution to  $\mathcal{W}$ ; and*
- (2) *for  $\nu$ -almost every  $\xi \in \partial\Gamma$ , the set of limit points of the sequence  $(\frac{S_n \xi}{2 \log \log n})_{n \in \mathbb{N}}$  of elements of  $C([0, 1])$  is equal to the following compact subset of  $C([0, 1])$ :*

$$\left\{ f \in C([0, 1]) : f \text{ is absolutely continuous, } f(0) = 0, \int_0^1 f'(t)^2 dt \leq 1 \right\}.$$

Two immediate corollaries of this result are the classical CLT and LIL with respect to the Patterson–Sullivan measure  $\nu$  (Corollary 7.2).

## 1.4 | Outline of the arguments

We briefly outline the overarching argument used to prove Theorem 1.12 which is also valid to some extent for the proof of Theorem 1.11 (see below for other limit theorems).

- (i) We begin by introducing multiple Markov chains based on the Cannon coding for our group  $\Gamma$  and generating set  $S$ .
- (ii) We formulate and, in some cases, further develop Bougerol’s results [10, 11] for random matrix products in Markovian dependence.
- (iii) Using work of Goldsheid–Margulis [36] and Gouëzel–Mathéus–Maucourant [39], we show that our assumptions on the representations (strongly irreducible and proximal) allow us to apply the results of (ii) to the Markovian products introduced in (i).
- (iv) We then use an argument of Calegari–Fujiwara [18] to show that the means  $\Lambda$  and variances  $\sigma^2$  coming from the limit theorems obtained in (iii) of different Markovian products introduced in (i) coincide (and equal the limiting average  $\Lambda$  obtained from Theorem 1.1).
- (v) We compare the stationary distributions on the Markov chains to a Patterson–Sullivan measures on the boundary of the group. With some additional work, this allows us to prove that along geodesic rays in the Cayley graph of  $(\Gamma, S)$ , the log-norm function satisfies the corresponding limit theorem with respect to Patterson–Sullivan measure on the Gromov boundary  $\partial\Gamma$ .

For counting limit theorems (Theorems 1.1, 1.5, and 1.7), we ultimately use the boundary results obtained in (v) above together with regularity estimates on Patterson–Sullivan measures to get counting limit theorems.

As discussed above, our tactic for proving counting CLT with error term (Theorem 1.9) is a little bit different (without passing by a boundary limit theorem to optimize the Berry–Esseen error term), see a more detailed description in Section 8.

## 1.5 | Previous works

Here we briefly comment on some previous related works in the literature.

In [46] (see also [47]), Kaimanovich–Kapovich–Schupp study generic asymptotic behavior (in the sense of Theorems 1.1 and 1.10) of elements in countable groups  $\Gamma$  where genericity is understood with respect to the uniform counting measure in a free group  $F$  (with a free generating set) when  $\Gamma$  is seen as a quotient of  $F$ . This point of view lies in between the two extremes, namely the symbolic counting point of view of probability theory (i.e., group invariant random walks on groups) and our deterministic counting viewpoint. In vague terms, our approach agree with that of Kapovich–Kaimanovich–Schupp when the underlying group is a free group and generating set is free<sup>†</sup> and these two together agree with the probabilistic (independent and identically distributed random walks) approach when the underlying algebraic object is a free semigroup.

Coming back to counting asymptotics, there has recently been significant interest in counting limit theorems on hyperbolic groups, see, for example, [18, 21, 23, 25, 33, 34, 44, 45, 58, 68]. In some of these works, techniques from thermodynamic formalism (tracing back to [15, 64], see also [49]) are used. Although these techniques are powerful and allow for stronger results to be obtained, they usually require strong assumptions on the studied potentials. In this work, our assumptions are too weak to allow us to apply techniques from thermodynamic formalism. In others, including ours, ideas from Markov chain or random walk theory (in a sense initiated in this context by [18]) are used instead of thermodynamic techniques. For example, in [34] the authors deduce a (qualitative) counting CLT for displacement functions on Gromov-hyperbolic spaces<sup>‡</sup> from a CLT for centerable cocycles. Using ideas of Benoist–Quint [6] relying on solving a cohomological equation, it might be possible to do so in our setting as well. However, our approach relying instead on results of Bougerol coming from analytic perturbation theory, yields more quantitative results (such as the Berry–Esseen bounds). It was indeed one of our goals to get quantitative results as it seems particularly in line with the spirit of counting problems.

Finally, we also mention that in the upcoming work with Cipriano and Dougall, we obtain more precise limit laws for the both the spectral radius and norm potentials under the assumption that our representation is Anosov (or dominated). In this setting, we will be able to exploit ideas from thermodynamic formalism.

### 1.5.1 | Further directions

In this work, we restricted our attention to Gromov-hyperbolic groups. Our approach relies on the existence of a nice combinatorial structure (Cannon coding) and stochastic results on this structure. Various generalizations of the notion of the Cannon coding have been studied, both from a combinatorial perspective [20]<sup>§</sup> and geometric perspective [34]. It seems possible to find examples of nonhyperbolic groups equipped with certain specific (in some cases abstract) generating sets (see [34]) for which our counting results will hold. In some settings (e.g., relatively hyperbolic groups) it may also be possible to find analogues of our boundary limit theorems. It would be interesting to characterize the widest class of groups (equipped with any generating set) for which our results hold.

<sup>†</sup> Accordingly, our Theorem 1.10 generalizes [46, Theorem A.1].

<sup>‡</sup> See Section 9 for a consequence of our large deviation results for the displacement function on symmetric spaces of noncompact type.

<sup>§</sup> See [46] for some considerations toward this direction, but the results therein do not readily yield counting estimates.

## 2 | HYPERBOLIC GROUPS AND AUTOMATIC STRUCTURES

### 2.1 | Gromov-hyperbolic groups

Let  $\Gamma$  be a finitely generated group and  $S$  a (finite, symmetric) generating set. As in the introduction, for  $g, h \in \Gamma$ ,  $|g|_S$  denotes the word length of  $g$  with respect to  $S$  and  $d_S(g, h) = |g^{-1}h|_S$  defines a left-invariant metric on  $\Gamma$ . The Gromov product of  $g, h \in \Gamma$  is defined as  $\langle g, h \rangle = \frac{1}{2}(|g|_S + |h|_S - |g^{-1}h|_S)$ . The group  $\Gamma$  is said to be Gromov-hyperbolic if  $(\Gamma, d_S)$  is a Gromov-hyperbolic metric space. We recall that a metric space  $(H, d)$  is said to be Gromov-hyperbolic if there exists  $\Delta > 0$  such that for every  $o, x, y, z \in H$ ,

$$(x, z)_o \geq \min\{(x, y)_o, (y, z)_o\} - \Delta,$$

where  $(x, y)_o := \frac{1}{2}(d(x, o) + d(y, o) - d(x, y))$  denotes the Gromov-product. Although the constant  $\Delta > 0$  may depend on the generating set  $S$ , Gromov-hyperbolicity of  $\Gamma$  does not depend on  $S$ .

Fix a Gromov-hyperbolic group  $\Gamma$  and a generating set  $S$ . A geodesic ray is a sequence of elements  $\xi_n \in \Gamma$  such that  $|\xi_n^{-1}\xi_m|_S = m - n$  for each  $m, n \in \mathbb{N}$  with  $m \geq n$ . The Gromov boundary  $\partial\Gamma$  is the set of equivalence classes of geodesic rays where two rays  $\xi$  and  $\xi'$  are equivalent if  $\sup_{n \geq 1} |\xi_n^{-1}\xi'_n|_S$  is finite. As the action of  $\Gamma$  on itself by left-multiplication is by isometries (with respect to  $d_S$ ), the natural action on the set of geodesic rays factors through this equivalence relation and defines an action of  $\Gamma$  on its Gromov boundary  $\partial\Gamma$ . It is well-known that the set  $\Gamma \cup \partial\Gamma$  carries a compact metrizable topology extending the (discrete) topology of  $\Gamma$  such that  $\Gamma$  is open and dense in  $\Gamma \cup \partial\Gamma$  and the  $\Gamma$ -action is by homeomorphisms.

We can extend the Gromov product to  $\partial\Gamma \times \Gamma$  by setting

$$\langle \xi, g \rangle = \sup\{\liminf_{n \rightarrow \infty} \langle \xi'_n, g \rangle : \xi'_n \rightarrow \xi\},$$

where the supremum is taken over geodesic rays  $\xi'_n$  with  $\xi'_n \rightarrow \xi$ ; the latter notation denotes the fact that  $\xi'_n$  is a geodesic ray in the equivalence class corresponding to  $\xi$ . Using this extended Gromov product, for  $R > 0$  and  $g \in \Gamma$ , we define the  $R$ -shadow based at  $g$  to be the following subset of  $\partial\Gamma$ :

$$O(g, R) = \{\xi \in \partial\Gamma : \langle \xi, g \rangle > |g|_S - R\}$$

To prove our law of large numbers (Theorem 1.1), a key ingredient will be the study of the growth rate of subadditive functions along typical geodesic rays in  $\Gamma$ . In particular, we will be interested in the behavior of our functions along Patterson–Sullivan typical geodesic rays. Recall that a Patterson–Sullivan measure for the length function  $|\cdot|_S$  on  $\Gamma$  is obtained as a weak limit of the following sequence of measures on the compact  $\Gamma \cup \partial\Gamma$

$$\frac{\sum_{g \in \Gamma} \lambda^{-s|g|_S} \delta_g}{\sum_{g \in \Gamma} \lambda^{-s|g|_S}} \quad (2.1)$$

as  $s \searrow 1$  where  $\lambda > 1$  is the exponential growth rate of the cardinality of  $S_n = \{g \in \Gamma : |g|_S = n\}$ . Alternatively, we can obtain a Patterson–Sullivan measure as the weak limit of the

sequence

$$\frac{\sum_{|g|_S \leq n} \lambda^{-|g|_S} \delta_g}{\sum_{|g|_S \leq n} \lambda^{-|g|_S}} \tag{2.2}$$

as  $n \rightarrow \infty$  (see [18, section 4]). Any measure  $\nu$  constructed using either of the above two methods yields a Radon measure supported on  $\partial\Gamma$  such that the  $\Gamma$  action preserves its measure class and is ergodic. In this setting, ergodic means that  $\Gamma$ -invariant subsets of  $\partial\Gamma$  have either full or null  $\nu$ -measure. An important property exhibited by Patterson–Sullivan measures is the so-called quasiconformal property: for each  $R > 0$  sufficiently large, there exists a constant  $C > 1$  depending only on  $R$  and the hyperbolicity constant of  $\Gamma$  such that

$$C^{-1} \lambda^{-|g|_S} \leq \nu(O(g, R)) \leq C \lambda^{-|g|_S} \tag{2.3}$$

for all  $g \in \Gamma$ .

Before we move on to discuss the strongly Markov structure of hyperbolic groups, we record a basic property of subadditive functions that we will use implicitly throughout our work. Recall that the left and right word metrics associated to a generating set  $S$  on  $\Gamma$  are

$$d_S(g, h) = |g^{-1}h|_S \quad \text{and} \quad d_R(g, h) := |gh^{-1}|_S.$$

We will repeatedly (and sometimes implicitly) use the fact that subadditive functions  $\varphi : \Gamma \rightarrow \mathbb{R}$  are Lipschitz in these metrics as noted in the next result.

**Lemma 2.1.** *Fix a finite symmetric generating set  $S$  for  $\Gamma$  and let  $\varphi : \Gamma \rightarrow \mathbb{R}$  be subadditive. Then,  $\varphi$  is Lipschitz in the left and right word metrics.*

*Proof.* It suffices to show that there exist a constant  $C > 0$  such that

$$|\varphi(g) - \varphi(sg)| \leq C \quad \text{and} \quad |\varphi(g) - \varphi(gs)| \leq C$$

for all  $g \in \Gamma$  and  $s \in S$ . It follows easily from the definition of subadditivity that  $C = \max_{s \in S} |\varphi(s)|$  works. □

## 2.2 | Markov structure of Gromov-hyperbolic groups

It was realized by Cannon [19] that certain Kleinian groups enjoy a strong coding property: the elements of metric spheres in the Cayley graph can be bijectively represented by admissible words of corresponding length in a finite automaton (which we will refer to as strongly Markov property, see Definition 2.2). It was indicated by Gromov [40] and proved by Coornaert–Delzant–Papadopoulos [27] and Ghys–de la Harpe [35] that general Gromov-hyperbolic groups with arbitrary finite generating sets enjoy the strongly Markov property. We now discuss this crucial property that will allow us to associate a subshift of finite type to a Gromov-hyperbolic group equipped with a generating set.

**Definition 2.2.** A group  $\Gamma$  is strongly Markov if given any generating set  $S$  for  $\Gamma$ , there exists a finite directed graph  $\mathcal{G}$  with vertex set  $V$  and directed edge set  $E \subset V \times V$  that exhibit the following properties.

- (i)  $V$  contains a vertex  $*$  such that  $(x, *)$  does not belong to  $E$  for any  $x \in V$ .
- (ii) There exists a labeling  $\lambda : E \rightarrow S$  such that the map sending a path (starting at  $*$ ) with concurrent edges  $(*, x_1), (x_1, x_2), \dots, (x_{n-1}, x_n)$  to the group element  $\lambda(*, x_1)\lambda(x_1, x_2) \dots \lambda(x_{n-1}, x_n)$ , is a bijection.
- (iii) The above bijection preserves word length; if  $|g| = n$ , then the finite path corresponding to  $g$  has length  $n$ .

To simplify notation later on, we augment the above strongly Markov structure by introducing an additional vertex labeled 0 to  $V$ . We also add directed edges from every vertex  $x \in V$  to 0 and define  $\lambda(x, 0) = \text{id}$  (the identity in  $\Gamma$ ) for every  $x \in V$ . We will assume that every strongly Markov structure has been augmented in this way and will abuse notation by labeling the augmented structure, its edge and vertex set by  $\mathcal{G}$ ,  $V$  and  $E$ , respectively. This directed graph  $\mathcal{G}$  allows us to introduce a subshift of finite type as we now explain.

### 2.2.1 | Shift spaces

Let  $A$  be a  $k \times k$  matrix consisting of zeros and ones. We use the notation  $A_{i,j}$  to denote the  $(i, j)$ th entry of  $A$ . The subshift of finite type associated to  $A$  is the space

$$\Sigma_A = \{(x_n)_{n=0}^\infty : x_n \in \{1, 2, \dots, k\}, A_{x_n, x_{n+1}} = 1, n \in \mathbb{Z}_{\geq 0}\}.$$

Given  $x$  in  $\Sigma_A$  we write  $x_n$  for the  $n$ th coordinate of  $x$ . The shift map  $\sigma : \Sigma_A \rightarrow \Sigma_A$  sends  $x$  to  $y = \sigma(x)$  where  $y_n = x_{n+1}$  for all  $n \in \mathbb{Z}_{\geq 0}$ .

The mixing properties of  $(\Sigma_A, \sigma)$  are determined by the structure of the matrix  $A$ .

**Definition 2.3.** We say that a  $k \times k$  zero-one matrix  $A$  is irreducible if for every  $(i, j)$  ( $i, j \in \{1, 2, \dots, k\}$ ), there exists  $n \in \mathbb{N}$  such that  $(A^n)_{i,j} > 0$ . We say that  $A$  is primitive (irreducible and aperiodic) if there exists  $n \in \mathbb{N}$  such that  $(A^n)_{i,j} > 0$  for all  $i, j$ .

It is a standard fact that if  $A$  is irreducible then  $(\Sigma_A, \sigma)$  is (topologically) transitive, and if  $A$  is primitive then  $(\Sigma_A, \sigma)$  is mixing. Further, if  $A$  is irreducible then there exists a natural number  $p \geq 1$  known as the period of  $A$  such that the alphabet  $\{1, \dots, k\}$  of  $A$  is partitioned into  $p$  disjoint subsets  $A_i$  and  $\Sigma_A$  has a cyclic decomposition

$$\Sigma_A = \bigsqcup_{k=0}^{i-1} \Sigma_A(i),$$

where  $\Sigma_A(i)$  is the subset of  $\Sigma_A$  starting with elements from  $A_i$ . The shift map  $\sigma : \Sigma_A \rightarrow \Sigma_A$  sends  $\Sigma_A(i)$  to  $\Sigma_A(i+1)$  where  $i, i+1$  are taken modulo  $p$  and for each  $i = 0, \dots, p-1$  the subshifts  $(\Sigma_A(i), \sigma^p)$  are mixing.

### 2.2.2 | Shift space associated to a Markov structure

Suppose now that  $\mathcal{G}$  is a strongly Markov structure associated to a Gromov-hyperbolic group  $\Gamma$  and generating set  $S$ . We can describe  $\mathcal{G}$  using a zero-one matrix  $A$ : we label the vertices of  $\mathcal{G}$  by  $0, *, 1, 2, \dots, k \in \mathbb{N}$  (where  $0$  and  $*$  are distinguished vertices described above) and set  $A_{i,j} = 1$  if and only if there is a directed edge from vertex  $i$  to vertex  $j$  and otherwise we set  $A_{i,j} = 0$ . We can then construct a subshift of finite type  $\Sigma_A$  as described in the previous paragraph. We will write  $A'$  for the matrix obtained from  $A$  by discarding the row and column corresponding to the vertex  $0$  and  $A''$  for the one where we also discard the vertex  $*$ .

We will write  $\Sigma_A^0$  for the collection of sequences in  $\Sigma_A$  that contains an occurrence of  $0$ . Note that, by construction, if a sequence  $(x_n)_{n=0}^\infty$  has  $x_k = 0$  for some  $k$  then  $x_l = 0$  for all  $l \geq k$ . We will use the notation  $(x_0, \dots, x_{n-1}, \dot{0})$  to express sequences that start with the vertices  $x_0, x_1, \dots, x_{n-1}$  and then end with infinitely many zeros. Note that  $\Sigma_A^0$  is dense in  $\Sigma_A$  when  $\Sigma_A$  is endowed with the restriction of the product topology on  $V^\mathbb{N}$ . We define a map  $i : \Gamma \rightarrow \Sigma_A^0$  by  $i(g) = (*, x_1, \dots, x_n, \dot{0})$  where  $(*, x_1, \dots, x_n, \dot{0})$  is the unique sequence belonging to  $\Sigma_A$  such that  $g = \lambda(*, x_1)\lambda(x_1, x_2) \cdots \lambda(x_{n-1}, x_n)$  (and  $|g|_S = n$ ).

For certain hyperbolic groups and generating sets (i.e., for a free group equipped with a free generating set) one can find a strongly Markov structure  $\mathcal{G}$  such that the corresponding matrix  $A''$  is primitive. However, for general hyperbolic groups and generating sets it is not known whether it is always possible to find a Markov structure such that the matrix  $A''$  is primitive or even irreducible. After relabeling (i.e., permuting) the columns and rows of  $A''$ , we may assume that  $A''$  has the form

$$A'' = \begin{pmatrix} B_{1,1} & 0 & \dots & 0 \\ B_{2,1} & B_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_{m',1} & B_{m',2} & \dots & B_{m',m'} \end{pmatrix},$$

where the matrices  $B_{i,i}$  are irreducible. The matrices  $B_{i,i}$  are known as the irreducible components of  $A''$  and the corresponding vertex sets in  $\mathcal{G}$  are the irreducible components of  $\mathcal{G}$ . By property (3) in Definition 2.2, it is easy to see that the spectral radius of each  $B_{i,i}$  is bounded above by the growth rate  $\lambda$  of the group  $\Gamma$ . Moreover, by the same token, there must be at least one component that has  $\lambda$  as an eigenvalue. We call an irreducible component *maximal* if the corresponding matrix  $B_{i,i}$  has spectral radius  $\lambda$ . We relabel the irreducible components so that maximal components correspond to  $B_{i,i}$  for  $i = 1, \dots, m$ , which we will denote as  $B_i$ . An important property of  $\mathcal{G}$  is that the maximal components of  $\mathcal{G}$  are disjoint. That is, there does not exist a path in  $\mathcal{G}$  from one maximal component to another. This is a consequence of a result of Coornaert [26] which asserts that for a nonelementary hyperbolic group (and any generating set  $S$ ) the growth of  $\#S_n$  is purely exponential, that is, for  $\Gamma, S$  as above there exist  $C > 1$  and  $\lambda > 1$  such that for all  $n \in \mathbb{Z}_{\geq 0}$ .

$$C^{-1}\lambda^n \leq \#S_n \leq C\lambda^n. \tag{2.4}$$

## 3 | LAW OF LARGE NUMBERS

The goal of this section is to prove Theorems 1.1 and 1.10. The former is a general (weak) law of large numbers for subadditive functions on hyperbolic groups, whereas the latter can be seen as



the corresponding strong law that asserts the existence of a common growth rate of a subadditive function along almost every geodesic with respect to the Patterson–Sullivan measure. The preliminary material introduced in Subsection 3.1 will serve as a tool throughout the article.

### 3.1 | Markov measures for Gromov-hyperbolic groups

We summarize here a construction of a Markov chain on the strongly Markov structure of a Gromov-hyperbolic group and its connection with the Patterson–Sullivan measure (both due to Calegari–Fujiwara [18] in this setting). We also include some further related observations from Cantrell [22]; other more specific ones will be included/proven in later sections where they are needed.

#### 3.1.1 | Patterson–Sullivan measures seen in the strongly Markov structure

We keep the notation from Section 2: let  $\Gamma$  be a nonelementary Gromov-hyperbolic group endowed with a generating set  $S$ . Fix a strongly Markov structure  $\mathcal{G}$ . Let  $\lambda > 1$  be the exponential growth rate of  $\Gamma$  with respect to  $S$ . Denote by  $\nu$  the Patterson–Sullivan probability measure (see [18, Definition 4.14]) obtained as the limit of the sequence of probabilities  $\nu_n$  on  $\Gamma \cup \partial\Gamma$ , where

$$\nu_n := \frac{\sum_{|g|_S \leq n} \lambda^{-|g|_S} \delta_g}{\sum_{|g|_S \leq n} \lambda^{-|g|_S}}. \quad (3.1)$$

Let  $Y = [*]$  be the (cylinder) set of sequences  $(x_n)$  in  $\Sigma_A$  that starts with the symbol  $*$ , that is,  $x_0 = *$ . Let  $Y_n \subseteq Y \cap \Sigma_A^0$  be the subset of  $Y$  consisting of sequences  $(x_m)$  such that  $x_m = 0$  for every  $m \geq n + 1$ . In view of Definition 2.2, the set  $Y_n$  is in bijection with the ball of radius  $n$  and hence the measures  $\nu_n$  on  $\Gamma$  defined in (3.1) can be considered as measures on  $Y_n$ , we denote them by  $\hat{\nu}_n$ . We note that this definition varies slightly from the one given in [18, Section 4]. Specifically our  $\hat{\nu}_n$  measures are normalized to be probability measures unlike in [18]. Passing to the limit  $\nu$  on  $\Gamma \cup \partial\Gamma$ , one gets a limiting measure  $\hat{\nu} = \lim_{n \rightarrow \infty} \hat{\nu}_n$  supported on  $Y$  and giving zero measure to each  $Y_n$ . The fact that the limit exists follows from a direct calculation: the  $\hat{\nu}_n$  measure of each cylinder set (which are open and closed sets the collection of which generates the algebra on  $Y_\infty$ ) converges to a finite limit. The fact that  $\hat{\nu}$  assigns zero measure to each  $Y_n$  corresponds to the fact that  $\nu$  on  $\Gamma \cup \partial\Gamma$  is supported on the compact  $\partial\Gamma$  (i.e., gives zero mass to  $\Gamma$ ). Alternatively, denoting  $Y_\infty := Y \setminus (\cup_{n \geq 0} Y_n)$  it is easy to see that there is a Borel map  $\Psi : Y_\infty \rightarrow \partial\Gamma$  that takes an infinite path not ending with 0's to the equivalence class of the corresponding (infinite) geodesic ray in  $\partial\Gamma$  and which pushes  $\hat{\nu}$  forward to  $\nu$  (see [17, section 3.5] for a similar description and more details). Simple topological observations show that  $\Psi$  is continuous, surjective and finite-to-one ([17, Lemma 3.5.1]). It follows that  $\nu = \lim_{n \rightarrow \infty} \Psi_* \hat{\nu}_n$  and so  $\nu$  is obtained as the weak limit of the sequence  $\nu_n$  defined in (3.1).

#### 3.1.2 | Parry measure of the strongly Markov structure

It is well-known since the work of Shannon [63] and Parry [57] that given an irreducible subshift of finite type  $\Sigma_B$  (i.e., a subshift associated to an irreducible matrix  $B$  consisting of zeros and ones,

see Subsection 2.2.1) there is a unique  $\sigma$ -invariant probability measure  $\mu$  on  $\Sigma_B$  for which the corresponding measure theoretic entropy  $h_\mu(\sigma)$  is maximal among all  $\sigma$ -invariant (Borel) probability measures, that is,  $h_\mu(\sigma) = \sup_m h_m(\sigma)$  where the supremum is over all  $\sigma$ -invariant probability measures on  $\Sigma_B$ . Moreover this measure is a Markov measure in the sense that denoting by  $E$  the alphabet of  $\Sigma_B$ , the matrix  $B$  gives rise to a transition kernel  $P$  and a probability  $\mu_\bullet$  on  $E$  such that  $\mu_\bullet$  is  $P$ -stationary (see Subsection 4.1.1) and  $\mu = \mathbb{P}_{\mu_\bullet}$ . It is an ergodic probability measure (with respect to the shift transformation  $\sigma$ ). We call this measure  $\mu = \mathbb{P}_{\mu_\bullet}$  the measure of maximal entropy (also called the Parry measure of  $\Sigma_B$ ).

Even though the shift space  $\Sigma_A$  associated to the strongly Markov structure  $\mathcal{G}$  of a Gromov-hyperbolic group  $\Gamma$  (endowed with a generating set  $S$ ) is not irreducible, one can run a Parry-like construction [18, section 4.2] to obtain a shift-invariant Markovian probability measure  $\mu$  on  $\Sigma_A$  with the properties discussed below. We do not include the standard construction to avoid repetition.

A key property of the Parry-like measure  $\mu$  is that by [18, Lemma 4.19] (more precisely by [22, Proposition 4.6]), it is closely related to the measure  $\hat{\nu}$  on  $Y \subseteq \Sigma_A$  constructed above using the Patterson–Sullivan measure  $\nu$ : we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sigma_*^k \hat{\nu} = \mu, \tag{3.2}$$

where the convergence holds (in fact with a speed estimate) in total variation distance.

On the other hand, by [22, Proposition 4.2] the Parry-like measure  $\mu$  is nothing but a linear combination of the Parry measures of maximal irreducible components of  $\Sigma_A$ : for each maximal component  $(B_j)_{j=1, \dots, m}$ , there exists  $\alpha_j > 0$  such that  $\sum_{j=1}^m \alpha_j = 1$  and

$$\mu = \sum_{j=1}^m \alpha_j \mu_j, \tag{3.3}$$

where  $\mu_j$  is the Parry measure of the maximal component  $B_j$  of  $\mathcal{G}$ .

As pointed out in [18, section 4.3], one can be more precise about the relation between  $\hat{\nu}$  and  $\mu$  (than the mere relation (3.2)). We record the following statement from [22] which is an instance of this more precise relation and which will be useful later on.

**Lemma 3.1** ([22, Lemma 4.5]). *For each  $v \in V$  with  $\mu[v] > 0$  and  $k \in \mathbb{Z}_{\geq 0}$  there exists  $\alpha_v^k \geq 0$  such that*

$$\sigma_*^k \hat{\nu}|_{[v]} = \alpha_v^k \mu|_{[v]}.$$

*There exists a length  $k$  path from  $*$  to  $v$  if and only if  $\alpha_v^k > 0$ . □*

### 3.2 | Proofs of Theorems 1.1 and 1.10

We first prove Theorem 1.1 using Theorem 1.10. The latter will be proven subsequently.

*Proof of Theorem 1.1.* We begin by establishing the existence of the limit in the theorem. Nonnegativity of the limit will then be deduced as a consequence. Let  $\Lambda \in \mathbb{R}$  be the constant given by

Theorem 1.10. For each  $\epsilon > 0$ , define

$$A_\epsilon = \left\{ g \in \Gamma : \left| \frac{\varphi(g)}{|g|_S} - \Lambda \right| > \epsilon \right\}.$$

We need to show that for each  $\epsilon > 0$  the density of  $A_\epsilon$  on  $S_n$  vanishes as  $n \rightarrow \infty$ . Fix a Patterson–Sullivan measure, that is, the one given by the limit of (2.2). Note that for any fixed sufficiently large  $R > 0$ , there exist positive constants  $C_1$  and  $C_2$  such that for every  $n \in \mathbb{N}$ , we have

$$\frac{\#(S_n \cap A_\epsilon)}{\#S_n} \leq C_1 \sum_{x \in S_n \cap A_\epsilon} \nu(O(x, R)) \leq C_2 \nu\left(\bigcup_{x \in S_n \cap A_\epsilon} O(x, R)\right). \tag{3.4}$$

The first inequality follows from (2.3) and (2.4) while the second follows from the fact that, due to hyperbolicity,  $O(x, R)$  for  $x \in S_n$  covers  $\partial\Gamma$  up to uniformly bounded multiplicity.

Now note that if  $\xi \in \bigcup_{x \in S_n \cap A_\epsilon} O(x, R)$ , then there is a representative geodesic ray  $(\xi_m)_{m=0}^\infty$  with  $\xi_0 = \text{id}$  and  $C > 0$  such that for every  $n \geq 1$

$$\left| \frac{\log \|\rho(\xi_n)\|}{n} - \Lambda \right| \geq \epsilon - \frac{C}{n}. \tag{3.5}$$

Using hyperbolicity, (the fact that geodesic rays with the same end point remain within bounded distance) by enlarging  $C$  we can assume that (3.5) holds for all geodesic representatives of  $\xi \in \bigcup_{x \in S_n \cap A_\epsilon} O(x, R)$ . Therefore, we deduce by Theorem 1.10 that

$$\nu\left(\bigcup_{x \in S_n \cap A_\epsilon} O(x, R)\right) \xrightarrow{n \rightarrow \infty} 0.$$

Plugging this into (3.4), the first conclusion of Theorem 1.1 follows.

To deduce the second conclusion, note that for every positive  $n \in \mathbb{N}$  and  $g \in S_n$ , by subadditivity of  $\varphi$ , we have  $|\varphi(g)| \leq Dn$ , where  $D = \max_{g \in S} |\varphi(g)|$ . We write

$$\begin{aligned} \left| \frac{1}{\#S_n} \sum_{g \in S_n} \frac{\varphi(g)}{n} - \Lambda \right| &\leq \left| \frac{1}{\#S_n} \sum_{g \in S_n \cap A_\epsilon} \left( \frac{\varphi(g)}{n} - \Lambda \right) \right| + \left| \frac{1}{\#S_n} \sum_{g \in S_n \setminus A_\epsilon} \left( \frac{\varphi(g)}{n} - \Lambda \right) \right| \\ &\leq \frac{\#(S \cap A_\epsilon)}{\#S_n} (D + \Lambda) + \frac{\#(S_n \setminus A_\epsilon)}{\#S_n} \epsilon. \end{aligned}$$

As  $\frac{\#(S_n \setminus A_\epsilon)}{\#S_n} \rightarrow 1$  and  $\frac{\#(S_n \cap A_\epsilon)}{\#S_n} \rightarrow 0$  as  $n \rightarrow \infty$  and  $\epsilon > 0$  is arbitrary, the second statement of Theorem 1.1 follows.

It remains to show that  $\Lambda$  is nonnegative. To see this note that as  $|g|_S = |g^{-1}|_S$  for every  $g \in \Gamma$ , we have

$$2 \frac{1}{\#S_n} \sum_{g \in S_n} \frac{\varphi(g)}{n} = \frac{1}{\#S_n} \sum_{g \in S_n} \frac{\varphi(g) + \varphi(g^{-1})}{n} \geq \frac{1}{\#S_n} \sum_{g \in S_n} \frac{\varphi(\text{id})}{n} = \frac{\varphi(\text{id})}{n},$$

where  $\text{id} \in \Gamma$  is the identity element. The result follows by taking the limit as  $n \rightarrow \infty$ . □

*Remark 3.2.* Notice that we proved that the constant  $\Lambda$  appearing in Theorems 1.1 and 1.10 is the same. One deduces that the constant  $\Lambda$  given by Theorem 1.10 is nonnegative.

We now prove Theorem 1.10. To do so, we follow the argument used Cantrell in [22] employing additionally the subadditive ergodic theorem and properties of subadditive functions. The general tactic is to exploit the ergodicity of the Patterson–Sullivan measure to connect the behavior of different maximal components for the strongly Markov structure, a key idea due to Calegari–Fujiwara [18]. Before starting the proof, note that as different constructions of Patterson–Sullivan measures yield measures in the same measure class, Theorem 1.10 does not depend on which construction we choose to work with.

*Proof of Theorem 1.10.* Let  $\mathcal{G}$  be a strongly Markov structure associated to the tuple  $(\Gamma, S)$  and  $\Sigma_A$  the shift space defined over symbols corresponding to the vertices of  $\mathcal{G}$ . Let  $B$  be a maximal component (among  $B_1, \dots, B_m$ , see Subsection 2.2.1) of  $\mathcal{G}$  and let  $(\Sigma_B, \sigma)$  denote the subshift defined over this component. Let  $\mu_B$  denote the measure of maximal entropy on this subshift.

For every  $n \in \mathbb{N}$ , let  $f_n : \Sigma_B \rightarrow \mathbb{R}$  be the function given by

$$f_n(x) = \varphi(\lambda(x_0, x_1)) \dots \lambda(x_{n-1}, x_n).$$

As  $\varphi : \Gamma \rightarrow \mathbb{R}$  is subadditive, the sequence  $(f_n)_{n \in \mathbb{N}}$  constitutes a subadditive cocycle in the sense that  $f_{n+m}(x) \leq f_n(x) + f_m(\sigma^n x)$ . As the Parry measure  $\mu_B$  is ergodic, Kingman’s subadditive ergodic theorem therefore yields that there exists  $\Lambda \in \mathbb{R}$  such that

$$\lim_{n \rightarrow \infty} \frac{f_n(x)}{n} = \Lambda \text{ for } \mu_B\text{-almost every } x \in \Sigma_B. \tag{3.6}$$

Now consider the set

$$E = \left\{ \xi \in \partial\Gamma : \lim_{n \rightarrow \infty} \frac{\varphi(\xi_n)}{n} = \Lambda \right\}.$$

By Lemma 2.1, this set is well-defined and  $\Gamma$ -invariant. The conclusion of our theorem is equivalent to the fact that  $\nu(E) = 1$ , where  $\nu$  denotes the Patterson–Sullivan probability measure (2.2). As the measure (class of)  $\nu$  is ergodic with respect to the  $\Gamma$ -action [26], it suffices to show that  $\nu(E) > 0$ .

Recall from Subsection 3.1.1 that we have a surjective, continuous map  $\Psi : Y_\infty \rightarrow \partial\Gamma$  which pushes  $\hat{\nu}$  forward to  $\nu$ . Now fix an integer  $k$  and vertex  $v \in B$  such that there exists a length  $k$  path in  $\mathcal{G}$  from  $*$  to a  $v$ . By combining (3.3) and Lemma 3.1, one gets that there exists a constant  $\alpha > 0$  such that  $\sigma_*^k \hat{\nu}|_{[v]} = \alpha \mu_B|_{[v]}$ . In particular, by (3.6) we have that

$$\sigma_*^k \hat{\nu}|_{[v]} \left\{ x \in \Sigma_B : \lim_{n \rightarrow \infty} \frac{f_n(x)}{n} = \Lambda \right\} > 0.$$

Using the basic relation  $|\phi(gh) - \phi(h)| \leq |\phi(g)| + |\phi(g^{-1})|$  valid for any subadditive function  $\phi : \Gamma \rightarrow \mathbb{R}$  and every  $g, h \in \Gamma$ , one readily sees that the convergence  $f_n(x)/n \rightarrow \Lambda$  only depends on the tail of the sequence  $x$ : if  $x$  and  $y$  are two sequences such that  $x_{n_1+k} = y_{n_2+k}$  for some  $n_1, n_2 \in \mathbb{N}$  and all  $k \in \mathbb{N}$ , then the convergence holds either for both of  $x, y$  or neither of them.

This implies that the set

$$E_Y := \left\{ x \in Y : \lim_{n \rightarrow \infty} \frac{f_n(x)}{n} = \Lambda \right\} \text{ satisfies } \widehat{\nu}(E_Y) > 0.$$

As  $\widehat{\nu}$  pushes forward to  $\nu$  under  $\Psi$  and  $\Psi^{-1}(E) = E_Y$ , we have that  $\nu(E) = \Psi_* \widehat{\nu}(E) = \widehat{\nu}(E_Y) > 0$  and the conclusion follows.  $\square$

The following is a consequence of the above proof regarding positivity of the constant  $\Lambda$  based on the Markov property (see, e.g., [33, 60] for similar uses of this idea in close contexts).

**Proposition 3.3.** *Let  $H$  be a group endowed with a seminorm  $|\cdot|$ ,  $\rho : \Gamma \rightarrow H$  a morphism and  $\varphi(\cdot) = |\rho(\cdot)|$  the associated subadditive function on  $\Gamma$ . Suppose that every probability measure with finite exponential moment and with support that generates a finite index subgroup of  $\Gamma$ , has strictly positive  $|\cdot|$ -drift. Then the constant  $\Lambda$  given by Theorem 1.1 is strictly positive.*

The hypotheses of this proposition are a little awkward but are satisfied in many cases:

- (Furstenberg [31])  $H = \mathrm{SL}_d(\mathbb{R})$  and the  $\rho(\Gamma)$  is a strongly irreducible and nonrelatively compact,
- (Guivarc'h [42]) the image of  $\Gamma$  is nonamenable and has at most exponential  $|\cdot|$ -growth (see more precisely Kaimanovich–Kapovich–Schupp [46, Proposition 2.5]),
- (Maher–Tiozzo [54]) the group  $\Gamma$  acts nonelementarily on a geodesic Gromov-hyperbolic space.

As we have already used similar arguments involving the renewal measures in Section 4 we will provide a brief proof of the above result.

*Proof.* It follows from (3.6) that  $\Lambda$  is realized as the linear growth rate of  $\varphi$  along almost-sure (with respect to  $\mu_B$ ) Markovian trajectories  $(\gamma_1, \gamma_2, \dots)$  on  $\Gamma^{\mathbb{N}}$ . Now fix an edge  $(v_0, v_1)$  in the maximal component  $B$  and consider the induced (renewal) measure  $\nu$  on  $\Gamma$  obtained by return times the vertices  $v_0$  and  $v_1$  consecutively (see, e.g., [60, Definition 3.4]). Then, by the Markov property, the induced law on  $\Gamma^{\mathbb{N}}$  of this Markovian random walk along the return times to  $(v_0, v_1)$  is the Bernoulli law  $\nu^{\mathbb{N}}$  on  $\Gamma^{\mathbb{N}}$  (see, e.g., [60, Lemma 3.5]). In particular, as the state space is finite,  $\nu^{\mathbb{N}}$  is absolutely continuous with respect to  $\mu_B$ . Denoting by  $\tau_0$  the expectation of return times (which has a finite exponential moment because the state space is finite and the chain is irreducible), the  $|\cdot|$ -drift  $\Lambda_\nu$  of  $\nu$  satisfies  $\Lambda_\nu = \tau_0 \Lambda$ . Finally, by the same argument that we will see in Lemma 5.1 (relying on [39, Theorem 4.3]), the support of  $\nu$  generates a finite index subgroup of  $\Gamma$  and hence  $\Lambda_\nu > 0$  by hypothesis. The result follows.  $\square$

We end this section by justifying Remark 1.2(2): it clear that if the conclusions of Theorems 1.1 and 1.10 hold for a function  $\varphi$  and  $\varphi' : \Gamma \rightarrow \mathbb{R}$  is such that  $|\varphi - \varphi'|$  is bounded, then they also hold for  $\varphi'$ . In particular, if  $\varphi'$  is almost-subadditive in the sense that  $\varphi'(gh) \leq \varphi'(g) + \varphi'(h) + C$  for some  $C > 0$  and for all  $g, h \in \Gamma$ , then the conclusions of Theorems 1.1 and 1.10 hold for  $\varphi'$ : indeed, in this case the function  $\varphi' + C$  is subadditive. As a quasi-morphism (i.e., a function  $f : \Gamma \rightarrow \mathbb{R}$  satisfying  $|f(gh) - f(g) - f(h)| \leq C$  for some  $C > 0$  and for every  $g, h \in \Gamma$ ) on a group  $\Gamma$  is almost-subadditive, Theorems 1.1 and 1.10 apply. Notice that for a quasi-morphism, the limit  $\Lambda$  is necessarily zero.

## 4 | MARKOVIAN RANDOM MATRIX PRODUCTS

This section is mostly independent of the rest of the paper and it is devoted to limit theorems for norms of Markovian random matrix products (which will be important ingredients of our counting results): simplicity of Lyapunov exponents, invariance principle, functional LIL, large deviation estimates and Berry–Esseen bounds.

- *Simplicity of Lyapunov exponents* (Subsection 4.2): We will briefly recall the work of Bougerol [12] (see also Virtser [66] and Royer [61]) generalizing previous work of Guivarc’h [43] and ultimately the key result of Furstenberg [31] on positivity of the top Lyapunov exponent.
- *Invariance principle and LIL* (Subsection 4.3): We will recall the work of Bougerol [10, 11] generalizing corresponding results in the independent and identically distributed setting due to Le Page [50].
- *Large deviation estimates* (Subsections 4.4 and 4.5): We will prove large deviation estimates both for Markovian matrix products and Markovian random walks on Gromov-hyperbolic spaces. The former result is contained in Bougerol’s work [11], however, we will give a different proof using an approach of Benoist–Quint [6]. We have two reasons for giving a different proof: the first one is that the tools developed for this proof will be used in the large deviation ingredient of the Berry–Esseen estimate (which we could not directly obtain from Bougerol’s work), the second one is that this approach is more general and gives also the corresponding results for Markovian random walks on Gromov-hyperbolic spaces. The latter will be used later to give another setting providing a positive answer to a question of Kaimanovich–Kapovich–Schupp [46] (see Section 9).
- *Berry–Esseen estimates* (Subsection 4.6): We will prove Berry–Esseen estimates for matrix norms  $\log \|M_n\|$  using the corresponding estimates of Bougerol [11] for  $\log \|M_n \nu\|$  by adapting the approach of Xiao–Grama–Liu [67] and using our large deviation estimates.

Before proceeding, we mention that we will restrict ourselves to Markovian random matrix products over countable state Markov chains. The general state space cases are typically treated using similar ideas but with heavier machinery (see Guivarc’h [43] and Bougerol [11, 12] for nice expositions). Although we will only need to apply these results to the finite state space case, we include the countable setting because it does not introduce any serious additional difficulties and as we believe that this generality could be useful for works in contexts close to ours (e.g., for extensions of our counting results).

### 4.1 | Basic definitions

We start by setting our notation, brief recalls and defining Markovian random walks on groups and Markovian random matrix products.

#### 4.1.1 | Reminders on the theory of Markov chains on countable state spaces

Let  $E$  be a countable set and  $P$  a probability transition kernel on  $E$ . By (standard) abuse of notation, let  $P$  also denote the associated Markov operator and its dual: given a real-valued function  $f$  on  $E$ ,  $Pf(x) = \int f(y)P(x, dy)$  whenever the integral makes sense. We shall write  $\mu P$  for the

action of  $P$  on probability measures on  $E$ . Given a probability measure  $\nu$  on  $E$ , the distribution of the associated Markov chain on  $E^{\mathbb{N}}$  is denoted by  $\mathbb{P}_\nu$ . We will usually denote the sequence of coordinate functions by  $z_n$  for  $n = 0, 1, \dots$ . We say that the probability kernel  $P$  is irreducible if for every  $x, y \in E$ , there exists  $n \in \mathbb{N}$  such that  $P^n(x, y) > 0$ . For an irreducible kernel  $P$ , its period is defined to be  $\gcd\{n \in \mathbb{N} : P^n(x, x) > 0\}$  for some (equivalently all)  $x \in E$ . An irreducible kernel is said to be aperiodic (or primitive) if its period is one. In general, if the period is  $p \in \mathbb{N}$ , there exists a partition  $E_1, \dots, E_p$  of the state space  $E$  such that for every  $x \in E_i$ ,  $P(x, E_{i+1}) = 1$  ( $i \pmod p$ ). If  $P$  is irreducible and has period  $p$ , then  $P^p$  defines an irreducible aperiodic kernel on  $E_i$  for every  $i = 1, \dots, p$ .

A probability measure  $\pi$  on  $E$  is called  $P$ -stationary if it satisfies  $\pi P = \pi$ . An irreducible transition kernel  $P$  is said to be *positively recurrent* if it admits a stationary probability measure  $\pi$ , in which case this probability measure is unique. If  $P$  has period  $p \in \mathbb{N}$ , we have  $\pi = \frac{1}{p} \sum_{i=1}^p \pi|_{E_i}$  and  $\pi|_{E_i} P = \pi|_{E_{i+1}}$  ( $i \pmod p$ ), where  $p\pi|_{E_i}$  is the unique stationary probability measure of the irreducible aperiodic kernel  $P^p$  on  $E_i$ . The Markov chain  $(z_n)$  is said to be *uniformly geometrically ergodic* if there exist a  $P$ -stationary probability measure  $\pi$  on  $E$  and constants  $C > 0$  and  $\rho \in (0, 1)$  such that for every  $z \in E$  and  $n \in \mathbb{N}$ , we have  $\|P^n(z, \cdot) - \pi(\cdot)\|_{TV} \leq C\rho^n$ , where  $\|\cdot\|_{TV}$  denotes the total variation or equivalently the  $\ell^1$ -norm (note that this condition forces  $P$  to be irreducible and aperiodic). This is automatically satisfied if  $P$  satisfies the Doeblin condition (i.e., there exist  $n \in \mathbb{N}$ ,  $z \in E$  and  $\delta > 0$  such that for every  $y \in E$ ,  $P^n(y, z) \geq \delta$ ) and in particular if  $E$  is finite and  $P$  is primitive.

### 4.1.2 | Markovian matrix products associated to a Markov chain

Let  $E$  be a countable state space,  $P$  a transition kernel on  $E$  and  $\Gamma$  a group. Given a map  $X : E \rightarrow \Gamma$ , the associated *Markovian random walk* on  $\Gamma$  is defined as the process  $M_n = X(z_n) \cdots X(z_1)$ . We will often write  $X_n = X(z_n)$  for the  $n$ th-step of the associated Markovian random walk. When  $\Gamma \leq \text{GL}_d(\mathbb{R})$ , we will mostly refer to it as a Markovian random matrix product. In this section, we will always require that the transition kernel  $P$  be irreducible and positive recurrent. For Markovian random matrix products, we will always ask that the map  $X$  has the following integrability condition with respect to the stationary probability measure  $\pi$  of  $P$ :  $\mathbb{E}_\pi[\log N(X_1)] < \infty$ , where for a matrix  $g \in \text{GL}_d(\mathbb{R})$  and a choice of norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , we write  $N(g) = \max\{\log \|g\|, \log \|g^{-1}\|\}$ . Note that the integrability condition does not depend on the choice of norm.

We will briefly refer to all these data as a Markovian random walk or Markovian (random matrix) product and denote it by  $(M_n)$ .

### 4.2 | Simplicity of Lyapunov exponents

Given a Markovian product  $(M_n)$ , it follows from the Furstenberg–Kesten theorem (or subadditive ergodic theorem) that for every  $k = 1, \dots, d$  there exist constants  $\lambda_1 \geq \dots \geq \lambda_d$  such that  $\mathbb{P}_\pi$ -almost surely we have

$$\frac{1}{n} \log \|\wedge^k M_n\| \xrightarrow{n \rightarrow \infty} \sum_{i=1}^k \lambda_i. \tag{4.1}$$

These numbers are called the *Lyapunov exponents* of the Markovian product  $(M_n)$ . Clearly, they do not depend on the choice of the norm on  $\text{Mat}(\wedge^k \mathbb{R}^d)$ .



We will now see a result characterizing certain situations where these exponents are distinct from each other. We first need some definitions.

We say that a subset  $T$  of  $GL_d(\mathbb{R})$  is  $r$ -proximal with  $r \in \{1, \dots, d - 1\}$  if there exists a sequence  $(g_n)$  of elements in  $T$  such that  $\frac{g_n}{\|g_n\|}$  converges in  $Mat_d(\mathbb{R})$  to a linear transformation of rank at most  $r$ . Sometimes, we shall simply write proximal instead of 1-proximal.

Given a Markovian product  $(M_n)$ , for  $x_0 \in E$ , let  $T_{x_0} := \{M \in GL_d(\mathbb{R}) : \exists n \in \mathbb{N}, x_1, \dots, x_n \in E \text{ such that } P(x_i, x_{i+1}) > 0 \text{ and } M = X(x_n) \dots X(x_1)\}$ , where the indices  $i$  in the condition  $P(x_i, x_{i+1}) > 0$  ranges from 0 to  $n - 1$ . Moreover, for  $x \in E$ , let  $T_x := \{M \in GL_d(\mathbb{R}) : \exists n \in \mathbb{N}, x_1, \dots, x_n = x \in E \text{ such that } P(x_i, x_{i+1}) > 0 \text{ and } M = X(x_n) \dots X(x_1)\}$ . Note that for every  $x \in E$ ,  $T_x$  is a semigroup in  $GL_d(\mathbb{R})$  contained in the set  $T_x$ .

The Markovian product  $(M_n)$  is said to be  $r$ -contracting if there exists  $x \in E$  such that  $T_x$  is  $r$ -proximal. We say that a Markovian product  $(M_n)$  is irreducible if for any  $r \in \{1, \dots, d - 1\}$  there does not exist a map  $V : E \rightarrow Gr_r(\mathbb{R}^d)$  (where  $Gr_r$  denotes the Grassmanian of  $r$ -dimensional subspaces) such that for every  $x_0 \in E$  and  $n \in \mathbb{N}$ ,  $\mathbb{P}_{x_0}$ -almost surely  $M_n V(z_0) = V(z_n)$ . Finally, we say that a Markovian product  $(M_n)$  is strongly irreducible if for any  $r \in \{1, \dots, d - 1\}$  there does not exist a finite number of maps  $V_i : E \rightarrow Gr_r(\mathbb{R}^d)$  (say,  $i = 1, \dots, t$ ) such that for  $x \in E$ , denoting  $W(x) = \cup_{i=1}^t V_i(x)$ , we have, for every  $x_0 \in E$  and  $n \in \mathbb{N}$ ,  $\mathbb{P}_{x_0}$ -almost surely  $M_n W(z_0) = W(z_n)$ .

The following particular case of a result of Bougerol [12, Theorem 1.6] gives a characterization of the so-called simplicity of Lyapunov spectrum in our setting.

**Theorem 4.1** (Simplicity of Lyapunov spectrum, Guivarc’h [43] and Bougerol [12]). *Let  $(M_n)$  be a strongly irreducible Markovian product. Then, for  $r = 1, \dots, d - 1$ , we have  $\lambda_1 > \lambda_{r+1}$  if and only if  $M_n$  is  $r$ -contracting.*

This result will be a crucial ingredient for the upcoming limit theorems (Theorems 4.2 and 4.8). It will also be used in the proof of positivity of  $\Lambda$  in Proposition 1.4 (however, this can alternatively be deduced from Proposition 3.3 relying directly on the earlier positivity result of Furstenberg).

### 4.3 | Invariance principle and functional LIL

Here we briefly discuss two limit theorems due to Bougerol [11]: the first one is an analogue of the classical invariance principle due to Donsker which is a generalization of the CLT. The second one is the analogue of Strassen’s functional LIL generalizing the Hartman–Wintner LIL. For these results (and others to follow), we will need further assumptions on the Markovian product  $(M_n)$  that we now discuss.

Following Bougerol [11, section 3] (see also [10]), we shall say that a Markovian random matrix product  $(M_n)$  satisfies

- Condition  $(A_1)$ : if the Markov chain  $(z_n)$  is uniformly geometrically ergodic; and
- Condition  $(A_2)$ : if there exist positive constants  $a, B$  such that  $\mathbb{E}_x[e^{aN(M_1)}] \leq B$  for every  $x \in E$ .

Notice that both are automatically satisfied if  $E$  is finite and  $(z_n)$  is primitive.

**Theorem 4.2** (Convergence to the Wiener process and LIL, Bougerol [11]). *Let  $(M_n)$  be a 1-contracting irreducible Markovian random matrix product satisfying condition  $(A_1)$  and  $(A_2)$ .*

For  $\sigma_0 > 0$  and  $t \in [0, 1]$  and  $n \in \mathbb{N}$ , let  $S_n(t)$  denote  $C([0, 1])$ -valued random variable defined by

$$S_n(t) = \frac{1}{(n\sigma_0^2)^{1/2}} (\log \|M_{\lfloor tn \rfloor}\| - nt\lambda_1 + (nt - \lfloor nt \rfloor)(\log \|M_{\lfloor tn \rfloor + 1}\| - \log \|M_{\lfloor tn \rfloor}\|)). \quad (4.2)$$

Then, there exists a constant  $\sigma_0 = \sigma > 0$  such that for every  $x \in E$

- (1) under  $\mathbb{P}_x$ , the sequence  $(S_n)_{n \in \mathbb{N}}$  of  $C([0, 1])$ -valued random variables converges in distribution to  $\mathcal{W}$ ; and
- (2) for  $\mathbb{P}_x$ -almost every  $\omega$ , the set of limit points of the sequence  $(\frac{(S_n(t))(\omega)}{2 \log \log n})_{n \in \mathbb{N}}$  of elements of  $C([0, 1])$  is equal to the following compact subset of  $C([0, 1])$ :

$$\left\{ f \in C([0, 1]) : f \text{ is absolutely continuous, } f(0) = 0, \int_0^1 f'(t)^2 dt \leq 1 \right\}.$$

We indicate how to deduce this version from Bougerol’s original statement which concerns  $\log \|M_n v\|$  for a nonzero vector  $v \in \mathbb{R}^d$ .

*Proof.* In view of Theorem 4.1, our assumptions on the Markovian product  $(M_n)$ , namely, 1-contracting and irreducible, imply that the condition [11, (A3)] is satisfied (see [11, Definition 2.7] and thereafter). Therefore, [11, Theorem 4.5] implies both statements when  $\log \|M_n\|$  is replaced by  $\log \|M_n v\|$  for some nonzero  $v \in \mathbb{R}^d$ . Note that positivity of the variance follows from [11, Proposition 4.9]. The statements for  $\log \|M_n\|$  then follow from [11, Proposition 2.8]: the second conclusion directly follows and the first one follows by appealing to a standard fact, see, for example, [48, Problem 4.16]. □

### 4.4 | Large deviation estimates for Markovian random matrix products

In this part, we prove the following theorem by using some ideas that we adapt from the work of Benoist–Quint [6]. The developed tools will also serve as an ingredient in the proof of Berry–Esseen estimates.

**Theorem 4.3** (Markovian random matrix products). *Let  $(M_n)$  be a strongly irreducible and 1-contracting random matrix product satisfying  $(A_1)$  and  $(A_2)$ . Let  $\|\cdot\|$  be a fixed norm on  $\mathbb{R}^d$ . Then, for every  $\epsilon > 0$ , there exist  $\alpha > 0$  and  $C > 0$  such that for every  $x \in E$  and  $n \in \mathbb{N}$ , and nonzero  $v \in V$ , we have*

$$\mathbb{P}_x(|\log \|M_n v\| - n\lambda_1| \geq n\epsilon) \leq Ce^{-\alpha n} \quad \text{and} \quad \mathbb{P}_x(|\log \|M_n\| - n\lambda_1| \geq n\epsilon) \leq Ce^{-\alpha n}.$$

This result is not new; it follows from Bougerol’s [11, Theorem 4.3]. However, we give a different proof. The tools developed for this proof, beyond their aforementioned utility in the Berry–Esseen estimate, will also allow us to prove Theorem 4.6 in the next part. The rest of Subsection 4.4 is devoted to its proof.

### 4.4.1 | Large deviations in Breiman’s LLN

Following Benoist–Quint [6], we adopt a slightly more general setting. Let  $C$  be a compact metrizable space and  $E$  a Polish space. We say that a Markov-Feller transition kernel  $Q$  on  $Y = E \times C$  covers a Markovian transition kernel  $P$  on  $E$ , if the following diagram commutes

$$\begin{array}{ccc}
 E \times C & \xrightarrow{Q} & \mathcal{P}(E \times C) \\
 \pi_1 \downarrow & & \downarrow \pi_{1*} \\
 E & \xrightarrow{P} & \mathcal{P}(E)
 \end{array} \tag{4.3}$$

Here,  $\mathcal{P}(E)$  (resp.,  $\mathcal{P}(E \times C)$ ) denotes the set of probability measures on  $E$  (resp., on  $E \times C$ ),  $\pi_1 : E \times C \rightarrow E$  is the projection map and  $\pi_{1*}$  is the induced pushforward map.

Given a bounded continuous function  $\varphi : Y \rightarrow \mathbb{R}$ , we set

$$\ell_\varphi^+ = \sup_\eta \int \varphi d\eta \quad \text{and} \quad \ell_\varphi^- = \inf_\eta \int \varphi d\eta$$

where the supremum and infimum are taken over  $Q$ -invariant probability measures on  $Y$ .

Let us say that a Markov–Feller kernel  $P$  on  $E$  is uniformly positive recurrent if for every  $\epsilon > 0$ , there exists a compact set  $K \subseteq E$  and  $N \in \mathbb{N}$  such that for every  $x \in E$  and  $n \geq N$ , we have  $\frac{1}{n} \sum_{j=1}^n (\delta_x P^j)(K) > 1 - \epsilon$ . Such a kernel has a unique stationary probability measure that we denote by  $\pi$ . The following result is a more general version of [6, Proposition 3.1] that one can derive from its proof with a small variation explained below.

**Proposition 4.4** (Benoist–Quint). *Let  $Q$  be a Markov–Feller transition kernel on  $Y = E \times C$  covering a transition kernel  $P$  on  $E$ . Suppose that  $P$  is uniformly positive recurrent. Then, for every bounded continuous function  $\varphi : Y \rightarrow \mathbb{R}$  and  $\epsilon > 0$ , there exists  $C_0 > 0$  and  $\alpha_0 > 0$  such that for every  $y \in Y$  and  $n \in \mathbb{N}$ , we have*

$$\mathbb{Q}_y \left\{ (y_0, \dots) \in Y^\mathbb{N} : \frac{1}{n} \sum_{k=1}^n \varphi(y_k) \in [\ell_\varphi^- - \epsilon, \ell_\varphi^+ + \epsilon] \right\} > 1 - C_0 e^{-\alpha_0 n}.$$

*Proof.* The proof of [6, Proposition 3.1] goes through: the uniform convergence [6, (3.2)] is the only point that needs care in our noncompact case and it follows from the uniform positive recurrence assumption we imposed on the transition kernel  $P$  on  $E$ . Indeed, suppose that the convergence  $\max(\ell_\varphi^+, \frac{1}{n} \sum_{k=1}^n Q^k \varphi) \rightarrow \ell_\varphi^+$  is not uniform. Then, one finds a sequence  $y_n$  of points and  $\epsilon_0 > 0$  such that for every  $n \in \mathbb{N}$ ,  $\frac{1}{n} \sum_{k=1}^n Q^k \varphi(y_n) \geq \ell_\varphi^+ + \epsilon_0$ . By the uniform positive recurrence property of  $P$ , any limit point  $\eta$  of  $\frac{1}{n} \sum_{k=1}^n \delta_{y_n} Q^k$  projects to the  $P$ -invariant measure  $\pi$  and as  $C$  is compact,  $\eta$  gives full mass to  $Y$ . Hence, it is a  $Q$ -invariant probability measure on  $Y$  satisfying  $\int \varphi d\eta \geq \ell_\varphi^+ + \epsilon_0$ , a contradiction. □

We now prove a large deviation result for cocycles associated to group actions (cf. [6, Proposition 3.2]). Let  $\Gamma$  be a locally compact second countable group acting continuously on  $C$ . Let  $X : E \rightarrow \Gamma$  be a continuous map and  $P$  be a Markov–Feller transition kernel on  $E$ . We consider the Markov–Feller transition kernel on  $Y$  defined as follows: for Borel subsets  $A \subset E$  and  $B \subset C$  and  $y = (x, c)$ ,

we set

$$Q(y, A \times B) := P(x, A) 1_B(X(x)c). \tag{4.4}$$

By construction the kernel  $Q$  covers the transition kernel  $P$  in the sense of (4.3). For  $y_0 = (x_0, c) \in Y$ , we will denote by  $\mathbb{Q}_{y_0}$  the probability measure on  $Y^{\mathbb{N}}$  determined by the kernel  $Q$  and the initial distribution  $\delta_{y_0}$ . Note that  $\mathbb{Q}_{y_0}$  is the pushforward of  $\mathbb{P}_{x_0}$  by

$$\begin{aligned} E^{\mathbb{N}} &\rightarrow Y^{\mathbb{N}} \\ (x_i) &\mapsto ((x_0, c), (x_1, X(x_0)c), (x_2, X(x_1)X(x_0)c), \dots, y_n, \dots), \end{aligned} \tag{4.5}$$

where  $y_n = (x_n, X(x_{n-1}) \dots X(x_0)c)$ .

A continuous cocycle  $\sigma : \Gamma \times C \rightarrow \mathbb{R}$  is said to have uniform exponential moment if there exists  $\alpha_2 > 0$  and  $C_2 > 0$  such that for every  $x \in E$ ,  $\mathbb{E}_x[\sup_{c \in C} e^{\alpha_2 \sigma(X(z_1), c)}] \leq C_2$ .

**Proposition 4.5.** *Under the assumptions of Proposition 4.4, given a continuous cocycle  $\sigma : \Gamma \times C \rightarrow \mathbb{R}$  with uniform exponential moment for every  $\epsilon > 0$ , there exist  $C > 0$  and  $\alpha > 0$  such that for every  $x \in E$ ,  $c \in C$ , and  $n \in \mathbb{N}$ , we have*

$$\mathbb{P}_x \left\{ (x_0, \dots) : \frac{1}{n} \sum_{k=1}^n \sigma(X(x_k), X(x_{k-1}) \dots X(x_0)c) \in [\ell^- - \epsilon, \ell^+ + \epsilon] \right\} \geq 1 - Ce^{-\alpha n},$$

where  $\ell^+ = \sup_{\eta} \int \sigma(X(x), c) d\eta(x, c)$ , with the supremum taken over Borel probability measures on  $Y$  that are  $Q$ -invariant (and the lower average  $\ell^-$  is defined similarly with  $\inf$  instead of  $\sup$ ).

*Proof.* We will write the sum  $\sum_{k=1}^n \sigma(X(x_k), X(x_{k-1}) \dots X(x_0)c)$  as a sum of two quantities for which we have exponential concentration bounds, one of them due to Proposition 4.4 and the other due to Azuma type concentration bounds for sums of martingale differences (see, e.g., [51, Theorem 1.1]).

To do this, let  $\xi : Y \rightarrow \mathbb{R}$  be defined for  $y = (x, c)$  as  $\xi(y) = \int \sigma(X(z), X(x)c) dP_x(z)$ . Note that  $\xi$  is continuous (as  $P$  is Markov-Feller and  $\sigma$  is continuous) and bounded (thanks to the uniform exponential moment hypothesis). Furthermore, having fixed  $c \in C$ , let  $\phi_n$  be the sequence of functions defined on  $E^{\mathbb{N}}$  by

$$\phi_n((x_i)) = \sigma(X(x_n), X(x_{n-1}) \dots X(x_0)c) - \int \sigma(X(z), X(x_{n-1}) \dots X(x_0)c) dP_{x_{n-1}}(z).$$

We then have

$$\sum_{k=1}^n \sigma(X(x_k), X(x_{k-1}) \dots X(x_0)c) = \sum_{k=1}^n \phi_k((x_i)) + \sum_{k=0}^{n-1} \xi(y_k), \tag{4.4}$$

where we recall that  $y_k = (x_k, X(x_{k-1}) \dots X(x_0)c)$  and  $y_0 = (x_0, c)$ .

One now readily checks that for every  $(x, c) \in Y$ , under  $\mathbb{P}_x$ ,  $\phi_k$  is a martingale difference sequence with respect to the canonical filtration  $\mathcal{F}_n$  on  $E^{\mathbb{N}}$ . Indeed, for  $\mathbb{P}_{x_0}$ -almost every  $(x_i)$ , we

have that  $\mathbb{E}_x(\phi_n | \mathcal{F}_{n-1})(x_i)$  is equal to

$$\mathbb{E}_{x_{n-1}}[\sigma(X(z_n), X(x_{n-1}) \dots X(x_0)c)] - \int \sigma(X(z), X(x_{n-1}) \dots X(x_0)c) dP_{x_{n-1}}(z) = 0 \tag{4.7}$$

Thanks to the uniform exponential moment assumption, we can apply [51, Theorem 1.1] and deduce that for every  $\epsilon > 0$ , there exists  $C_1 > 0$  and  $\alpha_1 > 0$  such that for every  $(x, c) \in Y$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x \left\{ (x_i) : \frac{1}{n} \sum_{k=1}^n \phi_k((x_i)) \geq \epsilon \right\} \leq C_1 e^{-\alpha_1 n}. \tag{4.8}$$

On the other hand, by Proposition 4.4 applied to the function  $\xi$  and thanks to the relation (4.5), we obtain that for every  $\epsilon > 0$ , there exists  $C_0 > 0$  and  $\alpha_0 > 0$  such that for every  $(x, c) \in Y$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x \left\{ (x_i) : \frac{1}{n} \sum_{k=0}^{n-1} \xi(x_k, X(x_{k-1}) \dots X(x_0)c) \in [\ell_\xi^- - \epsilon, \ell_\xi^+ + \epsilon] \right\} \geq 1 - C_0 e^{-n\alpha_0}, \tag{4.9}$$

where  $\ell_\xi^+ = \sup_\eta \int \int \sigma(X(z), X(x)c) dP_x(z) d\eta(x, c) = \sup_\eta \int \sigma(X(x), c) d\eta(x, c)$  because  $\xi = Q\sigma$  and  $\eta$  is  $Q$ -stationary. In view of (4.6), the result now follows by (4.8) and (4.9).  $\square$

We are now ready to prove Theorem 4.3.

*Proof of Theorem 4.3.* We start by proving the first inequality. We will basically show that we can apply Proposition 4.5 and that in that result, we have  $\ell^- = \ell^+ = \lambda_1$ .

Let  $C$  be the  $(d - 1)$ -dimensional real projective space  $\mathbf{P}(\mathbb{R}^d)$  endowed with the usual action of  $GL_d(\mathbb{R})$ . We take  $\Gamma = GL_d(\mathbb{R})$  and  $X$  as the map  $E \rightarrow GL_d(\mathbb{R})$  in the data of the Markovian product  $(M_n)$ . Let  $Q$  be the Markov-Feller transition kernel on  $Y := E \times C$  constructed as in (4.4) covering the kernel  $P$  on  $E$ . Moreover, let  $\sigma : \Gamma \times \mathbf{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$  be the continuous cocycle given by  $\sigma(g, [v]) = \log \frac{\|gv\|}{\|v\|}$  where  $v$  is any nonzero element in  $\mathbb{R}^d$  and  $[v]$  denotes its projection to  $\mathbf{P}(\mathbb{R}^d)$ . Thanks to condition  $(A_1)$ , the kernel  $P$  is uniformly positively recurrent and thanks to condition  $(A_2)$ ,  $\sigma$  has a uniform exponential moment. Therefore, Proposition 4.5 yields that for every  $\epsilon > 0$ , there exist  $C > 0$  and  $\alpha > 0$  such that for every  $x \in E$ , nonzero  $v \in \mathbb{R}^d$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x \left( \frac{1}{n} \log \frac{\|M_n v\|}{\|v\|} \notin [\ell^- - \epsilon, \ell^+ + \epsilon] \right) \leq C e^{-\alpha n}. \tag{4.10}$$

We will now see that there exists a unique  $Q$ -stationary probability measure on  $Y$  and deduce that  $\ell^- = \ell^+$ . To this end, let  $\eta$  be a  $Q$ -stationary probability measure on  $E \times C$ . As  $Q$  covers the kernel  $P$ , the projection of  $\eta$  to  $E$  is a  $P$ -stationary probability measure, which is therefore equal to  $\pi$  (because  $\pi$  is the unique  $P$ -stationary probability measure on  $E$ ). Hence, we can write  $\eta = \int \delta_x \otimes \nu_x d\pi(x)$ , where  $\nu_x$  is a probability measure on  $C$ . By [5, Lemma 3.4],  $\nu_x$  is the stationary probability measure for the Markov operator  $Q_x$  induced by  $Q$  on  $\{x\} \times C$ . But as  $x$  is a single state in the state space  $E$ ,  $Q_x$  is the Markov operator induced by a probability measure  $\mu_x$  on  $\Gamma$  (the renewal measure defined by  $\mu_x(g) = \mathbb{P}_x(X(z_{\tau_x-1}) \dots X(z_1)X(x) = g)$ , where  $\tau_x$  is the return time

to  $x$ , see [60, section 3.1]), that is,  $Q_x((x, c), A \times B) = \delta_x(A) \cdot (\mu_x * \delta_c(B))$  for  $A \subseteq E$  and  $B \subseteq C$ . By construction, the semigroup generated by the renewal measure  $\mu_x$  is precisely the semigroup  $X(x)^{-1}T_x(x)X(x)$  ([60, page 15]). It follows from [12, Theorem 5.3.(ii)] that for every  $x \in E$ , the semigroup  $T_x(x)$  in  $GL_d(\mathbb{R})$  is strongly irreducible and proximal. Therefore, by [13, section III, Theorem 3.1],  $\nu_x$  is the unique stationary probability measure of  $\mu_x$  and hence it does not depend on  $\eta$ . This shows that  $Q$  has a unique stationary probability measure  $\eta$  on  $Y$ . In particular,  $\ell^- = \ell^+ =: \ell$  in (4.10). Choosing a basis  $v_1, \dots, v_d$  with unit vectors, applying (4.10) with each  $v_i$ , one gets by Borel–Cantelli that for every  $x \in E$ ,  $\mathbb{P}_x$  almost surely

$$\sup_{i=1, \dots, d} \frac{1}{n} \log \|M_n v_i\| \xrightarrow[n \rightarrow \infty]{} \ell.$$

As the supremum over a basis with unit vectors defines a norm on  $\text{Mat}_d(\mathbb{R})$  comparable to an operator norm, one gets by the subadditive ergodic theorem (see (4.1)) that  $\ell = \lambda_1$ . □

### 4.5 | Large deviation estimates for Markovian random walks on Gromov-hyperbolic spaces

We introduce some basic definitions from metric geometry to state Theorem 4.6.

Let  $(H, d)$  be a separable Gromov-hyperbolic metric space and  $o \in H$  a basepoint. Given  $x \in H$ , let  $h_x \in \text{Lip}_0^1(H)$  denote the function defined by  $h_x(y) = d(x, y) - d(x, o)$ , where  $\text{Lip}_0^1(H)$  is the set of 1-Lipschitz functions on  $H$  vanishing at  $o$ , endowed with the pointwise topology. By taking the closure in  $\text{Lip}_0^1(H)$ , we get a compactification  $\overline{H}^h$  of  $H$ , called the *horofunction compactification*. The compact  $\overline{H}^h$  is metrizable because  $H$  is separable. The map  $x \mapsto h_x$  is injective on  $H$  and we usually identify  $H$  with its image in  $\overline{H}^h$ . The *Busemann cocycle*  $\sigma : \text{Isom}(H) \times \overline{H}^h \rightarrow \mathbb{R}$  is defined by  $\sigma(\gamma, h) = h(\gamma^{-1}o)$ . Note that for  $o \in H \subseteq \overline{H}^h$ ,  $\sigma(\gamma, o) = d(\gamma o, o) =: \kappa(\gamma)$  is the displacement functional. Recall finally that an element  $\gamma \in \text{Isom}(H)$  is called *loxodromic* if it has precisely two fixed points on the Gromov boundary  $\partial H$  of  $H$ .

We say that a Markovian random walk  $(M_n)$  on  $\text{Isom}(H)$  is *nonelementary* if there does not exist a finite number of maps  $V_i$  (say,  $i = 1, \dots, t$ ) from  $E$  to the Gromov boundary  $\partial H$  such that for  $x \in E$ , denoting  $W(x) = \cup_{i=1}^t V_i(x)$ , we have, for every  $x_0 \in E$  and  $n \in \mathbb{N}$ ,  $\mathbb{P}_{x_0}$ -almost surely  $M_n W(x_0) = W(z_n)$ . Moreover, we say that the Markovian random walk  $(M_n)$  satisfies *Condition  $(A'_2)$*  if there exist positive constants  $a$  and  $B$  such that  $\mathbb{E}_x[e^{a\kappa(M_1)}] \leq B$  for every  $x \in E$ .

**Theorem 4.6** (Markovian random walks on Gromov-hyperbolic spaces). *Let  $H$  be a separable geodesic Gromov-hyperbolic space and  $(M_n)$  be a nonelementary Markovian random walk on  $\text{Isom}(H)$  satisfying  $(A_1)$  and  $(A'_2)$ . Then, there exists a constant  $\ell_\Lambda \geq 0$  such that for every  $\epsilon > 0$ , there exists  $\alpha > 0$  and  $C > 0$  such that for every  $x \in E$ ,  $\xi \in \overline{H}^h$  and  $n \in \mathbb{N}$ , we have*

$$\mathbb{P}_x(|\sigma(M_n, \xi) - n\ell_\Lambda| \geq n\epsilon) \leq Ce^{-\alpha n}.$$

The constant  $\ell_\Lambda$  is called the *drift of the Markovian random walk  $(M_n)$* . We note that specializing to  $\xi = o \in H$ , the above statement boils down to large deviation estimates for the displacement function  $\kappa(\cdot)$ .

This result generalizes the assertion concerning the uniqueness of zero in the recent work [14] in the independent and identically distributed setting (let us note that even more recently, Gou zel [38] managed to get rid of the exponential moment assumption in the same setting). In the independent and identically distributed case, the dependence of  $\alpha$  on  $\epsilon$  has been specified and quantitative estimates have been recently obtained when  $H$  is proper (see [1, 3]). Finally, see also the recent work of Goldsborough–Sisto [37] for another perspective on Markovian random products of isometries.

*Proof.* We aim to apply Proposition 4.5. To this end, let  $C = \overline{H}^h$  and  $\Gamma$  be the countable group generated by the image of the map  $X : E \rightarrow \text{Isom}(H)$  in the data of the Markovian random walk  $(M_n)$ . Recall that the group of isometries  $\text{Isom}(H)$  acts on  $\overline{H}^h$  by homeomorphisms given, for  $\gamma \in \text{Isom}(H)$ ,  $h \in \overline{H}^h$  and  $y \in M$ , by  $(\gamma \cdot h)(y) = h(\gamma^{-1}y) - h(\gamma^{-1}o)$  and the Busemann cocycle is a continuous cocycle over this action. Let  $Q$  be the Markov–Feller transition kernel on  $Y := E \times C$  constructed as in (4.4) covering the kernel  $P$  on  $E$ . Thanks to condition  $(A_1)$ , the kernel  $P$  is uniformly positively recurrent and thanks to condition  $(A_2)$ ,  $\sigma$  has a uniform exponential moment. Therefore, Proposition 4.5 implies that for every  $\epsilon > 0$ , there exist  $C > 0$  and  $\alpha > 0$  such that for every  $x \in E$ ,  $\xi \in \overline{H}^h$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x\left(\frac{1}{n}\sigma(M_n, \xi) \notin [\ell^- - \epsilon, \ell^+ + \epsilon]\right) \leq Ce^{-\alpha n}.$$

We now let  $\ell_\Lambda$  be the constant given by the subadditive ergodic theorem as the  $\mathbb{P}_\pi$  almost sure limit of  $\frac{1}{n}\kappa(M_n)$  as  $n \rightarrow \infty$ . It remains to show that  $\ell^+ = \ell^- = \ell_\Lambda$ . If  $\ell_\Lambda = 0$  this equality is easy to see (using the fact that the cocycle  $\sigma$  is bounded in absolute value by  $\kappa$ , one controls the behavior of  $\sigma$  for any stationary measure covering  $P$  with the asymptotic behavior of  $\kappa$ ), so we suppose that  $\ell_\Lambda > 0$ . Here, we will employ a slightly different argument compared to the corresponding part in the proof of Theorem 4.3 replacing the proof of uniqueness of  $Q$ -stationary probability measure on  $Y$  (which may also be the case in the current setting). Let  $\eta$  be a  $Q$ -stationary probability measure on  $Y$ . By the Chacon–Ornstein ergodic theorem,

$$\eta\left\{(x, c) \in Y : \frac{1}{n} \sum_{k=0}^{n-1} Q^k \sigma(x, c) \xrightarrow{n \rightarrow \infty} \int \sigma d\eta\right\} = 1.$$

By specializing to such  $(x_0, c)$ , using the fact that for every  $y = (x', c')$ ,  $\mathbb{Q}_y$  is the pushforward of  $\mathbb{P}_{x'}$  by the map (4.5), we get that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{x_0}[\sigma(X(z_{n-1}) \dots X(z_1)X(x_0), c)] = \int \sigma d\eta. \tag{4.11}$$

For  $n \geq 1$ , let  $\tau_x(n)$  denote the random variable, defined on  $E^\mathbb{N}$ , which is given by  $n$ th-return time to  $x$ . Thanks to condition  $A_1$ ,  $\tau_x(1)$  (equivalently  $\tau_x(n)$  for every  $n \in \mathbb{N}$ ) has a finite exponential moment. For  $x \in E$  let  $\mu_x$  be the renewal measure  $\mu_x(g) = \mathbb{P}_x(X(z_{\tau_x(1)-1}) \dots X(z_1)X(x) = g)$ . It is easy to see that  $\mu_x$  has a finite exponential moment, that is,  $\int e^{\beta\kappa(g)} d\mu_x(g) < \infty$  for some  $\beta > 0$ . The support of  $\mu_x$  is the subsemigroup  $X(x)^{-1}T_x(x)X(x)$  of  $\Gamma$  defined in the same way as in Subsection 4.2. As the Markovian product  $(M_n)$  is nonelementary and has positive drift, the semigroup  $T_x(x)$  is clearly unbounded (as the random product, which grows at linear rate, belongs to this



semigroup at return times to  $x$ ). Moreover, it follows from the same argument as in the proof of [12, Lemma 4.6, Theorem 5.3.(ii)] that  $T_x(x)$  does not stabilize a finite collection of points in the Gromov boundary  $\partial H$ . Now, [24, Proposition 3.1] implies that the group generated by  $T_x(x)$  contains two independent loxodromics and then [30, Theorem 6.2.3 and Proposition 6.2.14] imply that the semigroup  $T_x(x)$  is nonelementary (i.e., contains two independent loxodromics).

For the rest, on the one hand, it is not hard to deduce from (4.11) that

$$\frac{1}{n} \mathbb{E}_{\mu_{x_0}} [\sigma(g_n \dots g_1, c)] \rightarrow \mathbb{E}[\tau_{x_0}(1)] \int \sigma \, d\eta \tag{4.12}$$

(where  $g_n \dots g_1$  is an independent and identically distributed random walk on  $\Gamma$  for  $\mu_{x_0}$ ) and as  $\mu_{x_0}$  is nonelementary and has a finite exponential moment, it follows from [1, Lemma 3.8] and (4.12) that

$$\frac{1}{n} \mathbb{E}_{\mu_{x_0}} [\kappa(g_n \dots g_1)] \rightarrow \mathbb{E}[\tau_{x_0}(1)] \int \sigma \, d\eta. \tag{4.13}$$

On the other hand, the left-hand side of (4.12) converges to  $\mathbb{E}[\tau_{x_0}(1)]\ell_\Lambda$  and  $\mathbb{E}[\tau_{x_0}(1)] > 0$  thanks to positive recurrence. This shows that  $\int \sigma \, d\eta = \ell_\Lambda$ . As  $\eta$  is an arbitrary  $Q$ -stationary probability measure, this shows that  $\ell^+ = \ell^- = \ell_\Lambda$ , completing the proof.  $\square$

*Remark 4.7.* We remark that in Theorem 4.6, we cannot exclude the possibility that  $\ell_\Lambda = 0$ . However, a handy characterization of when  $\ell_\Lambda > 0$  follows from the previous proof. Indeed, let  $x \in E$  and  $T_x(x)$  be the semigroup above. As in the proof above, as the Markovian product  $(M_n)$  is nonelementary, the semigroup  $T_x(x)$  does not stabilize a finite collection of points in the Gromov-boundary  $\partial H$ . Moreover, as shown above if  $T_x(x)$  is unbounded,  $T_x(x)$  contains two independent loxodromics. The linear escape result in [54] implies that the drift of  $\mu_x$  is positive and hence from the proof above, we get  $\ell_\Lambda > 0$ . On the other hand, it is easy to see that  $\ell_\Lambda = 0$  if for some (equivalently all)  $x \in E$ , the semigroup  $T_x(x)$  is bounded.

### 4.6 | Central limit theorem with Berry–Esseen type estimates

Specializing to  $t = 1$  in (4.2), Theorem 4.2(1) says that the CLT holds: for every  $x \in E$  and  $a \in \mathbb{R}$ , we have

$$\mathbb{P}_x \left( \log \|M_n\| - n\lambda_1 \leq a\sqrt{n} \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{s^2}{2\sigma^2}} \, ds. \tag{4.14}$$

In the following result, we give the Berry–Esseen type bound for the convergence rate in (4.14). Our main interest in the Berry–Esseen bound is that it will be used to obtain a *quantitative* counting CLT on spheres of the Cayley graph of a Gromov-hyperbolic group. Unlike Theorem 4.2, it is not simple to deduce the Berry–Esseen type bound for  $\log \|M_n\|$  from that of  $\log \|M_n v\|$ , the latter was proven by Bougerol [11]. Indeed, even in the independent and identically distributed case, although the Berry–Esseen bound for  $\log \|M_n v\|$  has been known since the work of Le Page [50], the bounds for the matrix norm were only recently studied [28, 29, 67]. Below, we give a version of these results for the Markovian case adapting the approach of Xiao–Grama–Liu [67] and

using our large deviation estimates (replacing the large deviation ingredient of [67] from [8] in the independent and identically distributed case).

**Theorem 4.8** (Berry–Esseen bound in CLT). *Let  $(M_n)$  be a strongly irreducible and 1-contracting Markovian product satisfying  $(A_1)$  and  $(A_2)$ . Let  $\|\cdot\|$  be a fixed norm on  $\mathbb{R}^d$ . Then, there exists a constant  $D > 0$  such that for every  $x \in E$ ,  $a \in \mathbb{R}$ , and  $n \in \mathbb{N}$ , we have*

$$\left| \mathbb{P}_x \left( \log \|M_n\| - n\lambda_1 \leq a\sqrt{n} \right) - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{s^2}{2\sigma^2}} ds \right| \leq \frac{D \log n}{\sqrt{n}}.$$

Regarding the CLT, we signal that in view of the more recent progress of Benoist–Quint [6, 7] (see also [29] for Berry–Esseen estimates) optimizing the moment hypothesis in the CLT for the independent and identically distributed case (respectively, improving the Berry–Esseen estimates), it is probable that (4.14) and some Berry–Esseen type estimates hold under a polynomial moment hypothesis (we do not pursue these directions).

To prove the Berry–Esseen estimate in Theorem 4.8, we will need some further results on large deviation estimates which are given in the next two lemmas.

**Lemma 4.9.** *Under the assumptions of Theorem 4.8, for every  $\epsilon > 0$ , we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_x \left( \frac{1}{n} \log \frac{\|M_n\|^2}{\|\wedge^2 M_n\|} \leq \lambda_1 - \lambda_2 - \epsilon \right) < 0,$$

uniformly in  $x \in E$ .

*Proof.* As the top Lyapunov exponent of the Markovian product  $(\wedge^2 M_n)$  is  $\lambda_1 + \lambda_2$  and  $(\wedge^2 M_n)$  satisfies  $(A_1)$  and  $(A_2)$ , by Theorem 4.3, it suffices to show that for every  $\epsilon > 0$ , we have the following uniformly in  $x \in E$ :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_x \left( \frac{1}{n} \log \|\wedge^2 M_n\| \geq \lambda_1 + \lambda_2 + \epsilon \right) < 0. \tag{4.15}$$

To prove this, we will apply Proposition 4.5. Let  $C = \mathbb{P}(\wedge^2 \mathbb{R}^d)$  and  $G = \text{GL}_d(\mathbb{R})$  and  $X$  as the map  $E \rightarrow \text{GL}_d(\mathbb{R})$  in the data of the Markovian product  $(\wedge^2 M_n)$ , let  $Q$  be the Markov–Feller transition kernel on  $Y := E \times C$  constructed as in (4.4) covering the kernel  $P$  on  $E$ . Moreover, let  $\sigma : G \times \mathbf{P}(\wedge^2 \mathbb{R}^d) \rightarrow \mathbb{R}$  be the continuous cocycle given by  $\sigma(g, [v \wedge w]) = \log \frac{\|\wedge^2 g(v \wedge w)\|}{\|v \wedge w\|}$  where  $v \wedge w$  is a line in  $\wedge^2 \mathbb{R}^d$  and  $[v \wedge w]$  denotes its projection to  $\mathbf{P}(\wedge^2 \mathbb{R}^d)$ . Thanks to condition  $(A_1)$ , the kernel  $P$  is uniformly positively recurrent and thanks to condition  $(A_2)$ ,  $\sigma$  has a uniform exponential moment. Therefore, we can apply Proposition 4.5 and deduce that for every  $\epsilon > 0$ , there exists  $C > 0$  and  $\alpha > 0$  such that for every  $x \in E$ , nonzero  $v \wedge w \in \wedge^2 \mathbb{R}^d$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x \left( \frac{1}{n} \log \frac{\|\wedge^2 M_n(v \wedge w)\|}{\|v \wedge w\|} \notin [\ell^- - \epsilon, \ell^+ + \epsilon] \right) \leq C e^{-\alpha n}.$$

By choosing a basis of  $\wedge^2 \mathbb{R}^d$  as in the proof of Theorem 4.3, we only need to show that  $\ell^+ \leq \lambda_1 + \lambda_2$ . Let  $\delta > 0$  be given and  $\eta$  be a  $Q$ -stationary and ergodic probability measure on  $Y$  with

$\int \sigma d\eta \geq \ell^+ - \delta$ . By the Chacon–Ornstein ergodic theorem, we have

$$\eta \left\{ (x, c) \in Y : \frac{1}{n} \sum_{k=0}^{n-1} Q^k \sigma(x, c) \xrightarrow[n \rightarrow \infty]{} \int \sigma d\eta \right\} = 1.$$

By specializing to such  $(x_0, c)$ , using the fact that for every  $y = (x', c')$ ,  $\mathbb{Q}_y$  is the pushforward of  $\mathbb{P}_x$  by the map (4.5), we get that

$$\begin{aligned} \ell^+ - \delta &\leq \int \sigma d\eta = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{x_0} [\sigma(X(x_{n-1}) \dots X(x_0), c)] \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{x_0} [\log \| \wedge^2 X(x_{n-1}) \dots X(x_0) \|]. \end{aligned}$$

It remains to observe that the last term is bounded above by  $\lambda_1 + \lambda_2$ . In fact we claim that it is equal to  $\lambda_1 + \lambda_2$ . Indeed, by subadditive ergodic theorem,  $\mathbb{P}_\pi$ -almost surely  $\frac{1}{n} \log \| \wedge^2 X(x_{n-1}) \dots X(x_0) \| \rightarrow \lambda_1 + \lambda_2$ . As  $\mathbb{P}_{x_0}$  is absolutely continuous with respect to  $\mathbb{P}_\pi$ , this convergence also holds true  $\mathbb{P}_{x_0}$ -almost surely. As the sequence  $\frac{1}{n} \log \| \wedge^2 X(x_{n-1}) \dots X(x_0) \|$  is uniformly integrable (thanks to condition  $(A_2)$ ), the result follows.  $\square$

The next lemma is the Markovian version of [8, Lemma 17.8] which was used to deduce a local limit theorem for the norms  $\log \|M_n\|$  from a local limit theorem for vector norms  $\log \|M_n v\|$  in the independent and identically distributed setting. Thanks to our above large deviation estimates, Benoist–Quint’s proof applies in our setting as we indicate below.

**Lemma 4.10.** *Under the assumptions of Theorem 4.8, for every  $\epsilon > 0$ , there exists  $\ell_0 \in \mathbb{N}$ , and  $c > 0$  such that for every  $n \geq \ell \geq \ell_0$  and nonzero  $v \in \mathbb{R}^d$  and  $x \in E$ , we have*

$$\mathbb{P}_x \left( \left| \log \|M_n\| - \log \frac{\|M_n v\|}{\|M_\ell v\|} - \log \|M_\ell\| \right| > e^{-c\ell} \right) < e^{-c\ell}.$$

*Proof.* We can apply the proof of the result [8, Lemma 17.8 and (17.9)] which makes use of various large deviation estimates all of which are established in our more general setting. Namely, the ingredient [8, Proposition 14.3] is similarly obtained in our setting using Theorem 4.3 and Lemma 4.9 which we proved for this purpose. Once equipped with this ingredient, Benoist–Quint only use the independent and identically distributed version of Lemma 4.9 (see [8, (17.10)]) and the linear algebraic lemma [8, Lemma 14.2] and hence this part of the proof goes through in our setting as well. We omit the details in order not to burden the exposition with many more linear algebraic notions that will only be used in a repetitive proof.  $\square$

Equipped with the previous two lemmas, we can give the proof of Theorem 4.8 adapting the approach of [67].

*Proof.* Let  $F$  denote the cumulative distribution function of the Gaussian  $\mathcal{N}(0, 1)$ , and for  $n \geq 1$ ,  $x \in E$  and  $t \in \mathbb{R}$ , set

$$I_{x,n}(t) = \mathbb{P}_x \left( \frac{\log \|M_n\| - n\lambda_1}{\sigma \sqrt{n}} \leq t \right).$$

As for any  $v \in \mathbb{R}^d$  of unit norm, we have  $\|M_n v\| \leq \|M_n\|$ , for any  $v \in \mathbb{R}^d$  with  $\|v\| = 1$ , by [11, Theorem 4.1], we have

$$I_{x,n}(t) \leq \mathbb{P}_x \left( \frac{\log \|M_n v\| - n\lambda_1}{\sigma\sqrt{n}} \leq t \right) \leq F(t) + \frac{C}{\sqrt{n}}.$$

The nontrivial bound is therefore the lower bound for  $I_{x,n}(t)$  which we now turn to. By Lemma 4.10, for any  $\epsilon > 0$ , there exist  $\ell_0 \in \mathbb{N}$  and  $c > 0$  such that for every  $n \geq \ell \geq \ell_0$ , we have

$$\begin{aligned} I_{x,n}(t) &\geq \mathbb{P}_x \left( \frac{\log \|M_n\| - n\lambda_1}{\sigma\sqrt{n}} \leq t \quad \text{and} \quad \left| \log \|M_n\| - \log \frac{\|M_n v\|}{\|M_\ell v\|} - \log \|M_\ell\| \right| \leq e^{-\epsilon\ell} \right) \\ &\geq \mathbb{P}_x \left( \frac{\log \|M_n v\| - \log \|M_\ell v\| + \log \|M_\ell\| - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} \leq t \right) - e^{-c\ell}. \end{aligned} \tag{4.16}$$

By Theorem 4.3, for every  $\lambda' > \lambda_1$ , there exists  $c' > 0$  such that for every  $\ell \geq 1$ , for every  $x \in E$ , we have

$$\mathbb{P}_x(\log \|M_\ell\| \geq \ell\lambda') \leq e^{-c'\ell}.$$

Therefore, it follows from (4.16) that for every  $x \in E$  and  $n \geq \ell \geq \ell_0$ , we have

$$I_{x,n}(t) \geq \mathbb{P}_x \left( \frac{\log \|M_n v\| - \log \|M_\ell v\| + \ell\lambda' - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} \leq t \right) - e^{-c'\ell} - e^{-c\ell}. \tag{4.17}$$

Notice now that we can rewrite  $\log \|M_n v\| - \log \|M_\ell v\|$  as  $\log \|X_n \dots X_{\ell+1} \tilde{v}\|$ , where  $\tilde{v} = \frac{M_\ell v}{\|M_\ell v\|}$ . To exploit this cocycle property, for  $\ell \in \mathbb{N}$ , let  $\mathcal{F}_\ell$  denote the  $\sigma$ -algebra generated by the first  $\ell$  steps  $z_1, \dots, z_\ell$  of the Markov chain on  $E$ . Conditioning on the first  $\ell$ -steps, we have

$$\begin{aligned} &\mathbb{P}_x \left( \frac{\log \|M_n v\| - \log \|M_\ell v\| + \ell\lambda' - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} \leq t \right) \\ &= \mathbb{E}_x \left( \mathbb{P}_x \left( \frac{\log \|M_n v\| - \log \|M_\ell v\| + \ell\lambda' - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} \leq t \mid \mathcal{F}_\ell \right) \right) \\ &\geq \mathbb{E}_x \left( \inf_{v \in \mathbb{R}^d, \|v\|=1} \mathbb{P}_{z_\ell} \left( \frac{\log \|M_{n-\ell} v\| + \ell\lambda' - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} \leq t \right) \right) \\ &= \mathbb{E}_x \left( \inf_{v \in \mathbb{R}^d, \|v\|=1} \mathbb{P}_{z_\ell} \left( \frac{\log \|M_{n-\ell} v\| - (n-\ell)\lambda_1}{\sigma\sqrt{n-\ell}} \leq T_n \right) \right), \end{aligned} \tag{4.18}$$

where  $T_n$  is the random variable

$$T_n = \frac{\sqrt{n} - \sqrt{n-\ell}}{\sigma\sqrt{n(n-\ell)}} \log \|M_{n-\ell} v\| - \frac{\ell\lambda' - n\lambda_1 + e^{-\epsilon\ell}}{\sigma\sqrt{n}} + t - \frac{\lambda_1\sqrt{n-\ell}}{\sigma}.$$

By Theorem 4.3, up to possibly reducing  $c' > 0$ , we have that for every  $n \geq \ell \geq 1$ , for every  $v \in \mathbb{R}^d$  with  $\|v\| = 1$  and  $y \in E$ ,

$$\mathbb{P}_y(\log \|M_{n-\ell} v\| > \lambda'(n-\ell)) \leq e^{-c'(n-\ell)}$$

so that by (4.18), we have

$$\begin{aligned} & \mathbb{P}_x \left( \frac{\log \|M_n v\| - \log \|M_\ell v\| + \ell \lambda' - n \lambda_1 + e^{-\epsilon \ell}}{\sigma \sqrt{n}} \leq t \right) \\ & \geq \mathbb{E}_x \left( \inf_{v \in \mathbb{R}^d, \|v\|=1} \mathbb{P}_{z_\ell} \left( \frac{\log \|M_{n-\ell} v\| - (n-\ell) \lambda_1}{\sigma \sqrt{n-\ell}} \leq t_n \right) \right) - e^{c'(n-\ell)}, \end{aligned} \tag{4.19}$$

where  $t_n$  is the constant

$$t_n = \frac{\sqrt{n} - \sqrt{n-\ell}}{\sigma \sqrt{n(n-\ell)}} \lambda'(n-\ell) - \frac{\ell \lambda' - n \lambda_1 + e^{-\epsilon \ell}}{\sigma \sqrt{n}} + t - \frac{\lambda_1 \sqrt{n-\ell}}{\sigma}.$$

Now applying once more [11, Theorem 4.1] to (4.19) and combining it with (4.17), we get that for every  $n \geq \ell \geq \ell_0$  and  $x \in E$

$$I_{x,n}(t) \geq F(t_n) - \frac{C}{\sqrt{n-\ell}} - e^{-c'(n-\ell)} - 2e^{-c''\ell}$$

with  $0 < c'' := \max\{c, c'\}$ . Now using the expression of  $t_n$  above, one gets that for any  $r > 0$  fixed, letting  $\ell = \lfloor r \log n \rfloor$ , we have  $|t - t_n| \leq \frac{D_r \log n}{\sqrt{n}}$  for some  $D_r \in (0, \infty)$  and every  $n \in \mathbb{N}$  and  $t \in \mathbb{R}$ . Using this and the fact that  $F$  is the cumulative distribution function of the standard Gaussian  $\mathcal{N}(0, 1)$ , one deduces by elementary calculus that choosing  $\ell = \lfloor \frac{1}{c''} \log n \rfloor$ , we have that there exists  $D \in (0, \infty)$  such that for every  $n \in \mathbb{N}$ ,  $x \in E$  and  $t \in \mathbb{R}$ , we have

$$I_{x,n}(t) \geq F(t) - \frac{D \log n}{\sqrt{n}},$$

as required. □

### 4.7 | Finite state versions without condition $(A_1)$

In our applications, we will need to deal with Markovian products associated to Markov chains on finite state spaces that are irreducible but not necessarily aperiodic. Such chains never satisfy the uniform recurrence condition  $(A_1)$ . However, it is not hard to deduce versions of above limit theorems for such finite state chains by considering the Markovian products along periodic times  $(pn)_{n \in \mathbb{N}}$ , where  $p \in \mathbb{N}$  denotes the period of the Markov chain. The goal of this part is to briefly record these versions of the above limit theorems for later use.

Let  $E$  be a finite state space and  $P$  an irreducible Markovian transition kernel on  $E$ . We denote by  $p \in \mathbb{N}$  the period of  $P$  and, for  $i = 0, \dots, p-1$ , by  $E_i$  the periodic components of  $E$ . Let a map

$X : E \rightarrow \Gamma$  to a group  $\Gamma$  be given and  $(M_n)$  be the associated Markovian product (recall that  $M_n = X(z_n) \dots X(z_1)$ , where  $(z_n)_{n \geq 0}$  denotes the Markov chain on  $E$ ). The Markovian product  $(M_n)$  does not necessarily satisfy condition  $(A_1)$ ; we will associate some auxiliary Markovian products that will satisfy it. To do this, for  $i = 0, \dots, p - 1$ , let  $\widehat{E}^i$  be the set of length  $p$ -paths based at  $E_i$ , that is,

$$\widehat{E}^i := \{(x_1, \dots, x_p) : P(x_j, x_{j+1}) > 0 \text{ for } j = 1, \dots, p - 1, \text{ and } x_1 \in E_{i+1}\},$$

where  $j$ 's are considered modulo  $p$ . We introduce a Markovian transition kernel  $\widehat{P}^i$  on  $\widehat{E}^i$  by setting, for  $(x_1, \dots, x_p)$  and  $(y_1, \dots, y_p)$  in  $\widehat{E}^i$ ,

$$\widehat{P}^i((x_1, \dots, x_p), (y_1, \dots, y_p)) = P(x_p, y_1) \prod_{j=1}^{p-1} P(y_j, y_{j+1}).$$

It is easily checked that  $\widehat{P}^i$  defines an irreducible and aperiodic Markovian kernel. Now, consider the map

$$\begin{aligned} \widehat{X} : \widehat{E}^i &\rightarrow \Gamma, \\ \widehat{X}(x_1, \dots, x_p) &\mapsto X(x_p) \dots X(x_1). \end{aligned}$$

We construct a Markovian product  $(\widehat{M}_n^i)$  for each  $i = 0, \dots, p - 1$  in the usual way. As  $\widehat{P}^i$  is primitive and  $\widehat{E}^i$  is finite, the Markovian product  $(\widehat{M}_n^i)$  automatically satisfies conditions  $(A_1)$  and  $(A_2)$  (and  $(A'_2)$ ) in the setting of Theorem 4.6). Moreover, for every  $i = 0, \dots, p - 1$  and  $(x_1, \dots, x_p) \in \widehat{E}^i$ , under  $\widehat{P}^i_{(x_1, \dots, x_p)}$ , the distribution of  $(\widehat{M}_n^i)$  is the same as the distribution of  $(M_{pn})$  allowing us to use the products  $(\widehat{M}_n^i)$  to control the product  $(M_n)$ .

We therefore aim to apply the above limit theorems to the Markovian products  $(\widehat{M}_n^i)$  and deduce the corresponding limit theorems for the products  $(M_{pn})_{n \in \mathbb{N}}$ , and then use the fact that the operator norm  $\log \|\cdot\|$  and the displacement  $\kappa(\cdot)$  (in the setting of Theorem 4.6) is subadditive to deduce the same limit theorems for the Markovian product  $(M_n)$  along all times  $n \in \mathbb{N}$ . To this end, we also need to relate 1-contracting and strong irreducibility assumptions on  $(M_n)$  and  $(\widehat{M}_n^i)$ .

**Lemma 4.11.** *The Markovian product  $(M_n)$  is 1-contracting/strong irreducible/nonelementary if and only if the Markovian product  $(\widehat{M}_n^i)$  is, respectively, 1-contracting/ strong irreducible/nonelementary for some (equivalently all)  $i = 0, \dots, p - 1$ .*

The proof is elementary, we briefly indicate the argument.

*Proof.* Suppose  $(M_n)$  is 1-contracting, then there exists  $x \in E$  and a sequence  $g_n \in T_x$  such that  $g_n / \|g_n\|$  converges to a rank one linear transformation. Writing  $g_n$  as a product of elements  $X(y)$  for  $y \in E$  and discarding the last elements to make the length divisible by the period  $p$ , we find a finite set  $F$  and for each  $n \in \mathbb{N}$  an element  $h_n \in F$  such that, if necessary passing to a subsequence,  $h_n^{-1} g_n$  belongs to  $T_{(x, x_1, \dots, x_{p-1})}$  for some  $x_1, \dots, x_{p-1} \in E$ . Up to further passing to a subsequence,  $h_n$  stabilizes and  $h_n^{-1} g_n / \|h_n^{-1} g_n\|$  converges to a rank one transformation. The converse implication (and the statement that some  $i$  is equivalent to all  $i$ ) is clear.

Suppose now that for some  $i = 0, \dots, p - 1$ ,  $(\widehat{M}_n^i)$  is not strongly irreducible. Then for every  $(x_1, \dots, x_p) \in \widehat{E}_i$ , there exists a union of a finite collection of proper subspaces  $W(x_1, \dots, x_p)$  such that  $\widehat{P}_{(x_1, \dots, x_p)_0}^i$  almost surely  $\widehat{M}_n^i W((x_1, \dots, x_p)_0) = W((x_1, \dots, x_p)_n)$ . Using this, first, one verifies that  $W(x_1, \dots, x_p)$  only depends on  $x_p$ . We set  $W(x_p) := W(x_1, \dots, x_p)$  for any  $(x_1, \dots, x_p) \in \widehat{E}^i$  which is hence well-defined. Second, one checks that for any  $x_k \in \widehat{E}^{i+k}$  ( $i + k$  considered modulo  $p$ ), the union of subspaces  $W(x_k) := X(x_k) \dots X(x_1)W(x_p)$  where  $P(x_p, x_1) > 0$  and  $P(x_i, x_{i+1}) > 0$  for every  $i = 1, \dots, k - 1$ , is well-defined (i.e., does not depend on the path  $(x_1, \dots, x_{p-1})$ ). Finally, one verifies that for every  $x_0 \in E$ ,  $\mathbb{P}_{x_0}$  almost surely  $M_n W(x_0) = W(x_n)$ , that is,  $(M_n)$  is not strongly irreducible. The other implications are clear and the statement about nonelementariness is proven in the same way as strong irreducibility.  $\square$

Combining the constructions above and the previous lemma, one readily deduces the following from Theorems 4.2 and 4.3.

**Theorem 4.12.** *Let  $E$  be a finite set,  $P$  an irreducible Markovian transition kernel on  $E$ ,  $X : E \rightarrow \text{GL}_d(\mathbb{R})$  a map, and  $(M_n)$  the associated Markovian product on  $\Gamma$ . Suppose that  $(M_n)$  is strongly irreducible and 1-contracting. Then there exist constants  $\Lambda \geq 0$  and  $\sigma > 0$  such that for every  $x \in E$*

(1) *the sequence of  $C([0, 1])$ -valued random variables defined by*

$$S_n(t) = \frac{1}{(n\sigma^2)^{1/2}} (\log \|M_{[tn]}\| - nt\Lambda + (nt - [nt])(\log \|M_{[tn]+1}\| - \log \|M_{[tn]}\|)) \quad (4.20)$$

*converges to the Wiener measure  $\mathcal{W}$  as  $n \rightarrow \infty$ ;*

(2) *for  $\mathbb{P}_x$ -almost every  $\omega$ , the set of limit points of the sequence  $(\frac{S_n(t)(\omega)}{2 \log \log n})_{n \in \mathbb{N}}$  is the compact set given in Theorem 4.2(2); and*

(3) *for every  $\epsilon > 0$ ,*

$$\mathbb{P}_x(|\log \|M_n\| - n\Lambda| \geq n\epsilon) \leq Ce^{-\alpha n}.$$

*Remark 4.13.* Similarly, using Theorem 4.6 one obtains the following statement for a nonelementary Markovian product on  $\text{Isom}(H)$  associated to a finite irreducible Markov chain: there exists  $\ell_\Lambda > 0$  such that for every  $\epsilon > 0$  and  $x \in E$ , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_x(|d(M_n \cdot o, o) - n\ell_\Lambda| \geq n\epsilon) < 0.$$

The proof is very similar to the proof of the previous result and it is omitted.

*Proof of Theorem 4.12.*

(1) Thanks to Lemma 4.11, we can apply Theorem 4.2 to each  $(\widehat{M}_n^i)$  and get that for every  $i = 0, \dots, p - 1$ , there exist constants  $\sigma_i > 0$  and  $\Lambda_i \in \mathbb{R}$  such that for every  $x = (x_1, \dots, x_p) \in \widehat{E}_i$ , we have that under  $\mathbb{P}_x$ ,

$$\frac{1}{(n\sigma_i^2)^{1/2}} (\log \|\widehat{M}_{[tn]}^i\| - nt\Lambda_i + (nt - [nt])(\log \|\widehat{M}_{[tn]+1}^i\| - \log \|\widehat{M}_{[tn]}^i\|)) \xrightarrow{\mathcal{L}} \mathcal{W}. \quad (4.21)$$



Recall that the distribution (denoted  $\mathcal{L}_{\mathbb{P}_{(x_1, \dots, x_p)}}(\widehat{M}_n^i)$ ) of  $\widehat{M}_n^i$  under  $\mathbb{P}_{(x_1, \dots, x_p)}$  is equal to that of  $M_{np}$  under  $\mathbb{P}_{x_p}$  for each  $n \in \mathbb{N}$  (i.e.,  $\mathcal{L}_{\mathbb{P}_{x_p}}(M_{np})$ ). Therefore, specializing to  $t = 1$  in the previous displayed equation, this implies that for  $x \in E_i$ , under  $\mathbb{P}_x$ , the sequence  $\frac{1}{(n\sigma_i)^2}(\log \|M_{np}\| - n\Lambda_i)$  converges in distribution to the Gaussian  $\mathcal{N}(0, 1)$  as  $n \rightarrow \infty$ . Thanks to the inequality

$$|\log \|gh\| - \log \|g\|| \leq \max\{\log \|h\|, \log \|h^{-1}\|\}, \tag{4.22}$$

we get that under  $\mathbb{P}_x$ ,  $\frac{1}{(n\sigma_i)^2}(\log \|gM_{np}\| - n\Lambda_i) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$  for any fixed  $g \in \text{GL}_d(\mathbb{R})$ . Now using the fact that  $\mathcal{L}_{\mathbb{P}_x}(M_{np+1}) = \sum_{y \in E_{i+1}} P(x, y) \mathcal{L}_{\mathbb{P}_y}(X(y)M_{np})$ , we easily deduce that  $\Lambda_i = \Lambda_0$  and  $\sigma_i = \sigma_0$  for every  $i = 0, \dots, p - 1$ . Therefore, (4.21) together with the equality  $\mathcal{L}_{\mathbb{P}_{(x_1, \dots, x_p)}}(\widehat{M}_n^i) = \mathcal{L}_{\mathbb{P}_{x_p}}(M_{np})$  implies that for every  $x \in E$ , under  $\mathbb{P}_x$

$$\frac{1}{(n\sigma_0^2)^{1/2}}(\log \|M_{p\lfloor nt \rfloor}\| - nt\Lambda_0 + (nt - \lfloor nt \rfloor)(\log \|M_{p\lfloor nt \rfloor + p}\| - \log \|M_{p\lfloor nt \rfloor}\|)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W}.$$

Once more using (4.22) together with the fact that the state space is finite, one gets that for every  $x \in E$ , under  $\mathbb{P}_x$ ,  $\|\widehat{S}_n - S_{np}\|_\infty \rightarrow 0$  in probability (as  $n \rightarrow \infty$ ), where  $S_n$  is defined in (4.20) with  $\sigma := \sigma_0/\sqrt{p}$  and  $\Lambda := \Lambda_0/p$ . This implies (see, e.g., [48, Problem 4.16]) that for every  $x \in E$ , under  $\mathbb{P}_x$

$$S_{np} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W}. \tag{4.23}$$

Once more using the inequality (4.22) and the fact that  $\sigma > 0$ , we observe that for every  $k \in \mathbb{N}$ , and  $x \in E$ , under  $\mathbb{P}_x$ ,  $\|S_{np} - S_{np+k}\| \rightarrow 0$  in probability and hence (4.23) implies that  $S_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W}$  as required.

The proofs of (2) and (3) are proven using the same ideas and are omitted to avoid repetition.  $\square$

## 5 | LIMIT THEOREMS FOR THE STRONGLY MARKOV STRUCTURE

In Subsection 5.1.1, we associate Markovian products to strongly Markov structure  $\mathcal{G}$  of a couple  $(\Gamma, S)$  and representation  $\rho : \Gamma \rightarrow \text{GL}_d(\mathbb{R})$ . We then deduce certain properties (1-contracting and irreducible) of the Markovian product from those of  $\rho$ . In Subsection 5.2, we single out the consequence of Theorem 4.12 for these chains. In Subsection 5.3, we prove that the Markovian products associated to different maximal components of  $\mathcal{G}$  satisfy limit theorems with the same parameters. Finally in Subsection 5.4, we indicate the analogous results for isometries.

### 5.1 | Markovian products associated to the Markov structure of a Gromov-hyperbolic group

#### 5.1.1 | Construction of Markovian products

We will extensively use the notation and terminology introduced in Sections 2 and 4. We fix a Gromov-hyperbolic group  $\Gamma$ , a generating set  $S \subset \Gamma$  and an associated strongly Markov structure

$\mathcal{G}$ . Denote by  $(B_j)_{j=1,\dots,m}$  the maximal components of  $\mathcal{G}$ . For each maximal component, let  $E_j$  denote the set of edges between two vertices of  $V_j$ , the set of vertices in  $B_j$ . The Parry construction discussed in Subsection 3.1.2 gives rise to an irreducible Markovian transition kernel  $P_{v,j}$  on the state space  $V_j$  and the Parry measure denoted  $\mu_j$  is the unique shift invariant probability measure of maximal entropy on the associated trajectory space (subshift). We denote its restriction to  $V_j$  by  $\mu_{v,j}$ . This is a  $P_{v,j}$ -stationary measure on  $V_j$  and in accordance with our notation of trajectory measures we have  $\mu_j = \mathbb{P}_{\mu_{v,j}}$ <sup>†</sup>. To define a Markovian product using the strongly Markov structure, we pass to the associated edge Markov chain: we consider the transition kernel  $P_{e,j}$  on  $E_j$  defined by  $P_{e,j}((v_1, v_2), (v_3, v_4)) = P_{v,j}(v_2, v_3)P_{v,j}(v_3, v_4)$ . It is also irreducible and has the unique stationary measure  $\mu_{e,j}$  given by  $\mu_{e,j}(v_1, v_2) = \mu_{v,j}(v_1)P_{v,j}(v_1, v_2)$ .

Having fixed a representation  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$ , we consider the map  $X$  defined on the state space  $E_j$  by the map  $\lambda(\cdot, \cdot)$  in the strongly Markov structure (see Definition 2.2) and transpose of the representation  $\rho$ , that is,  $X((v_1, v_2)) := {}^t\rho(\lambda(v_1, v_2))$ . These define the data of our Markovian product that we will denote by  $(M_n^j)$  for every  $j = 1, \dots, m$ . We denote the corresponding Lyapunov exponents by  $\Lambda_1(j) \geq \Lambda_2(j) \geq \dots \geq \Lambda_d(j)$  and will sometimes write  $\Lambda(j) = \Lambda_1(j)$  to simplify notation.

Finally, we write  $(M_n)$  for the stationary Markovian product obtained from the Markov chain  $(z_n)$  on the state space  $\cup_{j=1}^m E_j$  and with transition kernel  $P$  defined in the natural way from the  $P_{e,j}$ 's. Note that in general we will deal with the case  $m > 1$  so  $(z_n)$  is not an ergodic Markov chain with any starting distribution that is a nontrivial linear combination of  $\mu_{e,j}$ 's.

### 5.1.2 | Proximality and strong irreducibility of Markovian products

In the following lemma, we use the notation and constructions of the previous paragraph and show that proximal and strongly irreducible representations give rise to proximal and strongly irreducible Markovian products. This relies on key ingredients from the works of Goldsheid–Margulis [36] and Gouëzel–Mathéus–Maucourant [39].

**Lemma 5.1.** *Suppose that  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  is a proximal and strongly irreducible representation. Then, for each  $j = 1, \dots, m$ , the Markovian product  $(M_n^j)$  is 1-contracting and strongly irreducible.*

Recall that a semigroup  $\Lambda < \mathrm{GL}_d(\mathbb{R})$  is irreducible (resp., strongly irreducible) if there does not exist a nontrivial proper  $\Lambda$ -invariant subspace (resp., a finite collection of such subspaces whose union is  $\Lambda$ -invariant). It is not hard to see (see the proof of [12, Theorem 5.3 (ii)]) for every  $j = 1, \dots, m$ ,  $(M_n^j)$  is 1-contracting or strongly irreducible, if and only if, there exists  $x \in E_j$  such that the semigroup  $T_x(x) < \mathrm{GL}_d(\mathbb{R})$  (see Subsection 4.2) is, respectively, proximal or strongly irreducible.

*Proof.* Fix  $j \in \{1, \dots, m\}$  and  $x \in E_j$ . It suffices to show that the semigroup  $T_x(x)$  is proximal and strongly irreducible.

Let us first show that  $T_x(x)$  is strongly irreducible. Let  $v_1, v_2 \in V_j$  such that  $x = (v_1, v_2)$  and denote by  $p_j \in \mathbb{N}$  the period of the kernel  $P_{e,j}$  on  $E_j$ . Then by definition  $T_x(x) = {}^t\rho(\Gamma_x)$ , where

$$\Gamma_x := \{\lambda(v_2, v_3) \dots \lambda(v_{p_j n}, v_1) \lambda(v_1, v_2) : n \in \mathbb{N} \text{ and } (v_i, v_{i+1}) \in E_j \ \forall i = 1, \dots, p_j n\}.$$

<sup>†</sup> We stress this point to avoid any confusion. In the sequel, we will often use  $\mu_j$  instead of  $\mathbb{P}_{\mu_{v,j}}$  to simplify the notation.

We recall that in the preceding,  $\lambda(\cdot, \cdot)$  denotes the labeling map in the definition of strongly Markov structure  $\mathcal{G}$  (Definition 2.2 (ii)). By the property (iii) in Definition 2.2, we have that the subset of  $\Gamma_x$  consisting of elements of  $\Gamma_x$  of  $S$ -length  $p_j n$  is in bijection with the set  $\mathcal{G}_{p_j n}(x)$  paths of vertices of length  $p_j n$  in  $\mathcal{G}$  that are loops around the vertex  $v_1$ . As  $v_1$  belongs to the maximal component  $B_j$ , there exists  $c > 0$  such that  $\#\mathcal{G}_{p_j n}(x) \geq c\lambda^{p_j n}$ , where  $\lambda > 1$  is the growth rate of the group  $\Gamma$ . We therefore have that  $\#(\Gamma_x \cap S_{p_j n}) \geq c\lambda^{p_j n}$  and thanks to the purely exponential growth property (2.4) of Gromov-hyperbolic groups, it follows that the upper asymptotic density of the semigroup  $\Gamma_x$  over the spheres  $S_{p_j n}$  is strictly positive, that is,

$$\limsup_{n \rightarrow \infty} \frac{\#(S_{np_j} \cap \Gamma_x)}{\#S_{np_j}} > 0.$$

As a consequence, by a result of Gouëzel–Mathéus–Maucourant [39, Theorem 4.3], we get that the subgroup  $\Gamma_x^\pm$  generated by  $\Gamma_x$  has finite index in  $\Gamma$ . As  $\rho(\Gamma) < GL_d(\mathbb{R})$  is strongly irreducible and  $\Gamma_x^\pm < \Gamma$  is finite index,  $\rho(\Gamma_x^\pm)$  is also strongly irreducible. As the transpose semigroup  ${}^t\Lambda$  of a strongly irreducible semigroup  $\Lambda < GL_d(\mathbb{R})$  is also strongly irreducible, it follows that the semigroup  ${}^t\rho(\Gamma_x) = T_x(x)$  is strongly irreducible, as required.

It remains to show that  ${}^t\rho(\Gamma_x) = T_x(x) < GL_d(\mathbb{R})$  is a proximal semigroup. It suffices to show again that the transpose semigroup  $\rho(\Gamma_x)$  is proximal. By a result of Goldsheid–Margulis [36] (for a version we use, see [8, Lemma 6.23]), it suffices to show that the Zariski-closure  $H := \overline{\rho(\Gamma_x)}^Z < GL_d(\mathbb{R})$  is proximal. Recall that the Zariski-closure of a semigroup is a group and so  $H = \overline{\rho(\Gamma_x^\pm)}^Z$ . But as  $\Gamma_x^\pm$  has finite index in  $\Gamma$ , denoting by  $G = \overline{\rho(\Gamma)}^Z$ , we have the equality of connected components  $G^o = H^o$ . As  $\rho(\Gamma)$  is proximal by hypothesis so is  $G^o$  and consequently  $H$ , completing the proof. □

## 5.2 | Limit theorems for the maximal components of the strongly Markov structure

We now put together the construction of Markovian products  $(M_n^j)$  in Subsection 5.1.1, Lemma 5.1, and Theorem 4.12 to deduce the following limit laws on the maximal components of the strongly Markov structure. For easier referencing, we state them separately.

**Proposition 5.2** (Large deviations on maximal components). *Under the assumptions of Theorem 1.11, for each  $j = 1, \dots, m$  and for every  $x \in E_j$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_x(|\log \|M_n^j\| - n\Lambda(j)| > n\varepsilon) < 0,$$

where  $\Lambda(j)$  is the top Lyapunov exponent of the Markovian product  $(M_n^j)$ .

**Proposition 5.3** (Convergence to the Wiener measure on maximal components). *Under the assumptions of Theorem 1.12, for every  $j = 1, \dots, m$ ,  $\sigma > 0$ ,  $\Lambda \in \mathbb{R}$ ,  $t \in [0, 1]$  and  $n \in \mathbb{N}$ , let  $S_n^j(t)$*

denote  $C([0, 1])$ -valued random variable defined by

$$S_n^j(t) = \frac{1}{(n\sigma^2)^{1/2}} \left( \log \|M_{[tn]}^j\| - nt\Lambda + (nt - [nt])(\log \|M_{[tn]+1}^j\| - \log \|M_{[tn]}^j\|) \right). \quad (5.1)$$

Then, there exists  $\sigma = \sigma_j > 0$  such that for  $\Lambda = \Lambda(j) \in \mathbb{R}$  and for every  $x \in E_j$ , under  $\mathbb{P}_x$ , the sequence  $(S_n^j)_{n \in \mathbb{N}}$  of  $C([0, 1])$ -valued random variables converges in distribution to  $\mathcal{W}$ .

Finally, we record the following.

**Proposition 5.4** (Law of iterated logarithm on maximal components). *Keep the hypotheses and notation of Proposition 5.3 and let  $\sigma_j > 0$  and  $\Lambda(j) \in \mathbb{R}$  be the constants given by that result. Then, for every  $j = 1, \dots, m$ ,  $x \in E_j$ , for  $\mathbb{P}_x$ -almost every  $\omega$ , the set of limit points of the sequence  $(\frac{S_n^j(t)(\omega)}{2 \log \log n})_{n \in \mathbb{N}}$  of elements of  $C([0, 1])$  is equal to the following compact subset of  $C([0, 1])$ :*

$$\left\{ f \in C([0, 1]) : f \text{ is absolutely continuous, } f(0) = 0, \int_0^1 f'(t)^2 dt \leq 1 \right\}.$$

### 5.3 | Comparing means and variances

To upgrade Propositions 5.2, 5.3, and 5.4 to the corresponding limit theorems on the full strongly Markov structure  $\mathcal{G}$ , we first need to show that the Lyapunov exponents  $\Lambda_k(j)$  and variances  $\sigma_j^2$  obtained in the previous section all agree (i.e., they do not depend on  $j = 1, \dots, m$ ). This result is also needed to prove the positivity of the top Lyapunov exponent. These are the two goals of this paragraph.

To compare these Lyapunov exponents and variances across maximal components we implement the approach of Calegari–Fujiwara [18] (more precisely, its adaptation by Cantrell [21, Proposition 4.8]). The argument crucially relies on the ergodicity of the Patterson–Sullivan measure to compare typical growth rates of appropriately constructed functions along geodesic rays.

#### Proposition 5.5.

- (1) *There exists constants  $\Lambda_1 \geq \Lambda_2 \geq \dots \geq \Lambda_d$  such that for every  $j = 1, \dots, m$  and  $i = 1, \dots, d$  we have  $\Lambda_i(j) = \Lambda_i$ .*
- (2) *There exists a constant  $\sigma > 0$  such that for every  $j = 1, \dots, m$ , we have  $\sigma_j^2 = \sigma^2$ .*

In the sequel, whenever there is no risk of confusion, we will denote  $\Lambda_1$  by  $\Lambda$ .

*Proof.*

- (1) The proof is similar to that of Theorem 1.10 and hence omitted to avoid repetitive exposition of the same idea.
- (2) Fix a maximal component  $B_j$  and define  $S_j \subset \partial\Gamma$  to be Borel measurable subset of  $\partial\Gamma$  consisting of boundary elements  $\xi$  that have a geodesic representative  $(\xi_n)$  such that for

each  $t \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \limsup_{m \rightarrow \infty} \frac{1}{m} \# \left\{ 0 \leq k \leq m : \frac{\log \|\rho(\xi_k^{-1} \xi_{k+n})\| - \Lambda n}{\sqrt{n}} < t \right\} = \frac{1}{\sigma_j \sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2\sigma_j^2}} ds.$$

The set  $S_j$  is well-defined and  $\Gamma$ -invariant. Given  $t \in \mathbb{R}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we also define the set

$$F(t, n) = \left\{ x \in \Sigma_{B_j} : \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\| - \Lambda n}{\sqrt{n}} < t \right\}.$$

Here, without loss of generality, we choose the norm  $\|\cdot\|$  to be the operator norm induced by the Euclidean norm so that it is invariant under passing to the transpose of a matrix. Moreover, for  $z \in \Sigma_A$  and  $m \in \mathbb{Z}_{\geq 0}$ , we set

$$\mu_{(z,m)} = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{\sigma^k z}.$$

As the indicator functions  $1_{F(t,n)}$  are continuous and  $C(\Sigma_A)$  separable, using Birkhoff's ergodic theorem for the shift space  $(\Sigma_{B_j}, \mu_j)$  where  $\mu_j$  is the (ergodic) Parry measure of  $\Sigma_{B_j}$ , we find a set  $\Sigma'_{B_j} \subset \Sigma_{B_j}$  of full  $\mu_j$  measure such that for every  $z \in \Sigma'_{B_j}$ ,  $t \in \mathbb{R}$  and  $n \in \mathbb{N}$ , we have

$$\lim_{m \rightarrow \infty} \int_{\Sigma_{B_j}} 1_{F(t,n)} d\mu_{(z,m)} = \mu_j(F(t, n)). \tag{5.2}$$

Notice that, as in the first part of the proof, if  $z \in \Sigma_{B_j}$  satisfies the convergence (5.2), then any pre-image in  $\sigma^{-k}(z)$  (for any  $k \geq 1$ ) also satisfies the same convergence. Hence, thanks to (3.3) and Lemma 3.1, we can find  $k \geq 1$  and a subset  $\Sigma^o_{B_j} = \sigma^{-k}(\Sigma'_{B_j}) \subset \Sigma_A$  such that  $\hat{\nu}(\Sigma^o_{B_j}) > 0$  and for every  $z \in \Sigma^o_{B_j}$ ,  $t \in \mathbb{R}$  and  $n \in \mathbb{N}$ , we have

$$\lim_{m \rightarrow \infty} \int 1_{F(t,n)} d\mu_{(z,m)} = \mu_j(F(t, n)).$$

Notice that by construction of the edge-chain in Subsection 5.1.1 from the vertex chain, we have

$$\mu_j(F(t, n)) = \mathbb{P}_{\mu_{e,j}} \left\{ ((x_0, x_1), \dots) \in E_j : \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\| - \Lambda n}{\sqrt{n}} < t \right\}. \tag{5.3}$$

As the operator norm is invariant under transpose, by construction of the Markovian random product  $(M_n^j)$ , the right-hand side of (5.3) is equal to  $\mathbb{P}_{\mu_{e,j}}(\frac{\log \|M_n^j\| - \Lambda n}{\sqrt{n}} < t)$ . Using the CLT implied by Proposition 5.3 (e.g., by specializing to  $t = 1$  in (5.1) and using the definition of the Wiener measure) then implies that

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \int 1_{F(t,n)} d\mu_{(z,m)} = \lim_{n \rightarrow \infty} \mu_j(F(t, n)) = \frac{1}{\sigma_j \sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2\sigma_j^2}} ds$$

for  $t \in \mathbb{R}$ . Therefore, we deduce that  $\Psi(\Sigma_{B_j}^o) \subset S_j$ , where  $\Psi$  is the function  $\Psi : Y_\infty \rightarrow \partial\Gamma$  defined in Subsection 3.1.1. In particular, we have  $\nu(S_j) > 0$  and by the ergodicity of  $\nu$  this implies that  $\nu(S_j) = 1$  and so  $\sigma_j^2$  does not depend on  $j = 1, \dots, m$ , as required.  $\square$

A direct consequence of the previous result is the following.

**Corollary 5.6.** *Propositions 5.2, 5.3, and 5.4 hold when  $(M_n^j)$  is replaced by  $(M_n)$  (see Subsection 5.1.1) and the constants  $\Lambda(j)$  and  $\sigma_j^2$  are replaced by  $\Lambda$  and  $\sigma^2$  for each  $j = 1, \dots, m$ .*

In Proposition 5.5, we assumed that our representation  $\rho$  is both strongly irreducible and proximal. However, the argument used to prove the first part of this proposition does not require either the strongly irreducible or proximal assumption. We obtain the following which applies to any representation of a Gromov-hyperbolic group into  $\mathrm{GL}_d(\mathbb{R})$ .

**Lemma 5.7.** *Let  $\rho : \Gamma \rightarrow \mathrm{GL}_d(\mathbb{R})$  be a representation of a hyperbolic group  $\Gamma$  (which is equipped with a generating set). Then the Borel subset  $B \subseteq \partial\Gamma$  consisting of elements  $\xi$  having a geodesic representative  $(\xi_n)$  satisfying*

$$\frac{1}{n} \log \left\| \bigwedge^k \rho(\xi_n) \right\| \xrightarrow{n \rightarrow \infty} \sum_{i=1}^k \Lambda_i$$

for every  $k = 1, \dots, d$ , is well-defined,  $\Gamma$ -invariant and has full  $\nu$ -mass.  $\square$

We can now characterize the positivity of  $\Lambda$  for strongly irreducible representations as claimed in the introduction.

*Proof of Proposition 1.4.* To prove the necessity, note that if the image of  $\Gamma$  in  $\mathrm{PGL}_d(\mathbb{R})$  is relatively compact, then we can modify the norm  $\|\cdot\|$  so that the map  $\Gamma \ni \gamma \rightarrow \log \|\rho(\gamma)\| \in \mathbb{R}$  is additive. Now using Lemma 5.7 and Remark 3.2, we realize  $\Lambda$  as a counting average (as in Theorem 1.1) with respect to an additive function. The symmetry of  $S$  readily implies that this counting average is zero.

Let us now show the remaining implication. Suppose that the image of  $\Gamma$  is not-relatively compact in  $\mathrm{PGL}_d(\mathbb{R})$ . As  $\rho(\Gamma) < \mathrm{GL}_d(\mathbb{R})$  is (strongly) irreducible, it follows that the semi-group  $\rho(\Gamma)$  is  $r$ -proximal for some  $r \in \{1, \dots, d-1\}$  (this is standard, see, e.g., [56, Lemma 3.6]). It then follows by the same argument in Lemma 5.1 that  $(M_n^j)$  is  $r$ -contracting for each  $j = 1, \dots, m$ . Therefore, since by Proposition 5.5, the  $\Lambda_i$ 's are the Lyapunov exponents of the Markovian product  $(M_n^1)$  (which satisfies the assumptions of Theorem 4.1) there exists  $r \in \{1, \dots, d-1\}$  such that  $\Lambda_r > \Lambda_{r+1}$ . We now relate these Lyapunov exponents with spherical averages to exploit symmetry of the generating set  $S$  to get positivity. To this end, we apply Theorem 1.10 for the subadditive functions  $\phi_1(\cdot) = \log \|\rho(\cdot)\|$  and  $\phi_{\det}(\cdot) = \log \det(\rho(\cdot))$  and denote the corresponding averages by  $\tilde{\Lambda}_1$  and  $\tilde{\Lambda}_{\det}$ , respectively. In view of Lemma 5.7, we have  $\tilde{\Lambda}_1 = \Lambda_1$  and  $\tilde{\Lambda}_{\det} = \sum_{i=1}^d \Lambda_i$ . Now, on the one hand by Remark 3.2,  $\Lambda_1$  and  $\sum_{i=1}^d \Lambda_i$  are nonnegative, and on the other hand, we have  $\Lambda_1 \geq \dots \geq \Lambda_r > \Lambda_{r+1} \geq \dots \geq \Lambda_d$ . It follows that  $\Lambda = \Lambda_1 > 0$ .  $\square$

### 5.4 | The case of isometries

Here we briefly indicate how to associate a Markovian product (and the result corresponding to Proposition 5.2 and Corollary 5.6) in the analogous situation where, instead of a representation  $\Gamma \rightarrow GL_d(\mathbb{R})$ , we are given a nonelementary isometric action of  $\Gamma$  on a Gromov-hyperbolic space  $(H, d)$ .

For each maximal component  $B_j, j = 1, \dots, m$ , the underlying Markov chain  $(E_j, P_{e,j})$  described in Subsection 5.1.1 remains the same. One only modifies the map  $X$ . We define  $X : E_j \rightarrow \text{Isom}(H)$  by  $X((v_1, v_2)) = \lambda(v_1, v_2)^{-1}$ . We similarly denote by  $(M_n^j)$  the associated Markovian product on  $\text{Isom}(H)$  and  $(M_n)$  the Markovian product induced by the Markov chain  $(z_n)$  on the state space  $\cup_{j=1}^m E_j$  and with transition kernel  $P$  defined in the natural way from the  $P_{e,j}$ 's.

Fix  $j = 1, \dots, m$  and  $x \in E_j$ . One checks exactly as in the same way as Lemma 5.1 that the semigroup  $T_x(x) < \text{Isom}(H)$  is nonelementary. This implies that the Markovian product  $(M_n^j)$  in  $\text{Isom}(H)$  is nonelementary and has positive drift (see Remark 4.7). Moreover, being defined over an irreducible Markov chain with finite state space, it clearly satisfies Conditions  $(A_1)$  and  $(A'_2)$ . In view of Remark 4.13, we deduce that there exists a constant  $\ell_{\Lambda_j} > 0$  such that for every  $\epsilon > 0$ , there exists  $\alpha > 0$  and  $C > 0$  such that for every  $x \in E_j, \xi \in \overline{H}^h$  and  $n \in \mathbb{N}$ , we have

$$\mathbb{P}_x(|\sigma(M_n^j, \xi) - n\ell_{\Lambda_j}| \geq n\epsilon) \leq Ce^{-\alpha n}. \tag{5.4}$$

Specializing to  $\xi = o$ , the previous inequality shows by Borel–Cantelli (or directly by the subadditive ergodic theorem) that  $\mathbb{P}_x$ -almost surely  $\frac{1}{n}d(M_n^j \cdot o, o) \rightarrow \ell_{\Lambda_j}$ . Now, the proof of Proposition 5.5 goes through and shows that on one hand the constant  $\ell_{\Lambda_j}$  does not depend on the maximal component  $B_j$  for  $j = 1, \dots, m$  (and hence we denote this constant by  $\ell_{\Lambda}$ ), and on the other hand, it coincides with the constant  $\Lambda$  given by Theorem 1.10 applied with the subadditive function  $\phi(\gamma) = d(\gamma \cdot o, o)$  (as Lemma 5.7). The former fact together with (5.4) gives the following analogue of Corollary 5.6 (the result corresponding to large deviations, that is, Proposition 5.2):

**Proposition 5.8.** *There exists a constant  $\ell_{\Lambda} > 0$  such that for every  $\epsilon > 0$  and  $x \in \cup_{j=1}^m E_j$ , we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_x \left( \left| \frac{1}{n}d(M_n \cdot o, o) - \ell_{\Lambda} \right| > \epsilon \right) < 0. \quad \square$$

## 6 | LARGE DEVIATION THEOREMS

In Subsection 6.1, we first prove our counting large deviation theorem (Theorem 1.5) assuming Theorem 1.11. We subsequently prove our boundary large deviation result (Theorem 1.11) with respect to the Patterson–Sullivan measure, obtaining a quantitative version of our boundary strong law of large numbers (Theorem 1.10) in the current setting. In Subsection 6.2, we indicate the proofs of the analogous results in the case of isometries of Gromov-hyperbolic spaces.

### 6.1 | Large deviations for representations

We now give the proof of Theorem 1.5 using Theorem 1.11 which will be proven subsequently.



*Proof of Theorem 1.5.* Let  $\Lambda$  be given by Theorem 1.11 and let  $\epsilon > 0$  be fixed. Recall from (3.4) that for any sufficiently large  $R > 0$ , there exist positive constants  $C_1$  and  $C_2$  such that for every  $n \in \mathbb{N}$ ,

$$\frac{\#(S_n \cap A_\epsilon)}{\#S_n} \leq C_1 \sum_{x \in S_n \cap A_\epsilon} \nu(O(x, R)) \leq C_2 \nu\left(\bigcup_{x \in S_n \cap A_\epsilon} O(x, R)\right),$$

where

$$A_\epsilon = \left\{ g \in \Gamma : \left| \frac{\log \|\rho(g)\|}{|g|_S} - \Lambda \right| > \epsilon \right\}.$$

As in the proof of Theorem 1.1, if  $\xi \in \bigcup_{x \in S_n \cap A_\epsilon} O(x, R)$ , then there is  $C > 0$  such that for any representative geodesic ray  $(\xi_m)_{m=0}^\infty$  of  $\xi$  with  $\xi_0 = \text{id}$  and every  $n \geq 1$

$$\left| \frac{\log \|\rho(\xi_n)\|}{n} - \Lambda \right| \geq \epsilon - \frac{C}{n}. \tag{6.1}$$

In particular for all sufficiently large  $n$ ,

$$\begin{aligned} \frac{\#(S_n \cap A_\epsilon)}{\#S_n} &\leq C_2 \nu\left(\bigcup_{x \in S_n \cap A_\epsilon} O(x, R)\right) \\ &\leq C_2 \nu\left(\xi \in \partial\Gamma : \text{for all } \xi_m \rightarrow \xi \text{ with } \xi_0 = \text{id}, \left| \frac{\log \|\rho(\xi_n)\|}{n} - \Lambda \right| > \frac{\epsilon}{2}\right). \end{aligned}$$

The result now follows from Theorem 1.11. □

The rest of this section is devoted to the proof of Theorem 1.11 which will make key use of Bougerol’s Theorem 4.3 in the form of Proposition 5.2 (and Corollary 5.6).

*Proof of Theorem 1.11.* Let  $\Lambda$  be the constant given by Theorem 1.10 applied with the function  $\phi(\gamma) = \log \|\rho(\gamma)\|$ , where  $\|\cdot\|$  is the operator norm induced by the Euclidean norm on  $\mathbb{R}^d$  (in particular, it is invariant under the transpose). For any  $\epsilon > 0$  and  $n \geq 1$ , we define the sets

$$U_n(\epsilon) = \left\{ \xi \in \partial\Gamma : \text{for all } \xi_m \rightarrow \xi \text{ with } \xi_0 = \text{id}, \left| \frac{\log \|\rho(\xi_n)\|}{n} - \Lambda \right| > \epsilon \right\}$$

and

$$E_n(\epsilon) = \left\{ (x_m)_{m=0}^\infty \in Y_\infty : \left| \frac{\log \|\rho(\lambda(*, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \epsilon \right\}.$$

Note that  $\Psi^{-1}(U_n(\epsilon)) \subseteq E_n(\epsilon)$  and consequently  $\nu(U_n(\epsilon)) \leq \hat{\nu}(E_n(\epsilon))$ . Therefore, to prove Theorem 1.11 it suffices to show that for every  $\epsilon > 0$ ,  $\hat{\nu}(E_n(\epsilon)) \rightarrow 0$  exponentially quickly as  $n \rightarrow \infty$ , which is what we shall prove in the sequel.

To proceed, for every integer  $i \geq 1$ , we define

$$A_i = \left( \sigma^{-i} \left( \bigcup_{j=1}^m \Sigma_{B_j} \right) \setminus \bigcup_{k=0}^{i-1} \sigma^{-k} \left( \bigcup_{j=1}^m \Sigma_{B_j} \right) \right) \cap Y \tag{6.2}$$

where, as before,  $B_j$  for  $j = 1, \dots, m$  denote the maximal components. Intuitively, each  $A_i$  consists of elements in  $Y$  that correspond to a path in  $\mathcal{C}$  that starts at  $*$ , enters a maximal component exactly on its  $i$ th step and then never leaves this component. For each  $n \in \mathbb{N}$ , we let  $\tilde{\nu}_n$  to be the measure on  $Y$  given by the restriction of  $\hat{\nu}$  on  $\bigcup_{i=1}^n A_i$ , that is, for every Borel set  $R \subseteq Y$ ,

$$\tilde{\nu}_n(R) = \hat{\nu} \left( R \cap \bigcup_{i=1}^n A_i \right). \tag{6.3}$$

We then have the following.

**Lemma 6.1** (Lemma 4.8 [22]). *There exists  $0 < \theta < 1$  such that  $\|\tilde{\nu}_n - \hat{\nu}\|_{TV} = O(\theta^n)$ , as  $n \rightarrow \infty$ .*

It follows from this lemma that for any  $\epsilon' > 0$  there exist constants  $0 < \theta = \theta(\epsilon') < 1$  and  $C_0 > 0$  such that

$$\hat{\nu}(E_n(\epsilon)) \leq \tilde{\nu}_{\epsilon' n}(E_n(\epsilon)) + C_0 \theta^n \leq \sum_{k=1}^{n\epsilon'} \hat{\nu}(E_n(\epsilon) \cap A_k) + C_0 \theta^n \tag{6.4}$$

for every  $\epsilon > 0$  and  $n \in \mathbb{N}$ . Here and throughout the rest of this section we write  $\epsilon' n$  instead of  $\lfloor \epsilon' n \rfloor$  to simplify notation. We now turn our attention to studying each  $\hat{\nu}(E_n(\epsilon) \cap A_k)$ .

**Lemma 6.2.** *For every  $\epsilon > 0$ , there exist positive constants  $\epsilon'$  and  $C_1$  such that for all  $n \in \mathbb{N}$ ,  $n\epsilon' \geq k \geq 1$ , we have*

$$\hat{\nu}(E_n(\epsilon) \cap A_k) \leq C_1 \mu \left\{ x \in \bigcup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\}.$$

*Proof.* Given  $\epsilon > 0$ , fix  $\epsilon' > 0$  so that  $2\epsilon' \max_{s \in S} \{\log \|\rho(s)\|\} < \epsilon/2$ . Then, for each  $n\epsilon' \geq k \geq 1$ , we have that  $E_n(\epsilon) \cap A_k$  is given by

$$\begin{aligned} & \left\{ x \in Y_\infty : \left| \frac{\log \|\rho(\lambda(*, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \epsilon, \sigma^k x \in \bigcup_j \Sigma_{B_j}, \sigma^{k-1} x \notin \bigcup_j \Sigma_{B_j} \right\} \\ & \subseteq \left\{ x \in Y_\infty : \left| \frac{\log \|\rho(\lambda(x_k, x_{k+1}) \dots \lambda(x_{k+n-1}, x_{k+n}))\|}{n} - \Lambda \right| > \frac{\epsilon}{2}, \sigma^k x \in \bigcup_j \Sigma_{B_j} \right\} \tag{6.5} \\ & = Y_\infty \cap \sigma^{-k} \left\{ x \in \bigcup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\}. \end{aligned}$$

The inclusion above follows from Lemma 2.1 (due to the submultiplicativity of the operator norm) and the choice of  $\epsilon'$ . Letting  $V_{\max}$  denote the collection of vertices belonging to a maximal

component, it follows that

$$\begin{aligned} \widehat{\nu}(E_n(\epsilon) \cap A_k) &\leq \sigma_*^k \widehat{\nu} \left\{ x \in \cup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\} \\ &= \sum_{\nu \in V_{\max}} \sigma_*^k \widehat{\nu}|_{[\nu]} \left\{ x \in \cup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\} \\ &= \sum_{\nu \in V_{\max}} \alpha_\nu^k \mu|_{[\nu]} \left\{ x \in \cup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\}, \end{aligned}$$

where  $\alpha_\nu^k$  are the constants from Lemma 3.1. We recall now (from the construction of the Parry measure  $\mu$ ) that a vertex  $\nu$  belongs to  $V_{\max}$  if and only if  $\mu[\nu] > 0$ . In particular, for  $\nu \in V_{\max}$ ,

$$\alpha_\nu^k = \frac{\widehat{\nu}(\sigma^{-k}[\nu])}{\mu([\nu])} \leq \max_{\nu \in V_{\max}} \frac{1}{\mu([\nu])} < \infty$$

and so we deduce that there exists  $C_1 > 0$  such that

$$\widehat{\nu}(E_n(\epsilon) \cap A_k) \leq C_1 \mu \left\{ x \in \cup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\}$$

as required. □

We now complete the proof of Theorem 1.11. Fix  $\epsilon > 0$  and let  $\epsilon' > 0$  be as in Lemma 6.2. By (6.4), there exist constants  $0 < \theta < 1$  and  $C_0 > 0$  such that

$$\widehat{\nu}(E_n(\epsilon)) \leq \sum_{k=1}^{n\epsilon'} \widehat{\nu}(E_n(\epsilon) \cap A_k) + C_0 \theta^n$$

and so by Lemma 6.2 there is  $C_1 > 0$  such that

$$\widehat{\nu}(E_n(\epsilon)) \leq C_1 n \epsilon' \mu \left\{ x \in \cup_j \Sigma_{B_j} : \left| \frac{\log \|\rho(\lambda(x_0, x_1) \dots \lambda(x_{n-1}, x_n))\|}{n} - \Lambda \right| > \frac{\epsilon}{2} \right\} + C_0 \theta^n. \tag{6.6}$$

Recall that by Lemma 5.7 (and Corollary 5.6), the constant  $\Lambda$  is also the top Lyapunov exponent of the Markovian product  $(M_n)$ . We now apply Corollary 5.6 (statement corresponding to Proposition 5.2) which says precisely that the  $\mu$ -measure of the set in the first term of the right-hand side of (6.6) decays exponentially fast in  $n$ , concluding the proof. □

*Remark 6.3.* It is also possible to prove Theorem 1.5 using an approximation argument in which one compares the Markov measures on  $\mathcal{G}$  to the counting measures on  $S_n$ . This method, which would avoid proving Theorem 1.11, is used in Section 8 to prove our counting CLT. We presented the above proof instead as we believe Theorem 1.11 is interesting in its own right.

## 6.2 | Large deviations for isometries

This section is devoted to the proof of Theorem 1.7. As in the proof of Theorem 1.5, we deduce Theorem 1.7 from a boundary large deviation result: Theorem 6.4.

The proof of Theorem 1.7 (resp., Theorem 6.4) follows a very similar line as the proof of Theorem 1.5 (resp., Theorem 1.11). Therefore, for brevity, we will only point out the needed modifications in the proofs. Let us start with the boundary version.

**Theorem 6.4.** *Let  $\Gamma$  be a Gromov-hyperbolic group,  $S$  a generating set of  $\Gamma$  and  $(H, d)$  a geodesic Gromov-hyperbolic space and  $o \in H$  a basepoint. Suppose that  $\Gamma$  acts on  $H$  by isometries and that the action is nonelementary. Let  $\nu$  be a Patterson–Sullivan measure on  $\partial\Gamma$  for the  $S$  word metric. Then there exists a constant  $\Lambda > 0$  such that for any  $\epsilon > 0$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \nu \left( \xi \in \partial\Gamma : \text{for all } \xi_m \rightarrow \xi \text{ with } \xi_0 = \text{id}, \left| \frac{d(g \cdot o, o)}{n} - \Lambda \right| > \epsilon \right) < 0.$$

This result implies Theorem 1.7. The proof of this implication is precisely as in the proof of Theorem 1.5, one only needs to replace the occurrences of  $\log \|\rho(\star)\|$  by  $d(\star \cdot o, o)$ .

For Theorem 6.4, similarly, the proof of Theorem 1.11 goes through until the point at the end where we applied Corollary 5.6. One only has to replace this result by Proposition 5.8: the analogous Markovian limit law but for the isometric actions (instead of representations) that we now consider. This completes the proof.

## 7 | WIENER PROCESS AND THE LAW OF THE ITERATED LOGARITHM ON THE BOUNDARY

The goal of this section is to prove Theorem 1.12: convergence to the Wiener process and the functional LIL.

Before starting the proof, we recall the notion of tightness that will be used therein. For  $t \in [0, 1]$ , let  $E_t : C([0, 1]) \rightarrow \mathbb{R}$  denote the map that evaluates a function at  $t$ . We say that a sequence of probability measures  $\eta_n$  on  $C([0, 1])$  is tight if

- (i)  $\sup_{n \in \mathbb{N}} E_{0*} \eta_n(\mathbb{R} \setminus [-\lambda, \lambda]) \rightarrow 0$  as  $\lambda \rightarrow +\infty$ ; and
- (ii)  $\lim_{\delta \rightarrow 0} \sup_{n \in \mathbb{N}} \eta_n(\sup_{|t-s| \leq \delta} |X(t) - X(s)| \geq \epsilon) = 0$  for every  $\epsilon > 0$ , where  $X$  denotes a random variable with distribution  $\eta_n$  and  $t, s$  range over  $[0, 1]$ .

In the proof below, the distribution  $\eta_n$  will correspond to the pushforward of the Patterson–Sullivan measure  $\nu$  by the map  $S_n$  defined in (1.1).

*Proof of Theorem 1.12.*

- (1) To prove the first statement we need to show that the laws of the sequence  $(S_n)$  is a tight family and also that finite-dimensional distributions of this sequence converge to the finite-dimensional distributions of the Wiener measure on  $C([0, 1])$  (see, e.g., [48, Theorem 4.15]). Without loss of generality, we can assume that the operator norm in the definition of  $(S_n)$  is induced by the Euclidean norm.

Let us start by showing that the distributions of  $(S_n)$  constitute a tight family of measures on  $C([0, 1])$ . Notice that we only need to check the second condition in the definition of tightness above, since by construction  $S_n \xi(0) = 0$  for  $\nu$ -almost every  $\xi \in \partial\Gamma$ . Fix  $\epsilon > 0$ . For every  $\delta > 0$ , define

$$U_n(\epsilon, \delta) = \left\{ \xi \in \partial\Gamma : \sup_{|t-s|<\delta} |S_n \xi(t) - S_n \xi(s)| > \epsilon \right\}.$$

For  $x \in \Sigma_A$ , let  $\widehat{S}_n x$  denote the element of  $C([0, 1])$  defined in the same way as in (1.1) where for  $k \in \mathbb{N}$ ,  $\rho(\xi_k)$  is replaced by  $\rho(\lambda(x_0, x_1), \dots, \lambda(x_{k-1}, x_k))$ . Let us also similarly define

$$E_n(\epsilon, \delta) = \left\{ x \in Y_\infty : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon \right\}. \tag{7.1}$$

As any two representatives  $\xi_m$  and  $\xi'_m$  of an element  $\xi \in \partial\Gamma$  stay at bounded  $S$ -distance, it is easy to check that there exists  $n_0 = n_0(\epsilon) \in \mathbb{N}$  such that for every  $n \geq n_0$  and  $\delta > 0$ , we have  $\Psi^{-1}(U_n(\epsilon, \delta)) \subseteq E_n(\epsilon/2, \delta)$ . Consequently, for every  $n \geq n_0$  and  $\delta > 0$ , we have  $\nu(U_n(\epsilon, \delta)) \leq \widehat{\nu}(E_n(\epsilon/2, \delta))$ . Therefore, to show that the set of distributions of  $S_n$  is tight, it suffices to prove that  $\lim_{\delta \rightarrow 0} \sup_{n \in \mathbb{N}} \widehat{\nu}(E_n(\epsilon/2, \delta)) = 0$ .

We use the following strategy to complete the proof of tightness: we show that for large  $n \in \mathbb{N}$ , the distributions of  $\widehat{S}_n$  (under  $\widehat{\nu}$ ) are approximated by that of the Markovian products in Proposition 5.3 (or more generally Corollary 5.6) which themselves constitute a tight family (as they converge to the Wiener measure) and for small  $n \in \mathbb{N}$  we exploit the fact that jumps of  $S_n(\xi)(t)$  are bounded (for  $n$  bounded) because they are normalized matrix norms of bounded-length products of elements of the finite set  $S$ .

It follows from Lemma 6.1 that there exist constants  $C_0 > 0$  and  $\theta \in (0, 1)$  such that for every  $m \in \mathbb{N}$ , we have

$$\widehat{\nu}(E_n(\epsilon/2, \delta)) \leq \widetilde{\nu}_m(E_n(\epsilon/2, \delta)) + C_0 \theta^m = \sum_{k=1}^m \widehat{\nu}(E_n(\epsilon/2, \delta) \cap A_k) + C_0 \theta^m, \tag{7.2}$$

where the measures  $\widetilde{\nu}_m$  and sets  $A_k$  are as defined in (6.2) and (6.3). We will now require the following observation which is an analogue of Lemma 6.2.

**Lemma 7.1.** *There exist constants  $c > 0$  and  $C_1 > 0$  such that for every  $n \in \mathbb{N}$ ,  $cn^{1/2} \geq k \geq 1$  and  $\delta > 0$ , we have*

$$\widehat{\nu}(E_n(\epsilon/2, \delta) \cap A_k) \leq C_1 \mu \left\{ x \in \cup_j \Sigma_{B_j} : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon/4 \right\}.$$

*Proof.* Fix  $c > 0$  so that  $2c \max_{s \in S} \{\log \|\rho(s)\|\} < \epsilon/4$ . Then, for each  $cn^{1/2} \geq k \geq 1$ , the set  $E_n(\epsilon/2, \delta) \cap A_k$  satisfies

$$\left\{ x \in Y_\infty : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon/2, \sigma^k x \in \cup_j \Sigma_{B_j}, \sigma^{k-1} x \notin \cup_j \Sigma_{B_j} \right\}$$

$$\begin{aligned} &\subseteq \left\{ x \in Y_\infty : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon/4, \sigma^k x \in \cup_j \Sigma_{B_j} \right\} \\ &= Y_\infty \cap \sigma^{-k} \left\{ x \in \cup_j \Sigma_{B_j} : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon/4 \right\}. \end{aligned}$$

The inclusion in the second line above follows from Lemma 2.1 (due to the submultiplicativity of the operator norm) and the choice of  $c$ . From this point on, the proof follows the same lines as the proof of Lemma 6.2. We omit it to avoid repetition.  $\square$

From the previous lemma, we deduce the analogue of (6.6) which reads as follows: for every  $n \in \mathbb{N}$ ,  $\delta > 0$ ,  $cn^{1/2} \geq m \geq 1$ , we have

$$\widehat{\nu}(E_n(\epsilon/2, \delta)) \leq C_1 m \mu \left\{ x \in \cup_j \Sigma_{B_j} : \sup_{|t-s|<\delta} |\widehat{S}_n x(t) - \widehat{S}_n x(s)| > \epsilon/4 \right\} + C_0 \theta^m. \tag{7.3}$$

Let  $\eta > 0$  be arbitrary. Fix  $m \in \mathbb{N}$  large enough so that  $C_0 \theta^m < \eta/2$ . Now, by Corollary 5.6 (as the operator norm is invariant under the transpose) the pushforward of  $\mu$  by  $\widehat{S}_n$  converges to the Wiener measure. These pushforwards are tight and hence we can choose  $\delta_1 > 0$  small enough so that for every  $n \geq 1$ , the  $\mu$ -measure on the right-hand side of (7.3) is less than  $\frac{\eta}{2C_1 m}$  for every  $n \geq (m/c)^2$ . Now observe from the definition (7.1) of  $E_n(\epsilon/2, \delta)$  that for every  $n \in \mathbb{N}$  such that  $\epsilon/2 > \delta n^{1/2} (\frac{3M_0 + \Lambda}{\sigma})$ , we have  $E_n(\epsilon/2, \delta) = \emptyset$ , where  $M_0 = \max_{s \in S} \log \|\rho(s)\|$ . Therefore, up to reducing  $\delta_1 > 0$  to  $\delta_0 > 0$  so that any  $n \leq (m/c)^2$  satisfies  $\epsilon/2 > \delta_0 n^{1/2} (\frac{3M_0 + \Lambda}{\sigma})$ , we get that for every  $\delta \in (0, \delta_0)$ ,  $n \in \mathbb{N}$ , we have  $\widehat{\nu}(E_n(\epsilon/2, \delta)) \leq \eta$ , proving that the laws of  $S_n$  constitute a tight family.

We now turn to proving that the finite-dimensional distributions of  $(S_n)$  converge to those of the Wiener measure. Fix  $0 = t_0 < t_1 < t_2 < \dots < t_d \leq 1$ . Let  $F_{n,t_1,\dots,t_d}(x)$  for  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  denote the cumulative distribution function

$$F_{n,t_1,\dots,t_d}(x) = \nu \left( \xi \in \partial\Gamma : S_n \xi(t_1, \dots, t_d) \in \prod_{i=1}^d (-\infty, x_i] \right),$$

where

$$S_n \xi(t_1, \dots, t_d) := (S_n \xi(t_1), S_n \xi(t_2) - S_n \xi(t_1), \dots, S_n \xi(t_d) - S_n \xi(t_{d-1})).$$

We would like to prove that  $F_{n,t_1,\dots,t_d}(x)$  converges as  $n \rightarrow \infty$  to the cumulative distribution function  $F_{t_1,\dots,t_d}(x)$  of the multidimensional normal distribution  $N(0, \omega)$  with  $d \times d$  diagonal covariance matrix  $\omega$  with entries  $\omega_{ii} = t_i - t_{i-1}$ , [9, section 1]. Recall that the Patterson–Sullivan measure  $\nu$  on  $\partial\Gamma$  is given by  $\Psi_* \widehat{\nu}$ , where  $\Psi : Y_\infty \rightarrow \partial\Gamma$  is continuous, surjective and finite-to-one. Moreover, using the fact that any two geodesic ray representing  $\xi$  stays at bounded distance depending only on the hyperbolicity constant, it follows that there exists a sequence  $\eta_n$  converging to zero as  $n \rightarrow \infty$  such that for every  $\xi \in \partial\Gamma$ , for any  $y \in \Psi^{-1}(\xi)$ , we have  $\|S_n \xi(t_1, \dots, t_d) - \widehat{S}_n y(t_1, \dots, t_d)\|_\infty \leq \eta_n$ . As a consequence, it suffices to show that for every  $x \in \mathbb{R}^d$

$$\widehat{F}_{n,t_1,\dots,t_d}(x) = \widehat{\nu} \left( y \in Y : \widehat{S}_n y(t_1, \dots, t_d) \in \prod_{i=1}^{d-1} (-\infty, x_i] \right)$$

converges to  $F_{t_1,\dots,t_d}(x)$  as  $n \rightarrow \infty$ , where  $\widehat{S}_n y(t_1, \dots, t_d)$  is defined analogously to  $S_n \xi(t_1, \dots, t_d)$ .

We define  $E_{n,t_1,\dots,t_d}(x) = \{y \in \cup_j \Sigma_{B_j} : \widehat{S}_n y(t_1, \dots, t_d) \in \prod_{i=1}^d (-\infty, x_i]\} \subset \Sigma_A$ . Recall from (3.2) that the Cesaró averages of  $\widehat{\nu}$  under the shift map converges to the Parry-like measure  $\mu$  in the total variation distance. It follows from (3.2) and Corollary 5.6 (statement corresponding to Proposition 5.3) that for every  $x \in \mathbb{R}^d$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^{1/4}} \sum_{k=0}^{n^{1/4}} \sigma_*^k \widehat{\nu}(E_{n,t_1,\dots,t_d}(x)) &= \lim_{n \rightarrow \infty} \mu \left( y \in \Sigma_A : S_n y(t_1, \dots, t_d) \in \prod_{i=1}^d (-\infty, x_i] \right) \\ &= F_{t_1,\dots,t_d}(x). \end{aligned}$$

Defining

$$E_{n,t_1,\dots,t_d}^\pm(x) = E_{n,t_1,\dots,t_d}(x \pm Cn^{-1/4}(1, 1, \dots, 1)),$$

where  $C > 0$  is some positive constant, we see that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1/4}} \sum_{k=0}^{n^{1/4}} \sigma_*^k \widehat{\nu}(C_n^\star E_{n,t_1,\dots,t_d}(x)) = F_{t_1,\dots,t_d}(x), \quad (7.4)$$

for each  $x \in \mathbb{R}^d$  and  $\star \in \{+, -\}$ . Similarly to (6.5) if  $C > 0$  is taken sufficiently large (depending only on  $\max_{s \in S} \log \|\rho(s)\|$  and the variance  $\sigma^2 > 0$ ), by inclusion of the corresponding sets, we have

$$\sigma_*^k \widehat{\nu}(C_n^- E_{n,t_1,\dots,t_d}(x)) \leq \widetilde{\nu}_k(E_{n,t_1,\dots,t_d}(x)) \leq \sigma_*^k \widehat{\nu}(C_n^+ E_{n,t_1,\dots,t_d}(x)) \quad (7.5)$$

for all integers  $n \geq 1$  and  $n^{1/4} \geq k \geq 1$ . We deduce from (7.4) and (7.5) that

$$\frac{1}{n^{1/4}} \sum_{k=0}^{n^{1/4}} \widetilde{\nu}_k(E_{n,t_1,\dots,t_d}(x)) = F_{t_1,\dots,t_d}(x).$$

Finally, by Lemma 6.1, this implies that  $\widehat{F}_{n,t_1,\dots,t_d}(x)$  also converges to  $F_{t_1,\dots,t_d}(x)$  as  $n \rightarrow \infty$ . From our above discussion, this concludes the proof of 1.

(2) We need to show that the set  $U$  of  $\xi \in \partial\Gamma$  such that the conclusion of 2. holds has full  $\nu$ -measure. To this end, let  $E$  be the set of  $y \in Y_\infty$  such that the conclusion holds when  $S_n \xi$  is replaced by  $\widehat{S}_n y$  and  $B$  be the set of  $x \in \cup_j \Sigma_{B_j}$  such that the same conclusion again holds with  $\widehat{S}_n x$ . Note that the set  $U$  is well-defined because its defining property does not depend on the choice of the representing geodesic ray and all these sets are Borel measurable. As, given  $\xi \in U$ , we have that any  $\xi'$  with the property  $\xi_m = \xi'_{m+k}$  for certain  $k \in \mathbb{Z}$  and every  $m \in \mathbb{N}$  large enough also belongs to  $U$ , the set  $U$  is  $\Gamma$ -invariant. By  $\Gamma$ -ergodicity of  $\nu$ , all we need to show is  $\nu(U) > 0$ . As  $E \subset \Psi^{-1}(U)$  and  $\Psi_* \widehat{\nu} = \nu$ , it suffices to show that  $\widehat{\nu}(E) > 0$ . Let, as before,  $V_{\max}$  denote the set of vertices belonging to a maximal component and  $v \in V_{\max}$ . Let  $k \in \mathbb{N}$  be such that there exists a path of length  $k$  from  $*$  to  $v$ . By Lemma 3.1, there exists  $\alpha_v^k > 0$  such that

$$\sigma_*^k \widehat{\nu}|_{[v]} = \alpha_v^k \mu|_{[v]}. \quad (7.6)$$



Now thanks to Corollary 5.6 (statement corresponding to Proposition 5.4), the set  $B$  has full  $\mu$  measure. Therefore, by (7.6), we have  $\sigma_*^k \hat{\nu}(B) = \hat{\nu}(\sigma^{-k}(B)) > 0$ . But as for any  $k \in \mathbb{N}$ , we have  $\sigma^{-k}(B) \cap Y_\infty \subseteq E$ , we obtain  $\hat{\nu}(E) > 0$ , as desired.  $\square$

As an immediate consequence, we record the following more classical results, namely the CLT and LIL. The latter one provides a refinement of Theorem 1.10 in the current setting.

**Corollary 7.2** (Boundary CLT and LIL). *Let  $\rho : \Gamma \rightarrow \text{GL}_d(\mathbb{R})$  be a strongly irreducible proximal representation of a hyperbolic group  $\Gamma$ . Equip  $\Gamma$  with a finite generating set  $S$  and let  $\nu$  be the Patterson–Sullivan measure defined in (1.2). Let  $\Lambda, \sigma^2 > 0$  be the mean and variance from Theorem 1.12. Then,*

(1) *for each  $n \geq 1$  and  $x \in \mathbb{R}$ , denoting*

$$A_n(x) := \left\{ \xi \in \partial\Gamma : \text{for any representative } \xi_m \rightarrow \xi \text{ with } \xi_0 = o, \frac{\log \|\rho(\xi_n)\| - \Lambda n}{\sqrt{n}} \leq x \right\}$$

*we have*

$$\lim_{n \rightarrow \infty} \nu(A_n(x)) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^x e^{-t^2/2\sigma^2} dt ; \text{ and} \tag{7.7}$$

(2) *for  $\nu$  almost every  $\xi \in \partial\Gamma$  and for any representative  $\xi_m \rightarrow \xi$ ,*

$$\liminf_{n \rightarrow \infty} \frac{\log \|\rho(\xi_n)\| - n\Lambda}{\sqrt{2\sigma n \log \log n}} = -1 \text{ and } \limsup_{n \rightarrow \infty} \frac{\log \|\rho(\xi_n)\| - n\Lambda}{\sqrt{2\sigma n \log \log n}} = 1. \quad \square$$

*Remark 7.3* (Speed and uniformity in boundary CLT). Using Theorem 4.8, it is possible to give a speed estimate in (7.7) that is uniform over  $x \in \mathbb{R}$ . However, we will not pursue this direction as this would be a (somewhat lengthy and technical) diversion from the main goals in the article (see [22, sections 4 and 5]).

## 8 | COUNTING CLT AND ERROR TERMS

In this section, after briefly commenting on our approach, we prove Theorem 1.9.

Similar to the schemes we followed in the proofs of weak law of large numbers (Section 3) and large deviation results for counting (Section 6), one could try to obtain a corresponding counting CLT (with or without error term) directly from Corollary 7.2(1). However there are difficulties in implementing this approach for the CLT. The main issue stems from the fact that when we compare the asymptotic density of sets with the Patterson–Sullivan measure of certain boundary sets (e.g., as in the proof of Theorem 1.5), we do so up to a bounded multiplicative constant. Such a constant is inconsequential when proving large deviation type results, however it destroys the precise limiting behavior that we need for a CLT to hold. To overcome this issue (and, importantly, to prove a CLT with the Berry–Esseen type error term) we will directly compare the uniform counting measures on  $S_n$  with the Markov measures on  $\Sigma$ . This method will make use of a quantified version of an argument from a recent work of Gehlman–Taylor–Tiozzo [34].

More precisely, using the strongly Markov structure, up to a periodicity issue, we will consider a geodesic factorization of an element  $g \in \Gamma$  chosen uniformly from the sphere of length  $n$  as  $g_0 g_1 g_2$  with  $g_0$  and  $g_2$  of size approximately  $\log n$ . It will then suffice to show a CLT with error term for the middle factor  $g_1$ . We will show (Lemma 8.5) that the distribution of this middle factor  $g_1$  is approximated (with speed) by the length  $\sim (n - 2 \log n)$  path distribution of a Markov chain. We can then associate a Markovian random matrix product to this chain and bring back (Lemma 8.2) the relevant result (Theorem 4.8) from the Markovian matrix products to establish a counting CLT with error term (Proposition 8.8). The proof is then completed by resolving the periodicity issue.

We now start collecting the necessary ingredients for the proof of Theorem 1.9. We will heavily use the constructions from Section 2. Fix a nonelementary Gromov-hyperbolic group  $\Gamma$ , a generating set  $S \subset \Gamma$  and strongly Markov structure  $\mathcal{G}$ . Let  $A$  be the transition matrix as introduced in Subsection 2.2.2. Let  $p$  be an integer that is divisible by the periods of each maximal component of  $A$  so that the nonnegative matrix  $A^p$  has a unique (necessarily real) eigenvalue  $\lambda^p$  of maximal modulus. To deduce Theorem 1.9 we will first study the convergence of our counting distributions along the subsequence  $np$ .

For a positive  $k \in \mathbb{N}$ , we will define a Markov chain on the state space

$$\Omega_{kp} := \{(w_0, \dots, w_{kp}) : w_i \in \mathcal{G}, A_{w_i, w_{i+1}} = 1\}$$

of length- $kp$  paths in the strongly Markov structure  $\mathcal{G}$ . To define a transition kernel and a stationary measure on  $\Omega_{kp}$ , we let

$$p_i = \lim_{n \rightarrow \infty} \frac{e_i^T A^{np} \mathbf{1}}{\lambda^{np}} \quad \text{and} \quad u_i = \lim_{n \rightarrow \infty} \frac{e_*^T A^{np} e_i}{\lambda^{np}}, \tag{8.1}$$

where  $e_i$  and  $e_*$  correspond to the vectors that have the entry 1 in the index corresponding to the vertices  $v_i$  and  $*$ , respectively, and 0 elsewhere.

*Remark 8.1.* Before proceeding further, we remark that, by our choice of  $p$ , the limits above defining each  $p_i$  and  $u_i$  converge exponentially quickly. This is because the matrix  $A^p$  exhibits a spectral gap from its leading (real positive) eigenvalue to the rest of the spectrum.

Now let  $\pi_{kp}$  be the measure on  $\Omega_{kp}$ , defined by

$$\pi_{kp}(w_0, \dots, w_{kp}) = \frac{u_{w_0} p_{w_{kp}}}{\lambda^{kp} p_*}.$$

It is readily checked that  $\pi_{kp}$  defines a probability measure. Let  $N_{kp}$  be the transition kernel defined by

$$N_{kp}((w_0, \dots, w_{kp}), (w'_0, \dots, w'_{kp})) = \begin{cases} \frac{p_{w'_{kp}}}{\lambda^{kp} p_{w_0}} & \text{if } p_{w_0} > 0 \text{ and } w_{kp} = w'_0 \\ 0 & \text{otherwise.} \end{cases}$$

Unfolding the definitions, one also readily checks that  $N_{kp}$  is a stochastic matrix and  $\pi_{kp}$  is  $N_{kp}$ -stationary (i.e., a left eigenvector with eigenvalue one). Let  $\tilde{\Omega}_{kp} \subseteq \Omega_{kp}^{\mathbb{N}}$  be the subshift associated

to this Markov chain and  $\tilde{\mathbb{P}}_{kp}$  be the associated Markovian measure on  $\tilde{\Omega}_{kp}$ . Finally, for  $k \geq 1$  and  $z_1, \dots, z_k \in \Omega_p$ , let  $[z_1, \dots, z_k]$  be the associated cylinder set in  $\tilde{\Omega}_p$  and  $(z_1, \dots, z_k)$  be the corresponding element of  $\Omega_{kp}$ . Observe that by an easy calculation using the definitions of  $\pi_{kp}$ 's and  $N_p$ , we have

$$\pi_{kp}((z_1, \dots, z_k)) = \tilde{\mathbb{P}}_p([z_1, \dots, z_k]). \tag{8.2}$$

The nonnegative matrix  $A^p$  is not necessarily irreducible and hence we decompose it into connected components (as we did to obtain  $A''$  from  $A$  in Subsection 2.2.2). Some of these components will have spectrum with simple eigenvalue  $\lambda^p$ . We label these finitely many  $A^p$  maximal components  $C_1, \dots, C_{m_0}$ . Note that each of the vertex sets for  $C_1, \dots, C_{m_0}$  are subsets of the vertex sets of the maximal components of  $A$ . Notice from definitions of the constants  $p_i$  and  $u_i$ 's in (8.1) and that of the stationary measure  $\pi_{kp}$  that  $\pi_{kp}(w_0, \dots, w_{kp}) > 0$  if any only of  $w_0$  and  $w_{kp}$  belong to the same maximal component of  $A^p$ . Moreover, the transition kernel  $N_{kp}$  sends a path  $(w_0, \dots, w_{kp})$  in a maximal component  $C_i$  to a path in  $C_i$ . Therefore, the Markov chain defined above is not ergodic if  $m_0 \geq 2$ . Its ergodic components are simply given by the maximal components  $C_i$  for  $i = 1, \dots, m_0$ : the restriction of the transition kernel  $N_{kp}$  to the set  $\Omega_{kp}^i$  paths of length  $kp$  with initial and end vertex belonging to a single  $C_i$  (together with the normalized restriction of  $\pi_{kp}$  to  $\Omega_{kp}^i$ ) gives an ergodic Markov chain. Moreover, by the choice of  $p$  (a common multiple of the periods of maximal component of  $A$ ), these Markov chains are primitive.

We now proceed precisely as in Section 5 to deduce a CLT with Berry–Esseen bounds along periodic products from  $\Sigma_A$ . As the procedure is the same, we only outline the steps.

- (i) As in Subsection 5.1.1, we associate a Markovian random matrix product  $(M_n^i)$  to the primitive finite state Markov chains on  $\Omega_p^i$ .
- (ii) As in Lemma 5.1, we check the 1-contracting and strong irreducibility assumptions for these Markovian products.
- (iii) By applying Theorem 4.8, we deduce a CLT with mean  $\Lambda_i$  and variance  $\sigma_i^2 > 0$  and with Berry–Esseen error term of order  $O(\frac{\log n}{\sqrt{n}})$ .
- (iv) We check exactly as in Proposition 5.5 that the means  $\Lambda_i$  and variances  $\sigma_i^2$  do not depend on  $i = 1, \dots, m_0$ . Set  $\Lambda = \Lambda_i$  and  $\sigma^2 = \sigma_i^2$ .

From these, analogous to Corollary 5.6, we deduce the following.

**Lemma 8.2.** *There exists a constant  $D > 0$  such that for every  $n \geq 1$  and  $t \in \mathbb{R}$*

$$\left| \tilde{\mathbb{P}}_p \left( (z_1, \dots) \in \tilde{\Omega}_p : \frac{\log \|\rho(\lambda(z_1)) \dots \rho(\lambda(z_n))\| - np\Lambda}{\sqrt{np}} \leq t \right) - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2\sigma^2}} ds \right|$$

is bounded above by  $\frac{D \log n}{\sqrt{n}}$  where for  $z = (w_0, \dots, w_p)$  belonging to  $\Omega_p$ , we write  $\lambda(z) = \lambda(w_0, w_1) \dots \lambda(w_{p-1}, w_p)$ .

*Remark 8.3.* Keeping the notation of the previous lemma, notice that in view of (8.2), the first term in the previous lemma is equal to

$$\pi_{np} \left\{ (z_1, \dots, z_n) : \frac{\log \|\rho(\lambda(z_1)) \dots \rho(\lambda(z_n))\| - np\Lambda}{\sqrt{np}} \leq t \right\}.$$

Having obtained Lemma 8.2, to prove Theorem 1.9, we now follow the ideas used in [34, sections 6–7]. However, we need to quantify various rates of convergence to obtain the error term in Theorem 1.9.

We start by defining a probability measure  $\mu_q$  on each sphere  $S_q$  that will help us deal with the periodicity issue at the end. Fix an integer  $0 \leq r \leq p - 1$ . We define a measure on the set  $S_r$  (or equivalently the set of paths of length  $r$  in  $\mathcal{G}$  starting from the vertex  $*$ ) in the following way. For  $g \in S_r$ , we set

$$\mu_r(g) = \frac{e_i^T A_\infty \mathbf{1}}{e_*^T A^r A_\infty \mathbf{1}} = \lim_{n \rightarrow \infty} \frac{e_i^T A^{np} \mathbf{1}}{e_*^T A^r A^{np} \mathbf{1}},$$

where  $A_\infty = \lim_{n \rightarrow \infty} A^{np} / \lambda^{np}$  and  $i$  is the end vertex of the path in  $\mathcal{G}$  starting at  $v_*$  corresponding to  $g$ . Here the limit defining  $A_\infty$  exists because by choice of  $p \in \mathbb{N}$  so that  $\lambda^p$  is the unique eigenvalue of maximal modulus of  $A^p$ . For the same reason, the limit defining  $\mu_r(g)$  converges exponentially quickly. One easily checks that  $\sum_{|g|=r} \mu_r(g) = 1$ . We extend the definition of  $\mu_q$  on  $S_q$  for  $q \geq p$  as follows: given integers  $n \geq 1$  and  $q = np + r$  with  $0 \leq r \leq p - 1$ , we define  $\mu_{np+r}$  on  $S_{np+r}$  (equivalently, on the set of paths of length  $np + r$  in  $\mathcal{G}$  starting from the vertex  $*$ ) as follows: given  $g \in S_{np+r}$ , let  $(*, w_1, \dots, w_{np+r})$  be the unique path in  $\mathcal{G}$  such that  $g = \lambda(*, w_1) \dots \lambda(w_{np+r-1}, w_{np+r})$ . Set  $h = \lambda(*, w_1) \dots \lambda(w_{r-1}, w_r)$  and let  $\mu_{np+r}(g) = \mu_r(h) \frac{1}{e_{w_r}^T A^{np} \mathbf{1}}$ . Note that the denominator in the last expressions is the number of length  $np$ -paths starting at the vertex  $w_r$ . Let  $\tau_n$  denote the uniform probability measure on the sphere  $S_n$ . We have the following

**Lemma 8.4.** *For each  $r = 0, \dots, p - 1$  we have that  $\|\tau_{np+r} - \mu_{np+r}\|_{TV} = O(\theta^n)$  for some  $0 < \theta < 1$  as  $n \rightarrow \infty$ .*

*Proof.* Take a set  $R \subset \Gamma$ , let  $R_{np+r} = R \cap S_{np+r}$  and write  $R_{np+r}^+ = R_{np+r} \cap \bigcup_{g \in S_r^+} [g]$  where  $[g]$  denotes all group elements that have corresponding path in  $\mathcal{G}$  that start with  $g$  and  $S_r^+ = S_r \cap \{g \in S_r : \mu_r(g) > 0\}$ . From the definition of  $\mu_{np+r}$  we see that if  $\mu_{np+r}(R) = 0$  then  $R_{np+r}^+ = \emptyset$  and  $\tau_{np+r}(R)$  decays to 0 exponentially quickly, independently of  $R$ . Otherwise,  $\mu_{np+r}(R) \neq 0$  and denoting by  $v_g$  the last vertex in  $\mathcal{G}$  of the path from  $*$  corresponding to  $g$ , we have

$$\begin{aligned} \mu_{np+r}(R) &= \sum_{g \in S_r^+} \frac{\mu_r(g) \#(R_{np+r} \cap [g])}{e_{v_g}^T A^{np} \mathbf{1}} \\ &= \frac{1}{e_*^T A^{np+r} \mathbf{1}} \left( \sum_{g \in S_r^+} \frac{\mu_r(g) \#(R_{np+r} \cap [g])}{e_{v_g}^T A^{np} \mathbf{1} / e_*^T A^{np+r} \mathbf{1}} \right) \\ &= \left( \frac{1}{e_*^T A^{np+r} \mathbf{1}} \sum_{g \in S_r^+} \#(R_{np+r} \cap [g]) \right) + O(\theta^n) \\ &= \tau_{np+r}(R_{np+r}^+) + O(\theta^n) \end{aligned}$$

for some  $0 < \theta < 1$  independent of  $R$ . In the penultimate line, we used the fact that  $e_{v_g}^T A^{np} \mathbf{1} / e_*^T A^{np+r} \mathbf{1}$  converges to  $\mu_r(g)$  exponentially quickly as  $n \rightarrow \infty$ . To conclude the proof

we note that, from the construction of  $\mu_r$ ,  $|\tau_{np+r}(R) - \tau_{np+r}(R_{np+r}^+)|$  converges to 0 exponentially quickly and that this rate of convergence is independent of  $R$ .  $\square$

We now, following [34], define probability measures that will determine the law of the middle factor  $g_1$  of a  $n$ -long product  $g$  written in geodesic factorization  $g_0g_1g_2$  where  $g_0$  and  $g_2$  are of logarithmic length. Consequently, we show that these measures can be approximated by the path distribution of a Markov chain. Let  $c > 0$  be a positive constant. For a path  $\gamma$  in  $\mathcal{G}$  of length  $np - 2p\lfloor c \log n \rfloor$  starting at  $v_i$  ending at  $v_j$ , we set

$$\tilde{\tau}_{np}^c(\gamma) = \frac{e_*^T A^{p\lfloor c \log n \rfloor} e_i e_j^T A^{p\lfloor c \log n \rfloor} 1}{e_*^T A^{np} 1}.$$

Intuitively  $\tilde{\tau}_{np}^c$  assigns a path  $\gamma$  probability  $s$  if the proportion of length  $np$  paths starting at  $*$  that have  $\gamma$  as a sub-path from the  $p\lfloor c \log n \rfloor$  to the  $np - p\lfloor c \log n \rfloor$  vertex is  $s$ .

**Lemma 8.5.** *For every fixed  $c > 0$  sufficiently large, we have*

$$\|\pi_{pn-2p\lfloor c \log n \rfloor} - \tilde{\tau}_{np}^c\|_{TV} = O\left(\frac{1}{\sqrt{n}}\right)$$

as  $n \rightarrow \infty$ .

*Proof.* By Remark 8.1, there exists  $\delta > 0$  such that for every vertex  $v_i$  in  $\mathcal{G}$  we have

$$p_i = \frac{e_i^T A^{np} 1}{\lambda^{np}} + O(\lambda^{-\delta n}) \quad \text{and} \quad u_i = \frac{e_*^T A^{np} e_i}{\lambda^{np}} + O(\lambda^{-\delta n})$$

as  $n \rightarrow \infty$ . It follows that for any  $c > (2\delta \log \lambda)^{-1}$ , we have

$$p_i = \frac{e_i^T A^{\lfloor c \log n \rfloor p} 1}{\lambda^{\lfloor c \log n \rfloor p}} + O(n^{-1/2}) \quad \text{and} \quad u_i = \frac{e_*^T A^{\lfloor c \log n \rfloor p} e_i}{\lambda^{\lfloor c \log n \rfloor p}} + O(n^{-1/2}) \tag{8.3}$$

as  $n \rightarrow \infty$ . Fix such a constant  $c > 0$ . Let  $v_i$  and  $v_j$  be two vertices in  $\mathcal{G}$  that belong to the same maximal component of  $A^p$  and such that  $u_i > 0$ . Let  $\gamma$  be a path of length  $n' = np - 2p\lfloor c \log n \rfloor$  from  $v_i$  to  $v_j$ . Then,

$$\begin{aligned} \frac{\pi_{n'}(\gamma)}{\tilde{\tau}_{np}^c(\gamma)} &= \frac{u_i p_j}{\lambda^{n'} p_*} \frac{e_*^T A^{np} 1}{e_*^T A^{p\lfloor c \log n \rfloor} e_i e_j^T A^{p\lfloor c \log n \rfloor} 1} \\ &= \frac{u_i p_j}{p_*} \frac{\lambda^{p\lfloor c \log n \rfloor}}{e_*^T A^{p\lfloor c \log n \rfloor} e_i e_j^T A^{p\lfloor c \log n \rfloor} 1} \frac{e_*^T A^{np} 1}{\lambda^{np}}. \end{aligned}$$

Now by the estimates (8.3) and Remark 8.1, we see that this quotient is equal to

$$\frac{u_i p_j}{p_*} \cdot \left(\frac{1}{u_i} + O(n^{-1/2})\right) \cdot \left(\frac{1}{p_j} + O(n^{-1/2})\right) \cdot (p_* + O(\theta^n)) = 1 + O(n^{-1/2}) \tag{8.4}$$

for some  $0 < \theta < 1$ . Here we have used that  $u_i > 0$  and  $p_j > 0$  (the former is assumed, the latter follows because  $v_j$  is assumed to belong to maximal component of  $A^p$ ). As there are only finitely many vertices in  $\mathcal{G}$  and the left-hand side of (8.4) only depends on vertices of  $\mathcal{G}$ , we deduce that

$$\sup_{\gamma} \left| \frac{\pi_{n'}(\gamma)}{\tilde{\tau}_{np}^c(\gamma)} - 1 \right| = O(n^{-1/2}) \quad (8.5)$$

where the supremum is taken over all paths of length  $n'$  that lie entirely in a single  $A^p$  maximal component. Now note that, given arbitrary two vertices  $v_i$  and  $v_j$ , if  $S_{n'}^{ij}$  denotes the set of paths of length  $n'$  from  $v_i$  to  $v_j$  then

$$\begin{aligned} \tilde{\tau}_{np}^c(S_{n'}^{ij}) &= \frac{(e_i^T A^{n'} e_j)(e_*^T A^{p \lfloor c \log n \rfloor} e_i)(e_j^T A^{p \lfloor c \log n \rfloor} \mathbf{1})}{e_*^T A^{np} \mathbf{1}} \\ &\leq \frac{(e_i^T A^{n'} \mathbf{1})(e_*^T A^{p \lfloor c \log n \rfloor} e_i)(e_j^T A^{p \lfloor c \log n \rfloor} \mathbf{1})}{e_*^T A^{np} \mathbf{1}} = p_i u_i p_j / p_* + O(n^{-1/2}), \end{aligned} \quad (8.6)$$

as  $n \rightarrow \infty$ , where the last equality follows (as in (8.4)) by our estimates (8.3) and Remark 8.1. The limit  $p_i u_i p_j / p_*$  in (8.6) is equal to 0 unless both  $v_i, v_j$  belong to the same  $A^p$  maximal component and  $u_i > 0$ . Letting  $L_n$  denote all paths of length  $n'$  that lie entirely in an  $A^p$  maximal component and start at any vertex  $v_i$  with  $u_i > 0$ . We have, for any set  $R$  consisting of length  $n'$  paths,

$$|\pi_{n'}(R) - \tilde{\tau}_{np}^c(R)| \leq \sum_{\gamma \in R \cap L_n} |\pi_{n'}(\gamma) - \tilde{\tau}_{np}^c(\gamma)| + \tilde{\tau}_{np}^c(R \setminus L_n).$$

Here we have used that  $\pi_{n'}(R \setminus L_n) = 0$  which we can see holds from the definition of  $\pi_{n'}$ . To conclude the proof we note that, from (8.6),  $\tilde{\tau}_{np}^c(R \setminus L_n) = O(n^{-1/2})$  and

$$\begin{aligned} \sum_{\gamma \in R \cap L_n} |\pi_{n'}(\gamma) - \tilde{\tau}_{np}^c(\gamma)| &= \sum_{\gamma \in R \cap L_n} \left| \frac{\pi_{n'}(\gamma)}{\tilde{\tau}_{np}^c(\gamma)} \tilde{\tau}_{np}^c(\gamma) - \tilde{\tau}_{np}^c(\gamma) \right| \\ &\leq \sup_{\gamma \in R \cap L_n} \left| \frac{\pi_{n'}(\gamma)}{\tilde{\tau}_{np}^c(\gamma)} - 1 \right| = O(n^{-1/2}), \end{aligned}$$

where we used (8.5) in the last equality and implied error term constants are independent of the  $R$ . This completes the proof.  $\square$

*Remark 8.6.* So far this section has been concerned with comparing the measures  $\pi_{np}$ ,  $\tau_{np}$  and  $\tilde{\tau}_{np}^c$ . Each of these measures are constructed with the  $*$  vertex as their “base point,” that is,  $\pi_{np}$  is constructed using the  $e_*$  vector and  $\tau_{np}$ ,  $\tilde{\tau}_{np}^c$  can be seen as counting measures on the paths in  $\mathcal{G}$  starting at  $*$  (as indicated in their constructions). If we replace the  $*$  vertex with any other vertex  $v_0$  of large growth, that is a vertex  $v_0$  such that  $e_{v_0}^T A^n \mathbf{1} \geq C \lambda^n$  for some  $C > 0$  and all  $n \geq 1$ , then we can construct measures analogous to  $\pi_{np}$ ,  $\tau_{np}$  and  $\tilde{\tau}_{np}^c$  but with  $v_0$  being the new “base point”. To do this, one replaces  $e_*$  with  $e_{v_0}$  in the construction of  $\pi_{np}$  and alters  $\tau_{np}$  and  $\tilde{\tau}_{np}^c$  so that they count with respect to paths starting at  $v_0$  instead of  $*$ . This new construction will yield

different measures however all of the results that we have seen so far in this section will also hold for these measures.

We can now prove a counting CLT with error term for the sequence of spheres  $(S_{np})_{n \in \mathbb{N}}$ .

**Proposition 8.7.** *There exists  $\Lambda, \sigma^2 > 0$  such that*

$$\tau_{np} \left( g \in \Gamma : \frac{\log \|\rho(g)\| - \Lambda|g|}{\sqrt{|g|}} \leq t \right) = N(t, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right)$$

as  $n \rightarrow \infty$ .

*Proof.* Let  $\Omega$  denote the set of finite paths in  $\mathcal{G}$  and for  $g \in \Omega$  let  $\bar{g}$  denote the group element corresponding to  $g$  via the labeling map. For  $t \in \mathbb{R}$ , let  $E(t)$  and  $\hat{E}(t)$  be the sets

$$\left\{ g \in \Gamma : \frac{\log \|\rho(g)\| - \Lambda|g|}{\sqrt{|g|}} \leq t \right\} \text{ and } \left\{ g \in \Omega : \frac{\log \|\rho(\bar{g})\| - \Lambda|\bar{g}|}{\sqrt{|\bar{g}|}} \leq t \right\},$$

respectively, and let  $c > 0$  be a constant given by Lemma 8.5. For each  $n \in \mathbb{N}$ , we factorize each path (or element)  $g$  of length  $np$  as a concatenation (resp., product)  $g_0g_1g_2$  where  $g_0, g_1$  and  $g_2$  are the sub-paths (resp., factors) of  $g$  of length  $p\lfloor c \log n \rfloor, np - 2p\lfloor c \log n \rfloor$  and  $p\lfloor c \log n \rfloor$ , respectively. Writing  $\tau_{np}(E(t)) = \tau_{np}(g = g_0g_1g_2 \in E(t))$  and using submultiplicativity of the matrix norm  $\|\cdot\|$  we deduce that there exists  $C > 0$  such that  $\tau_{np}(E(t))$  is bounded above and below by

$$\tau_{np}(g = g_0g_1g_2 : g_1 \in E(t + Cn^{-1/2} \log n)) \text{ and } \tau_{np}(g = g_0g_1g_2 : g_1 \in E(t - Cn^{-1/2} \log n)),$$

respectively. Now note that by the definition of  $\tilde{\tau}_{np}^c$  and by Lemma 8.5

$$\begin{aligned} \tau_{np}\{g = g_0g_1g_2 : g_1 \in E(t \pm Cn^{-1/2} \log n)\} &= \tilde{\tau}_{np}^c(\hat{E}(t \pm Cn^{-1/2} \log n)) \\ &= \pi_{pn-2p\lfloor c \log n \rfloor}(\hat{E}(t \pm Cn^{-1/2} \log n)) + O(n^{-1/2}). \end{aligned}$$

On the other hand, by Lemma 8.2 and Remark 8.3, we get that

$$\begin{aligned} \pi_{pn-2p\lfloor c \log n \rfloor}(\hat{E}(t \pm Cn^{-1/2} \log n)) &= N(t \pm Cn^{-1/2} \log n, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right) \\ &= N(t, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right) \end{aligned}$$

as  $n \rightarrow \infty$  uniformly in  $t \in \mathbb{R}$ . The last line follows from the fact that the normal distribution has uniformly bounded derivative. The proof is completed by combining the last two displayed equations. □



Using the same ideas, we can also prove the following. Given a vertex  $v$  in  $\mathcal{G}$  recall that we say that  $v$  is of large growth if the number of length  $n$  paths in  $\mathcal{G}$  starting at  $v$  grows at least like  $C\lambda^n$  for some  $C > 0$ , that is,  $e_v^T A^n 1 \geq C\lambda^n$ .

**Proposition 8.8.** *Suppose  $v$  is a vertex of large growth. Suppose  $\tau_{np}^v$  is the uniform counting measure on the paths in  $\mathcal{G}$  of length  $np$  starting at  $v$ . Let  $\Lambda, \sigma^2 > 0$  be the constants in Proposition 8.7. Then*

$$\tau_{np}^v \left( g \in \Omega : \frac{\log \|\rho(\bar{g})\| - \Lambda|\bar{g}|}{\sqrt{|\bar{g}|}} \leq x \right) = N(x, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right)$$

as  $n \rightarrow \infty$  where  $\Omega$  represents the set of finite paths in  $\mathcal{G}$  and for  $g \in \Omega$ ,  $\bar{g} \in \Gamma$  is the group element corresponding to multiplying the edge labelings in  $g$ .

*Proof.* When  $v = *$  this proposition is precisely Proposition 8.7. The proof of this more general result follows the same method used to prove Proposition 8.7 but we consider the “initial vertex” to be  $v$  instead of  $*$ . We define the counting measures  $\tilde{\tau}_{np}^v$  and  $\pi_{kp}$  as before, but we replace the vector  $e_*$  with the vector  $e_v$  in their definitions, see Remark 8.6. We can then prove analogous results, such as Lemma 8.5 for these measures and then carry out the same proof.  $\square$

Finally, we are in a position to prove our CLT.

*Proof of Theorem 1.9.* For  $t \in \mathbb{R}$ , let  $E(t)$  denote the set defined in Proposition 8.7. Fix  $r \in \{0, \dots, p - 1\}$ . For each  $g_0 \in \Gamma$  with  $|g_0|_S = r$  let  $t(g_0)$  denote the terminal vertex in the path corresponding to  $g_0$  that begins with  $*$  in  $\mathcal{G}$ . It follows from the definition of  $\mu_r$  that if  $t(g_0)$  is not a vertex of large growth then  $\mu_r(g_0) = 0$ . Then, by definition of the measure  $\mu$  and Proposition 8.8

$$\begin{aligned} \mu_{np+r}(E(t)) &= \sum_{|g_0|=r} \mu_r(g_0) \tau_{np}^{t(g_0)}(g_1 : g_0 g_1 \in E(t)) \\ &= \sum_{|g_0|=r} \mu_r(g_0) \left( N(t, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right) \right) \\ &= N(t, \sigma) + O\left(\frac{\log n}{\sqrt{n}}\right) \end{aligned}$$

$n \rightarrow \infty$  and where the implied constant is independent of  $t \in \mathbb{R}$ . It then follows from Lemma 8.4 that  $\tau_{np+r}(E(t)) = N(t, \sigma) + O(n^{-1/2} \log n)$ . As this holds for each  $r = 0, \dots, p - 1$ , our theorem follows.  $\square$

## 9 | ON A QUESTION OF KAIMANOVICH–KAPOVICH–SCHUPP

Here we briefly discuss some consequences of our results which pertain to the growth indicator functions, and make a connection between these and a result of Lubotzky–Mozes–Ragunathan [52, 53]. These consequences provide an affirmative answer to a question of

Kaimanovich–Kapovich–Schupp [46] that was also raised in our precise setting in Sert’s thesis [62].

### 9.1 | Unique maximum of growth indicator

We will formulate the consequences using the language of reductive real linear algebraic groups. For definitions of the notions and objects we use, we refer the reader to [8]. The reader is invited to consider the case  $G = \text{SL}_d(\mathbb{R})$  or  $\text{GL}_d(\mathbb{R})$  in which case we will specify the relevant objects.

Let  $G$  be the group of real points of a connected reductive affine algebraic group defined over  $\mathbb{R}$  (we will shortly refer to such a group as a real reductive Lie group). Let  $\mathfrak{a}^+$  be a Weyl chamber in a Cartan subspace of the Lie algebra of  $G$  and  $\vec{\kappa} : G \rightarrow \mathfrak{a}^+$  the associated Cartan projection. For the case of  $G = \text{GL}_d(\mathbb{R})$  or  $\text{SL}_d(\mathbb{R})$ , one can define for  $g \in G$ ,

$$\vec{\kappa}(g) = (\log \sigma_1(g), \dots, \log \sigma_d(g)),$$

where  $\sigma_i(g)$ ’s are the singular values of  $g$  in decreasing order and  $\mathfrak{a}^+$  to be the cone in  $\mathbb{R}^d$  given by  $x_1 \geq \dots \geq x_d$  in the case of  $\text{GL}_d(\mathbb{R})$  and the intersection of this cone with the subspace  $x_1 + \dots + x_d = 0$  in the case of  $\text{SL}_d(\mathbb{R})$ . Denote by  $\mathfrak{a}^{++}$  the interior of  $\mathfrak{a}^+$ . We clarify that these definitions only differ by a linear change of coordinates from the more standard definitions in [8] and the results discussed below are independent of the choice of coordinates up to affine transformations.

A direct corollary of Theorem 1.1 is the following.

**Corollary 9.1.** *Let  $G$  be a real reductive Lie group,  $\vec{\kappa} : G \rightarrow \mathfrak{a}^+$  a Cartan projection of  $G$ . Let  $\Gamma$  be a Gromov-hyperbolic group,  $\rho : \Gamma \rightarrow G$  a representation with Zariski-dense image. For every finite symmetric generating set  $S$  of  $\Gamma$ , there exists  $\vec{\Lambda} \in \mathfrak{a}^{++}$  such that*

$$\frac{1}{n} \sum_{\gamma \in S_n} \frac{1}{\#S_n} \vec{\kappa}(\rho(\gamma)) \xrightarrow{n \rightarrow \infty} \vec{\Lambda}.$$

*Proof.* The convergence is a straightforward consequence of [8, Lemmas 8.15 and 8.17] together with Theorem 1.1. The fact that  $\vec{\Lambda} \in \mathfrak{a}^{++}$  is obtained using additionally Proposition 1.4. The details are standard and omitted. □

It might be possible to prove the above convergence under the same assumptions when we replace the Cartan projection  $\vec{\kappa}$  with the Jordan projection  $\vec{\lambda}$ . However, even in the case of Markovian random matrix products, the law of large numbers for the spectral radius may fail (see [2]) and one has to deal with this difficulty. On the other hand, in ongoing work with Cipriano and Dougall, we show that the above convergence holds for both  $\vec{\kappa}$  and  $\vec{\lambda}$  with a speed estimate under the assumption that the representation  $\rho$  is Anosov (with respect to an appropriate sense parabolic subgroup). Finally, in the previous result, one may prove the stronger statement that  $\vec{\Lambda}$  belongs to the interior of the joint spectrum  $J(S)$  of  $S$  (see [16]). We will however content with the above version for brevity.

We now turn to a consequence of our large deviation estimate Theorem 1.5, its connection to the uniqueness of the maximum of the growth indicator function and the connection between the latter and a question of Kaimanovich–Kapovich–Schupp [46, Problem 9.3]. In the latter, the authors proved (see also an earlier related consideration in [47]) that if in Theorem 1.1, one considers  $\Gamma$  to

be a free group with a free generating set  $S$  and having fixed an automorphism  $\phi : F \rightarrow F$ , one takes  $\varphi : F \rightarrow \mathbb{R}$  to be the function  $w \mapsto |\phi(w)|_S$ , then the convergence in Theorem 1.1 is exponential. In [46, Problem 9.3], the authors ask the question of whether there are other examples where this convergence is exponential for a map on a free group. Theorems 1.5 and 1.7 clearly provide positive answers in a more general (both for underlying groups and generating sets) setting. The existence of this kind of phenomenon was also asked in [62, Introduction 7.4.3] with the language of growth indicator of a finite set, a notion that was introduced therein (see also a forthcoming paper by Sert). We now briefly recall this notion and formulate the consequence of our counting large deviation result.

Let  $G$  be a real reductive Lie group,  $\mathfrak{a}^+$  a Weyl chamber of  $G$ ,  $\Gamma < G$  a finitely generated subgroup and  $S$  be a finite generating set  $\Gamma$ . We define the growth indicator of  $S$  as:

$$\varphi_S : \mathfrak{a}^+ \rightarrow [0, \infty) \cup \{-\infty\}$$

$$\alpha \mapsto \inf_{\alpha \in O} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \# \left\{ g \in S_n \mid \frac{1}{n} \vec{\kappa}(g) \in O \right\},$$

where  $O$  ranges over neighborhoods of  $\alpha$  in the Weyl chamber  $\mathfrak{a}^+$ . If  $\Gamma$  is Zariski-dense, the closure of the locus of points  $x \in \mathfrak{a}^+$  on which  $\varphi_S$  takes values in  $[0, \infty)$  is contained in the joint spectrum of  $S$  ([16]), which is a convex body in  $\mathfrak{a}^+$ . On the other hand, denoting by  $\lambda_S > 1$  the exponential growth rate of the cardinality of  $S_n$ , the function  $\varphi_S$  is bounded above by  $\log \lambda_S$ . Moreover, it is not hard to see that the value  $\log \lambda_S$  is always attained by  $\varphi_S$ . In this general setting, the locus of maxima, that is, the description of the set  $\varphi_S^{-1}(\{\log \lambda_S\})$  remains to be studied. Thanks to our Theorem 1.5, we can describe it in the setting of Corollary 9.1. Indeed, the conclusion of the latter implies that  $\varphi_S(\vec{\Lambda}) = \log \lambda_S$  where  $\vec{\Lambda} \in \mathfrak{a}^{++}$  is given by that corollary and the following consequence of Theorem 1.5 says that  $\varphi_S$  attains its maximum only on  $\vec{\Lambda}$  which is precisely the aforementioned positive answer to [46, Problem 9.3].

**Corollary 9.2.** *Let  $G$  be a real reductive Lie group,  $\vec{\kappa} : G \rightarrow \mathfrak{a}^+$  a Cartan projection,  $\Gamma < G$  a Zariski-dense Gromov-hyperbolic subgroup and  $S$  a finite symmetric generating set of  $\Gamma$ . Let  $\varphi_S : \mathfrak{a}^+ \rightarrow [0, \infty) \cup \{-\infty\}$  be the growth indicator of  $S$ . Then, the Weyl chamber element  $\vec{\Lambda} \in \mathfrak{a}^{++}$  given by Corollary 9.1 is the unique point where  $\varphi_S$  reaches its maximum value  $\log \lambda_S$ .  $\square$*

## 9.2 | A connection to the work of Lubotzky–Mozes–Raghunathan

Here we let  $G$  be a connected semisimple real Lie group and  $\Gamma < G$  a finitely generated Zariski-dense subgroup, endowed with a finite symmetric generating set  $S$ . Let  $K < G$  be a maximal compact subgroup and  $d_G$  a left- $G$ -invariant and bi- $K$ -invariant Riemannian metric on  $G$  induced by the Killing form. If  $\Gamma$  is a uniform lattice in  $G$ , then it is not hard to see that the word-metric  $d_S$  is Lipschitz equivalent  $d_G$  (see, e.g., [53, Proposition 3.2]). The situation is much less clear for nonuniform lattices. Confirming a conjecture of Kazhdan (see [40]), Lubotzky–Mozes–Raghunathan [53] have shown that if  $G$  has  $\mathbb{R}$ -rank at least two and  $\Gamma$  is an irreducible lattice in  $G$ , then  $d_S$  and  $d_G$  are Lipschitz equivalent. In other words, there is a constant  $C > 1$  such that for every  $n \in \mathbb{N}$  and  $g \in S_n$ , we have

$$C^{-1}n \leq d_G(g, \text{id}) \leq Cn.$$

This equivalence breaks down for rank-one simple Lie groups in which case the word-metric  $d_S$  of a (nonuniform) lattice can be exponentially distorted in the terminology of [40, section 3]. This is for example the case for  $SL_2(\mathbb{Z}) < SL_2(\mathbb{R})$ . When  $\Gamma$  is only required to be Zariski-dense, the connection between  $d_S$  and  $d_G$  is much less clear. In many cases (e.g., if  $\Gamma$  is not discrete), it does not make sense to ask for Lipschitz equivalence of every element of  $\Gamma$  as there can be elements of  $\Gamma$  with arbitrarily large  $d_S$ -length but small  $d_G$ -length.

One way to study the connection between  $d_S$  and  $d_G$ , despite the fact they can not be Lipschitz equivalent, is to ask whether there is an equivalence  $d_S \sim d_G$  for most of the elements of  $\Gamma$ . Our counting large deviation Theorem 1.5 (and Corollary 9.2) then have the following consequence that establishes such a statistical relation between  $d_S$  and  $d_G$  for Gromov-hyperbolic groups.

**Corollary 9.3.** *Let  $\Gamma$  be a Zariski-dense, nonelementary Gromov-hyperbolic subgroup of a real semisimple Lie group  $G$ . Then, for every finite symmetric generating set  $S$  of  $\Gamma$  and constant  $\epsilon > 0$ , there exists a subset  $T_\epsilon$  of  $\Gamma$  with the property that*

$$\frac{\#(S_n \setminus (T_\epsilon \cap S_n))}{\#S_n} = O(e^{-\alpha n}) \quad (9.1)$$

for some  $\alpha > 0$ , and there exists a constant  $\Lambda = \Lambda(S) > 0$  such that for every  $n \in \mathbb{N}$  and  $g \in S_n \cap T_\epsilon$ , we have

$$n(\Lambda - \epsilon) \leq d_G(g, \text{id}) \leq n(\Lambda + \epsilon). \quad (9.2)$$

A subset of  $\Gamma$  satisfying (9.1) can be called *S-exponentially generic* in  $\Gamma$  in the terminology of [46].

*Proof.* It suffices to work with the symmetric space  $G/K$  and the  $G$ -invariant metric  $d_{G/K}$  induced by the Killing form. Let  $\mathfrak{a}^+$  be a Weyl chamber in a Cartan subspace  $\mathfrak{a}$  of the Lie algebra  $\mathfrak{g}$  of  $G$  such that we have the Cartan decomposition  $K \exp(\mathfrak{a}^+)K$ . Denoting by  $\|\cdot\|$  the norm induced by the Killing form on  $\mathfrak{a}$ , by [8, section 6.7.4], for any  $g \in G$ , we have  $\|\vec{\kappa}(g)\| = d_{G/K}(g \cdot o, o)$ . The result now follows from Corollary 9.2.  $\square$

*Remark 9.4.* By replacing the use of Theorem 1.5 (in the form of Corollary 9.2) by Theorem 1.1, one can obtain a version of Corollary 9.3 valid for any left- $G$ -invariant Riemannian metric  $d$  on  $G$  but  $T_\epsilon$  being only *S-generic* for  $\Gamma$  instead of *S-exponentially generic*. Here, by *S-generic* for  $\Gamma$ , we understand a subset satisfying (9.1) with  $O(e^{-\alpha n})$  replaced by  $o(1)$ .

## ACKNOWLEDGMENTS

The authors thank Emmanuel Breuillard and Ryokichi Tanaka for useful discussions and suggestions. They also thank the anonymous referee for helpful suggestions/corrections and for a careful reading of their work. Cagri Sert is supported by the Grant SNF Ambizione 193481.

## JOURNAL INFORMATION

The *Proceedings of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and

mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

## REFERENCES

1. R. Aoun, P. Mathieu, and C. Sert, *Random walks on hyperbolic spaces: second order expansion of the rate function at the drift*, to appear in JEP, 2021.
2. R. Aoun and C. Sert, *Law of large numbers for the spectral radius of random matrix products*, Amer. J. Math. **143** (2021), no. 3, 995–1010.
3. R. Aoun and C. Sert, *Random walks on hyperbolic spaces: concentration inequalities and probabilistic Tits alternative*, arXiv:2101.08222 v3, 2021.
4. H. Baik, I. Choi, and D. M. Kim, *Linear growth of translation lengths of random isometries on Gromov hyperbolic spaces and Teichmüller spaces*, arXiv:2103.13616, 2021.
5. Y. Benoist and J.-F. Quint, *Stationary measures and invariant subsets of homogeneous spaces (III)*, Ann. Math. **178** (2013), 1017–1059.
6. Y. Benoist and J.-F. Quint, *Central limit theorem for linear groups*, Ann. Probab. **44** (2016), 1308–1340.
7. Y. Benoist and J.-F. Quint, *Central limit theorem on hyperbolic groups*, Izv. Math. **80** (2016), no. 1, 3–23.
8. Y. Benoist and J.-F. Quint, *Random walks on reductive groups*, Springer, Cham, 2016.
9. P. Billingsley, *Convergence of probability measures*, John Wiley & Sons, Hoboken, NJ, 2013.
10. P. Bougerol, *Théorèmes de la limite centrale pour les produits de matrices en dépendance Markovienne. Résultats récents*, Probability measures on groups VIII, Springer, Berlin, Heidelberg, 1986, pp. 225–240.
11. P. Bougerol, *Théorèmes limite pour les systèmes linéaires à coefficients markoviens*, Probab. Theory Related Fields **78** (1988), no. 2, 193–221.
12. P. Bougerol, *Comparaison des exposants de Lyapounov des processus markoviens multiplicatifs*, Ann. de l'IHP Poincaré Probab. Stat. **24** (1988), no. 4, 439–489.
13. P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, vol. 8, Springer Science & Business Media, Berlin, 2012.
14. A. Boulanger, P. Mathieu, C. Sert, and A. Sisto, *Large deviations for random walks on Gromov-hyperbolic spaces*, to appear in Annales Scientifiques de l'École Normale Supérieure, 2020.
15. R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–459.
16. E. Breuillard and C. Sert, *The joint spectrum*, J. Lond. Math. Soc. **103** (2021), no. 3, 943–990.
17. D. Calegari, *The ergodic theory of hyperbolic groups: Geometry and topology down under*, Contemp. Math. **597** (2013), 1343–1369.
18. D. Calegari and K. Fujiwara, *Combable functions, quasimorphism, and the central limit theorem*, Ergodic Theory Dynam. Systems **30** (2009), 1343–1369.
19. J. Cannon, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata **16** (1984), 123–148.
20. J. Cannon, D. Epstein, D. Holt, S. Levy, M. Paterson, and W. Thurston, *Word processing in groups*, Jones and Barlett Publ., Boston, MA, 1992.
21. S. Cantrell, *Statistical limit laws for hyperbolic groups*, Trans. Amer. Math. Soc. **374** (2021), 2687–2732.
22. S. Cantrell, *Typical behaviour along geodesic rays in hyperbolic groups*, Math. Z. **297** (2021), 711–727.
23. S. Cantrell and R. Tanaka, *The Manhattan curve, ergodic theory of topological flows and rigidity*, arXiv:2104.13451, 2021.
24. P. E. Caprace, Y. Cornuier, N. Monod, and R. Tessera, *Amenable hyperbolic groups*, J. Eur. Math. Soc. **17** (2015), no. 11, 2903–2947.
25. I. Choi, *Pseudo-Anosovs are exponentially generic in mapping class groups*, arXiv:2110.06678, 2021.
26. M. Coornaert, *Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math. **159** (1993), 241–270.
27. M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et théorie des groupes: les groupes hyperboliques de Gromov*, vol. 1441, Springer, Berlin, 2006.
28. C. Cuny, J. Dedecker, and C. Jan, *Limit theorems for the left random walk on  $GL_d(\mathbb{R})$* , Ann. Inst. H. Poincaré Probab. Statist. **53** (2017), no. 4, 1839–1865.

29. C. Cuny, J. Dedecker, F. Merlevède, and M. Peligrad, *Berry–Esseen type bounds for the left random walk on  $GL_d(\mathbb{R})$  under polynomial moment conditions*, arXiv: <https://arxiv.org/abs/2110.10937>, 2021.
30. T. Das, D. Simmons, and M. Urbański, *Geometry and dynamics in Gromov hyperbolic metric spaces*, vol. 218, Amer. Math. Soc., Providence, RI, 2017.
31. H. Furstenberg, *Noncommuting random products*, Trans. Amer. Math. Soc. **108** (1963), no. 3, 377–428.
32. H. Furstenberg and H. Kesten, *Products of random matrices*, Ann. Math. Statist. **31** (1960), no. 2, 457–469.
33. I. Gekhtman, S. J. Taylor, and G. Tiozzo, *Counting loxodromics for hyperbolic actions*, J. Topol. **11** (2018), no. 2, 379–419.
34. I. Gekhtman, S. J. Taylor and G. Tiozzo, *Central limit theorems for counting measures in coarse negative curvature*, arXiv:2004.13084, 2020.
35. É. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d’après Mikhael Gromov*, Progr. Math. **83** (1990), 103–165.
36. I. Goldsheid and G. Margulis, *Lyapunov indices of a product of random matrices*, Russian Math. Surveys **44** (1989), 11–81.
37. A. Goldsborough and A. Sisto, *Markov chains on hyperbolic-like groups and quasi-isometries*, arXiv:2111.09837, 2021.
38. S. Gouëzel, *Exponential bounds for random walks on hyperbolic spaces without moment conditions*, arXiv: <https://arxiv.org/abs/2102.01408>, 2021.
39. S. Gouëzel, F. Mathéus, and F. Maucourant, *Entropy and drift in word hyperbolic groups*, Invent. Math. **211** (2018), 1201–1255.
40. M. Gromov, *Hyperbolic groups*, Essays in group theory, MSRI 8, Springer, New York, 1987, pp. 75–263.
41. M. Gromov, *Asymptotic invariants of infinite groups*, Geometric group theory, vol. 2, London Mathematical Society Lecture Note Series, vol. 182, Cambridge University Press, Cambridge, 1993, pp. 1–295.
42. Y. Guivarc’h, *Sur la loi des grands nombres et le rayon spectral d’une marche aléatoire*, Conference on Random Walks (Kleebach, 1979), 3, Astérisque, vol. 74, Soc. Math., France, Paris, 1980, pp. 47–98.
43. Y. Guivarc’h, *Exposants caractéristiques des produits de matrices aléatoires en dépendance Markovienne*, Probability measures on groups VII, Springer, Berlin, Heidelberg, 1984, pp. 161–181.
44. M. P. Horsham, *Central limit theorems for quasi-morphisms of surface groups*, Ph.D. thesis, The University of Manchester (United Kingdom), 2007.
45. M. Horsham and R. Sharp, *Lengths, quasi-morphisms and statistics for free groups*, in “Spectral analysis in geometry and number theory”, Contemp. Math. **484** (2009), 219–237.
46. V. Kaimanovich, I. Kapovich, and P. Schupp, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel J. Math. **157** (2007), no. 1, 1–46.
47. I. Kapovich, P. Schupp, and V. Shpilrain, *Generic properties of Whitehead’s algorithm and isomorphism rigidity of random one-relator groups*, Pacific J. Math. **223** (2006), no. 1, 113–140.
48. I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113, Springer Science & Business Media, Berlin, 2012.
49. S. Lalley, *Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tessellations and their fractal limits*, Acta Math. **163** (1989), no. 1, 1–55.
50. É. Le Page, *Théorèmes limites pour les produits de matrices aléatoires*, Probability measures on groups, Springer, Berlin, Heidelberg, 1982, pp. 258–303.
51. Q. Liu and F. Watbled, *Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment*, Stochastic Process. Appl. **119** (2009), no. 10, 3101–3132.
52. A. Lubotzky, S. Mozes, and M. S. Raghunathan, *Cyclic subgroups of exponential growth and metrics on discrete groups*, Comptes Rendus-académie des sciences, Paris Série **1** (1993), no. 317, 735–735.
53. A. Lubotzky, S. Mozes, and M. S. Raghunathan, *The word and Riemannian metrics on lattices of semisimple groups*, Publ. Math. Inst. Hautes Études Sci. **91** (2000), no. 1, 5–53.
54. J. Maher and G. Tiozzo, *Random walks on weakly hyperbolic groups*, J. Reine Angew. Math. **2018** (2018), no. 742, 187–239.
55. I. Mineyev, *Flows and joins of metric spaces*, Geom. Topol. **9** (2005), 403–482.
56. I. D. Morris and C. Sert, *A converse statement to Hutchinson’s theorem and a dimension gap for self-affine measures*, to appear in JEMS, arXiv:1909.08532, 2019.
57. W. Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–66.



58. M. Pollicott and R. Sharp, *Comparison theorems and orbit counting in hyperbolic geometry*, Trans. Amer. Math. Soc. **350** (1998), 473–499.
59. M. Pollicott and R. Sharp, *Statistics of matrix products in hyperbolic geometry*, Dynamical numbers: interplay between dynamical systems and number theory, Contemp. Math. **532** (2011), 213–230.
60. R. Prohaska and C. Sert, *Markov random walks on homogeneous spaces and Diophantine approximation on fractals*, Trans. Amer. Math. Soc. **373** (2020), no. 11, 8163–8196.
61. G. Royer, *Croissance exponentielle de produits markoviens de matrices aléatoires*, Ann. I.H.P **16** (1980), 49–62.
62. C. Sert, *Joint spectrum and large deviation principle for random matrix products*, Ph.D. thesis, Université Paris-Saclay, 2016.
63. C. E. Shannon and W. Weave, *The mathematical theory of communication*, University of Illinois Press, Illinois, 1963.
64. G. Sinai, *The asymptotic behaviour of the number of closed orbits on a compact manifold of negative curvature*, Transl. A. M. S. **73** (1968), 227–250.
65. R. Tanaka, *Topological flows for hyperbolic groups*, Ergodic Theory Dynam. Systems **41** (2020), 1–47.
66. A. D. Virtser, *On products of random matrices and operators*, Theory Prob. Appl. **24** (1979), 367–377.
67. H. Xiao, I. Grama, and Q. Liu, *Berry–Esseen bounds and moderate deviations for the random walk on  $GL_d(\mathbb{R})$* , arXiv: <https://arxiv.org/abs/2010.00557>, 2020.
68. W. Y. Yang, *Genericity of contracting elements in groups*, Math. Ann. **376** (2020), no. 3, 823–861.