# Tunisian Journal of Mathematics

an international publication organized by the Tunisian Mathematical Society


## Exponential bounds for random walks on hyperbolic spaces without moment conditions


Sébastien Gouëzel



2022 vol. 4 no. 4

msp

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 4, No. 4, 2022
https://doi.org/10.2140/tunis.2022.4.635

# Exponential bounds for random walks on hyperbolic spaces without moment conditions

Sébastien Gouëzel

We consider nonelementary random walks on general hyperbolic spaces. Without any moment condition on the walk, we show that it escapes linearly to infinity, with exponential error bounds. We even get such exponential bounds up to the escape rate of the walk. Our proof relies on an inductive decomposition of the walk, recording times at which it could go to infinity in several independent directions, and using these times to control further backtracking.

## 1. Introduction

Let $X$ be a Gromov-hyperbolic space, with a fixed basepoint $o$. Fix a discrete probability measure $\mu$ on the space of isometries of $X$. We assume that $\mu$ is *nonelementary*: in the semigroup generated by the support of $\mu$, there are two loxodromic elements with disjoint fixed points. Let $g_0, g_1, \dots$ be independent isometries of $X$ distributed according to $\mu$. One can then define a random walk on $X$ given by $Z_n \cdot o$, where $Z_n = g_0 \cdots g_{n-1}$.

In general, results in the literature fall into two classes, qualitative and quantitative, where the second class requires more stringent assumptions on the walk.

Without any moment assumption, it is known that $Z_n \cdot o$ converges almost surely to a point on the boundary $\partial X$, thanks to a beautiful nonconstructive argument originally due to Furstenberg [1963] in a matrix setting but that works in our setting when $X$ is proper, and extended to the general situation above by Maher and Tiozzo [2018]. The idea is to use a stationary measure on the boundary of $X$ and the martingale convergence theorem there to obtain the convergence of the random walk. When $X$ is not proper, the boundary is not compact, and showing the existence of a stationary measure on it is a difficult part of [Maher and Tiozzo 2018]. In that article, the authors also show linear progress, in the following sense: there exists $\kappa > 0$ such that, almost surely, $\liminf \frac{1}{n} d(o, Z_n \cdot o) \geqslant \kappa$.

Assuming additional moment conditions, one gets stronger results. From [Maher and Tiozzo 2018], if $\mu$ has finite support, then $\mathbb{P}(d(o, Z_n \cdot o) \leqslant \kappa n)$ is exponentially

*MSC2020:* 20F67, 20P05, 60B15, 60F10.
*Keywords:* large deviations, random walks, hyperbolic groups.

small, for some $\kappa > 0$, and we say that the walk makes linear progress with exponential decay. The finite support assumption has been weakened to an exponential moment condition in [Sunderland 2020]. More recently, still under an exponential moment condition, [Boulanger et al. 2021] shows (among many other results) that the exponential bound holds for any $\kappa$ strictly smaller than the escape rate $\ell = \lim \frac{1}{n}\mathbb{E}(d(o, Z_n \cdot o))$.

When $X$ is a hyperbolic group, one has in fact linear progress with exponential decay without any moment assumption: this follows from nonamenability of the group, and the fact that the cardinality of balls is at most exponential. This argument breaks down when the space is nonproper, though, as in many interesting examples such as the curve complex.

Our goal in this paper is to show that, to have linear progress with exponential decay, even in its strongest versions, there is no need for any moment condition. Define the escape rate of the walk $\ell(\mu) = \lim \frac{1}{n}\mathbb{E}(d(o, Z_n \cdot o))$ if $\mu$ has a moment of order 1, i.e., $\sum \mu(g) d(o, g \cdot o) < \infty$, and $\ell(\mu) = \infty$ otherwise.

Our first result is that the escape rate is positive, with an exponential error term.

**Theorem 1.1.** *Consider a discrete nonelementary measure on the space of isometries of a Gromov-hyperbolic space $X$ with a basepoint $o$. Then there exists $\kappa > 0$ such that, for all $n$,*

$$\mathbb{P}(d(o, Z_n \cdot o) \leqslant \kappa n) \leqslant e^{-\kappa n}.$$

One recovers in particular that $\ell(\mu) > 0$, a fact already proved in [Maher and Tiozzo 2018]. The control in the previous theorem can in fact be established up to the escape rate:

**Theorem 1.2.** *Under the assumptions of Theorem 1.1, consider $r < \ell(\mu)$. Then there exists $\kappa > 0$ such that, for all $n$,*

$$\mathbb{P}(d(o, Z_n \cdot o) \leqslant rn) \leqslant e^{-\kappa n}.$$

In particular, when $\mu$ has no moment of order 1, this implies that $\frac{1}{n} d(o, Z_n \cdot o) \to +\infty$ almost surely.

We also get the corresponding statement concerning directional convergence to infinity. For $\xi \in \partial X$ and $x, y \in X$, denote the corresponding Gromov product by

$$(x, \xi)_y = \inf_{z_n \to \xi} \liminf_n (x, z_n)_y, \tag{1-1}$$

where $(x, z_n)_y = \frac{1}{2}(d(y, x) + d(y, z_n) - d(x, z_n))$ is the usual Gromov product inside the space (see Section 3 for more background on Gromov-hyperbolic spaces). The limit only depends on the choice of the sequence $z_n$ up to $2\delta$. Intuitively, $(x, \xi)_y$ would be roughly the distance from $y$ to a geodesic between $x$ and $\xi$ if the space were geodesic. It is also the amount that $x$ has moved in the direction of $\xi$ compared to $y$. A sequence $x_n$ converges to $\xi$ if and only if $(x_n, \xi)_o \to \infty$.

**Theorem 1.3.** *Under the assumptions of Theorem 1.2, $Z_n \cdot o$ converges almost surely to a point $Z_\infty \in \partial X$. Moreover, for any $r < \ell(\mu)$, there exists $\kappa > 0$ such that, for all $n$,*

$$\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant rn) \leqslant e^{-\kappa n}.$$

Theorem 1.3 readily implies Theorem 1.2 as $(Z_n \cdot o, Z_\infty)_o \leqslant d(o, Z_n \cdot o)$, which follows directly from the definition.

The convergence statement in Theorem 1.3 is due to [Maher and Tiozzo 2018]. The novelty is the quantitative exponential bound, without any moment assumption. Note that, in both theorems, when $\mu$ has no moment of order 1, one may take any $r \geqslant 0$, so the conclusion is superlinear growth with exponential decay.

It follows from subadditivity that the sequence $-\frac{1}{n}\log(\mathbb{P}(d(o, Z_n \cdot o) \leqslant rn))$ converges to a limit $I(r)$, for any $r \leqslant \ell$. This is a rate function in the classical sense of large deviations in probability theory. Theorem 1.2 shows that the rate function is strictly positive for $r < \ell$, recovering part of [Boulanger et al. 2021, Theorem 1.1] while removing their exponential moment assumption. Note that [Boulanger et al. 2021] also obtains exponential estimates for upper deviation inequalities $\mathbb{P}(d(o, Z_n \cdot o) \geqslant rn)$ for $r > \ell$. These estimates can not hold without exponential moments, since exponential controls for lower and upper deviation probabilities imply an exponential moment for the measure, see [Boulanger et al. 2021, Section 3.1].

**Remark 1.4.** The fact that we use discrete measures in the above theorems is for convenience only, to avoid discussing measurability issues and conditioning on zero measure sets. Suitable versions removing discreteness, but adding measurability and separability conditions, hold with the same proofs.

Also, it is not essential to our argument that the isometries we consider are onto: if the measure $\mu$ is supported on the space of isometric embeddings of $X$ into itself, our proof goes through since we never use the reversed random walk. Note however that, to define the fact that $\mu$ is nonelementary, we still need to have two loxodromic (surjective) isometries with disjoint fixed points in the semigroup generated by $\mu$.

**Remark 1.5.** The explicit nature of our estimates makes them robust under perturbations: it follows from our proof that Theorem 1.3 holds uniformly, with the same $\kappa$, over all measures in a neighborhood of $\mu$. We state this in Proposition 5.15, and recover from this the (already known) continuity of the escape rate as a function of the driving measure.

Our approach is elementary, in the spirit of [Mathieu and Sisto 2020] and [Boulanger et al. 2021], the latter article being a strong inspiration for our work, and does not rely on any boundary theory. The main intuition is the following. In the hyperbolic plane, we define a path as follows: walk straight on during a distance $d_1$, then turn by an angle $\theta_1 \leqslant \bar{\theta} < \pi$, then walk straight on during a distance $d_2$, then turn by an angle $\theta_2 \leqslant \bar{\theta}$, and so on. If all the lengths $d_i$ are larger than a constant $D = D(\bar{\theta})$, then this path is essentially going straight to infinity, and at time $n$ it is roughly at distance $d_1 + \cdots + d_n$ of the origin. The problem when doing a random walk is that the analogues of the angles $\theta_i$ could be equal to $\pi$, i.e., the walker could come back exactly along its footsteps. But this should not happen often. Our main input is a technical way to justify that indeed it does not happen often, in a precise quantitative version. We will keep track of some times (called *pivotal times* below) at which the random walk can choose some direction, with most choices leading to progress towards infinity, which will be implemented through the notion of Schottky set coming from [Boulanger et al. 2021], and at which we will keep some degree of freedom in an inductive construction. Of course, backtracking can happen later on, and we will spend the degree of freedom we had kept to still control the behavior after backtracking.

We could give directly the proof of Theorem 1.3, but it would be very hard to follow. Instead, we will start with proofs of easier statements, and add new ingredients in increasingly complicated proofs. Section 2 is devoted to the simplest instance of our proof, in the free group, where everything is as transparent as possible. Then, Section 3 introduces some tools of Gromov-hyperbolic geometry (notably chains, shadows and Schottky sets) that will be used to extend the previous proof to a nontree setting. Section 4 uses these tools in a crude way to prove Theorem 1.1, that is, linear escape with exponential decay, and also convergence at infinity with exponential bounds. Section 5 follows the same strategy but in a more refined way, to get Theorems 1.2 and 1.3.

## 2. Linear escape with exponential decay on free groups

The goal of this section is to illustrate the concept of pivotal times in the simplest possible setting. We show that, for a class of measures without moments on the free group, there is linear escape with exponential decay. Of course, this follows from nonamenability. Instead of the result, what matters here is the proof: the rest of the paper is an extension of the same idea to technically more involved contexts (general measures, Gromov-hyperbolic spaces), but the main insight can be explained much more transparently in a tree setting.

**Theorem 2.1.** *Let $d \geqslant 3$. Let $\mu$ be a probability measure on $\mathbb{F}_d$ that can be written as $\mu_S * \nu$, where $\mu_S$ is the uniform probability measure on the canonical generators of $\mathbb{F}_d$, and $\nu$ is a probability measure with $\nu(e) = 0$. Let $Z_n = g_1 \cdots g_n$, where the $g_i$ are independent and distributed according to $\mu$. There exists $\kappa > 0$ (independent* *of* $\nu$ *and of* $d$) *such that, for all* $n$,

$$\mathbb{P}(|Z_n| \leqslant \kappa n) \leqslant e^{-\kappa n}.$$

**Remark 2.2.** The fact that $\kappa$ can be chosen independently of $\nu$ and of $d$ does not follow from nonamenability, and is really a byproduct of our proof technique. Indeed, for fixed $d$ and $\nu$ the nonamenability shows that the conclusion of Theorem 2.1 holds for some $\kappa > 0$, but it does not provide any a priori bound on $\kappa$ without further work.

**Remark 2.3.** The restrictions $d \geqslant 3$ and $\nu(e) = 0$ are simplifying assumptions to have a proof that is as streamlined as possible. In the next sections, we will prove analogous theorems but for general measures, on general hyperbolic spaces.

The key point in the proof of Theorem 2.1 is the next lemma.

**Lemma 2.4.** *There exists* $\kappa > 0$ *satisfying the following. Consider* $d \geqslant 3$ *and* $n \geqslant 0$. *Fix* $w_1, \dots, w_n$ *nontrivial words in* $\mathbb{F}_d$, *and let* $Z_n = s_1 w_1 \cdots s_n w_n$, *where the* $s_i$ *are generators of* $\mathbb{F}_d$, *chosen uniformly and independently. Then* $\mathbb{P}(|Z_n| \leqslant \kappa n) \leqslant e^{-\kappa n}$.

This lemma directly implies Theorem 2.1, by conditioning with respect to the realizations of $\nu$ and just keeping the randomness coming from the factor $\mu_S$ in $\mu = \mu_S * \nu$.

To prove the lemma, one wants to argue that the walk does not backtrack too much. Of course, the walk can backtrack completely: as the size of $w_i$ is not controlled, it may happen that $w_n$ is exactly inverse to $s_1 w_1 \cdots s_n$ and therefore that $Z_n = e$. However, this is unlikely to happen for most choices of $s_1, \dots, s_n$.

A difficulty is that the control of the distance to the origin is not well behaved under the walk. For instance, assume that $Z_{n-2} = e$, that $w_{n-1}$ is very long (of length $2n$, say) and that for some generators $s$ and $t$, one has $t w_n = (s w_{n-1})^{-1}$. Then $Z_{n-1}$ is far away from the origin, and in particular it satisfies the inequality $|Z_{n-1}| > n$. However, $Z_n$ is equal to the origin if $s_{n-1} = s$ and $s_n = t$, which happens with probability $1/(2d)^2$. This is not exponentially small, even though the distance control at time $n-1$ is good.

For this reason, we will not try to control inductively the distribution of the distance to the origin. Instead, we will control a number of branching points of the random walk up to time $n$, that we call *pivotal points*. In the general case of random walks in hyperbolic spaces, the definition will be quite involved, but for trees one can give a direct definition as follows. Denote by $\gamma_n$ the path in the Cayley graph of $\mathbb{F}_d$ corresponding to the walk up to $Z_n$, i.e., the concatenation of the geodesics from $e$ to $s_1$ then to $s_1 w_1$ then to $s_1 w_1 s_2$ and so on until $s_1 w_1 s_2 w_2 \cdots s_n w_n = Z_n$.

**Definition 2.5.** A time $k \in [1, n]$ is a pivotal time (with respect to $n$) if $s_k$ is the inverse neither of the last letter of $Z_{k-1}$, nor of the first letter $(w_k)_0$ of $w_k$, so that the path $\gamma_n$ is locally geodesic of length 3 around $Z_{k-1}$, and moreover the path $\gamma_n$ does not come back to $Z_{k-1}s_k$ afterwards.

We will denote by $P_n$ the set of pivotal times with respect to $n$.

In other words, $k$ is pivotal if the walk at time $k$ goes away from the origin during two steps ($s_k$ and then $(w_k)_0$) and then remains stuck in the subtree based at $Z_{k-1}s_k(w_k)_0$.

The evolution of the set of pivotal times is not monotone: if the walk backtracks a lot, then many times that were pivotal with respect to $n$ will not be anymore pivotal with respect to $n+1$, since the nonbacktracking condition is not satisfied anymore. On the other hand, the only possible new pivotal point is the last one: $P_{n+1} \subseteq P_n \cup \{n+1\}$.

We will say that a sequence $(s'_1, \dots, s'_n)$ is *pivoted* from $\bar{s} = (s_1, \dots, s_n)$ if they have the same pivotal times and, additionally, $s'_k = s_k$ for all $k$ which is not a pivotal time. This is an equivalence relation. Moreover, a sequence has many pivoted sequences: if $k$ is a pivotal time and one changes $s_k$ to $s'_k$ which still satisfies the local geodesic condition (the inverse of $s'_k$ is different from the last letter of $Z_{k-1}$ and from the first letter of $w_k$), then we claim that $(s_1, \dots, s'_k, \dots, s_n)$ is pivoted from $(s_1, \dots, s_n)$. Indeed, the part of $\gamma_n$ originating from $Z_{k-1}s_k(w_k)_0$ never comes back on the edge from $Z_{k-1}$ to $Z_{k-1}s_k$, not even on its endpoints, so changing $s_k$ to $s'_k$ does not change this fact. Thus the behavior of $\gamma'_n$ after $Z_{k-1}$ is exactly the same as that of $\gamma_n$, but in a different subtree — one has pivoted the end of $\gamma_n$ around $Z_{k-1}s_k$, hence the name. In particular, subsequent pivotal times are the same. Moreover, since the trajectory never comes back before $Z_{k-1}s_k$, pivotal times before $k$ are not affected, and are the same for $\gamma_n$ and $\gamma'_n$.

More generally, denoting the pivotal times by $p_1 < \cdots < p_q$, then changing the $s_{p_i}$ to $s'_{p_i}$ still satisfying the local geodesic condition gives a pivoted sequence. Let $\mathcal{E}_n(\bar{s})$ be the set of sequences which are pivoted from $\bar{s}$. Conditionally on $\mathcal{E}_n(\bar{s})$, the previous discussion shows that the random variables $s'_{p_i}$ are independent, but not identically distributed as each of them is drawn from some subset of the generators depending on $i$, of cardinality $|S|-1$ or $|S|-2$.

**Proposition 2.6.** *Let $A_n = |P_n|$ be the number of pivotal times. Then, in distribution, $A_{n+1} \geqslant A_n + U$ where $U$ is a random variable independent from $A_n$ and distributed as*

$$\mathbb{P}(U=-j) = \frac{2d-3}{d(2d-2)^j} \textit{ for } j>0, \quad \mathbb{P}(U=0)=0 \quad \textit{and} \quad \mathbb{P}(U=1) = \frac{d-1}{d}.$$

*In other words, $\mathbb{P}(A_{n+1} \geqslant i) \geqslant \mathbb{P}(A_n + U \geqslant i)$ for all $i$.*

*Proof.* Let us fix a sequence $\bar{s} = (s_1, \dots, s_n)$, and let $q = |P_n|$ be its number of pivotal times. We will prove the estimate by conditioning on $\mathcal{E}_n(\bar{s})$. Let $\bar{s}' \in \mathcal{E}_n(\bar{s})$.

First, assume there are no pivotal points, so $q=0$. Then for each $\bar{s}'$ there are at least $2d-2$ generators whose inverses are different from the last letter of $Z'_n$ and from the first letter of $w_{n+1}$, giving rise to one pivotal time in $P'_{n+1}$ with probability at least $(2d-2)/(2d) = \mathbb{P}(U=1)$. Otherwise, $|P'_{n+1}| = 0$. Conditionally on $\mathcal{E}_n(\bar{s})$, it follows that the conclusion of the lemma holds.

Assume now that there is at least one pivotal point. From the last pivotal time onward, the behavior is the same over all the equivalence class $\mathcal{E}_n(\bar{s})$, so the last letter of $Z'_n$ does not depend on $\bar{s}'$. There are at least $2d-2$ generators of $\mathbb{F}_d$ whose inverses are different from the last letter of $Z'_n$ and from the first letter of $w_{n+1}$. If $s'_{n+1}$ is such a generator, then $P'_{n+1} = P'_n \cup \{n+1\}$. Therefore,

$$\mathbb{P}(A_{n+1} \geqslant q+1 \mid \mathcal{E}_n(\bar{s})) \geqslant \frac{2d-2}{2d}.$$

We have adjusted the definition of $U$ so that the right-hand side is $\mathbb{P}(U \geqslant 1)$.

Fix now $s'_{n+1}$ which is not such a nice generator. Then $s'_{n+1} w_{n+1}$ may backtrack, possibly until the last pivotal point $Z'_{p_q}$, thereby decreasing the number of pivotal points with respect to $n+1$. However, it may only backtrack further if the generator $s'_{p_q}$ is exactly the inverse of the corresponding letter in $w_{n+1}$. This can happen for $s'$, but then it will not happen for all the pivoted configurations of $s'$ obtained by changing $s'_{p_q}$ to another generator still satisfying the local geodesic condition. Therefore,

$$\mathbb{P}(A_{n+1} \leqslant q-2 \mid \mathcal{E}_n(\bar{s})) \leqslant \frac{2}{2d} \cdot \frac{1}{2d-2},$$

where the first factor corresponds to the choice of a generator $s'_{n+1}$ which does not satisfy the local geodesic condition, and the second factor corresponds to the choice of the specific generator for $s'_{p_q}$ to make sure that one backtracks further.

More generally, to cross $j$ pivotal times, there is one specific choice of generator at each of these pivotal times, which can only happen with a probability at most $1/(2d-2)$ at each of these times. Therefore, for $j \geqslant 1$,

$$\mathbb{P}(A_{n+1} \leqslant q-j \mid \mathcal{E}_n(\bar{s})) \leqslant \frac{2}{2d} \cdot \frac{1}{(2d-2)^{j-1}}.$$

We have adjusted the distribution of $U$ so that the right-hand side is exactly $\mathbb{P}(U \leqslant -j)$.

Finally, for $j > 0$, we obtain the inequalities

$$\mathbb{P}(A_{n+1} \leqslant q-j \mid \mathcal{E}_n(\bar{s})) \leqslant \mathbb{P}(U \leqslant -j) \quad \text{and} \quad \mathbb{P}(A_{n+1} \geqslant q+1 \mid \mathcal{E}_n(\bar{s})) \geqslant \mathbb{P}(U \geqslant 1).$$

Taking the complement in the first inequality yields $\mathbb{P}(A_{n+1} \geqslant q+k \mid \mathcal{E}_n(\bar{s})) \geqslant \mathbb{P}(U \geqslant k)$ for all $k \in \mathbb{Z}$. As $A_n$ is constant equal to $q$ on $\mathcal{E}_n(\bar{s})$, the right-hand side is $\mathbb{P}(A_n + U \geqslant q+k \mid \mathcal{E}_n(\bar{s}))$. Writing $i = q+k$, we have obtained for all $i$ the inequality

$$\mathbb{P}(A_{n+1} \geqslant i \mid \mathcal{E}_n(\bar{s})) \geqslant \mathbb{P}(A_n + U \geqslant i \mid \mathcal{E}_n(\bar{s})).$$

As this inequality is uniform over the conditioning, it gives the conclusion of the proposition. □

*Proof of Lemma 2.4.* Let $U_1, U_2, \dots$ be a sequence of i.i.d. random variables distributed like $U$ in Proposition 2.6. Iterating the proposition, one gets $\mathbb{P}(A_n \geqslant k) \geqslant \mathbb{P}(U_1 + \cdots + U_n \geqslant k)$. The random variables $U_i$ have an exponential moment. Moreover, when $d \geqslant 3$, their expectation equals $(2d-5)\cdot(d-1)/((2d-3)\cdot d)$ and is therefore positive. Large deviations for sums of i.i.d. real random variables with an exponential moment ensure the existence of $\kappa > 0$ such that $\mathbb{P}(U_1+\cdots+U_n \leqslant \kappa n) \leqslant e^{-\kappa n}$ for all $n$. Then $\mathbb{P}(A_n \leqslant \kappa n) \leqslant e^{-\kappa n}$. As the distance to the origin is bounded from below by the number of pivotal points, this proves Lemma 2.4, except that the constant $\kappa$ depends on the number of generators $d$. However, the random variables $U = U(d)$ depending on $d$ increase with $d$, in the sense that if $d \geqslant d'$, then $\mathbb{P}(U(d) \geqslant k) \geqslant \mathbb{P}(U(d') \geqslant k)$ for all $k$. Therefore, one can use the random variables $U(3)$ to obtain a lower bound in all free groups $\mathbb{F}_d$ with $d \geqslant 3$. □

The rest of the paper is devoted to the extension of this argument to general measures and general Gromov-hyperbolic spaces. While the intuition will remain the same, the definition of pivotal times will need to be adjusted, as there is no well-defined concept of subtree. Instead, we will use a suitable notion of shadow, and require that the walk after the pivotal time remain in the shadow. Also, to separate possible directions, we will rely on the notion of Schottky sets introduced by [Boulanger et al. 2021], instead of just using the generators as in the free group. These notions are explained in the next section.

## 3. Prerequisites on Gromov-hyperbolic spaces

Let $X$ be a metric space, and $x, y, z \in X$. Their Gromov product is defined by

$$(x, z)_y = \tfrac{1}{2}(d(x, y) + d(y, z) - d(x, z)).$$

Let $\delta \geqslant 0$. A metric space is $\delta$-Gromov hyperbolic if, for all $x, y, z, a$,

$$(x, z)_a \geqslant \min((x, y)_a, (y, z)_a) - \delta. \tag{3-1}$$

When the space is geodesic, this is equivalent (up to changing $\delta$) to the fact that geodesic triangles are thin, i.e., each side is contained in the $\delta$-neighborhood of the other two sides.

In the rest of the paper, $X$ is a $\delta$-hyperbolic metric space (without any geodesicity or properness or separability condition). We also fix a basepoint $o \in X$. We will sometimes mention geodesics to give some geometric intuition, but all our definitions and results work as well when the space is not geodesic.

**3A. *Boundary at infinity.*** We recall a few basic facts on the boundary at infinity of a Gromov-hyperbolic space that we will need later on.

A sequence $(x_n)_{n\in\mathbb{N}}$ is converging at infinity if $(x_n, x_m)_o$ tends to infinity when $m, n \to \infty$. Two sequences $(x_n)$ and $(y_n)$ which are converging at infinity are converging to the same limit if $(x_n, y_n)_o \to \infty$. This is an equivalence relation, thanks to the hyperbolicity inequality. Quotienting by this equivalence relation, one gets the boundary at infinity of the space $X$ denoted $\partial X$.

The $C$-shadow of a point $x$, seen from $o$, is the set of points $y$ such that $(y, o)_x \leqslant C$. We denote it with $\mathcal{S}_o(x; C)$. Geometrically, this would mean that a geodesic from $o$ to $y$ would go within distance $C + O(\delta)$ of $x$ if the space were geodesic, but it makes sense even when the space is not geodesic. Let us record a few classical properties of shadows.

**Lemma 3.1.** *For $y \in \mathcal{S}_o(x; C)$, one has $d(y, o) \geqslant d(x, o) - C$.*

*Proof.* We have

$$d(y, o) = d(y, x) + d(x, o) - 2(y, o)_x \geqslant (d(x, o) - d(y, o)) + d(x, o) - 2C.$$

Passing $-d(y, o)$ from the right-hand side to the left-hand side and dividing by 2 gives the conclusion. □

**Lemma 3.2.** *Let $C > 0$, and let $x_n \in X$ be such that $d(o, x_n) \to \infty$. Consider another sequence $y_p$ such that, for all $n$, eventually $y_p \in \mathcal{S}_o(x_n; C)$. Then $y_p$ converges at infinity.*

*Proof.* Fix $n$ large. For large enough $p$, one has $y_p \in \mathcal{S}_o(x_n; C)$, thus $(o, y_p)_{x_n} \leqslant C$. As $(o, y_p)_{x_n} + (x_n, y_p)_o = d(o, x_n)$, this gives $(x_n, y_p)_o \geqslant d(o, x_n) - C$.

For large enough $p, q$, we get, using hyperbolicity for the first inequality,

$$(y_p, y_q)_o \geqslant \min((y_p, x_n)_o, (y_q, x_n)_o) - \delta \geqslant d(o, x_n) - C - \delta. \tag{3-2}$$

As $d(o, x_n) \to \infty$ by assumption, it follows that $(y_p, y_q)_o \to \infty$, as claimed. □

**Lemma 3.3.** *Let $C > 0$ and $x \in X$. Consider $y \in \mathcal{S}_o(x; C)$, and a point $\xi \in \partial X$ which is a limit of points in $\mathcal{S}_o(x; C)$. Then*

$$(y, \xi)_o \geqslant d(o, x) - C - 3\delta.$$

*Proof.* Let $z_n \in \mathcal{S}_o(x; C)$ be a sequence converging to $\xi$. As the Gromov product at infinity does not depend on the sequence up to $2\delta$, we have $(y, \xi)_o \geqslant \liminf (y, z_n)_o - 2\delta$. Moreover, as both $y$ and $z_n$ belong to $\mathcal{S}_o(x; C)$, the inequality (3-2) gives $(y, z_n)_o \geqslant d(o, x) - C - \delta$. The conclusion follows. □

**3B.** ***Chains and shadows.*** In a geodesic hyperbolic space, $(x, z)_y$ is roughly the distance from $y$ to a geodesic between $x$ and $z$. In particular, if $(x, z)_y \leqslant C$ for some constant $C$, this means that the points $x, y, z$ are roughly aligned in this order, up to an error $C$. We will say that the points are *$C$-aligned* when they satisfy this condition, even when the space is not geodesic.

In a hyperbolic space, if in a sequence of points all consecutive points are $C$-aligned, and the points are separated enough, then the sequence is progressing linearly, and all points in the sequence are $C + O(\delta)$-aligned (see for instance [Ghys and de la Harpe 1990, Theorem 5.3.16]). We will need variations around this classical idea.

We start with distance estimates for three points.

**Lemma 3.4.** *Consider $x, y, z$ with $(x, z)_y \leqslant C$. Then $d(x, z) \geqslant d(x, y) - C$ and $d(x, z) \geqslant d(y, z) - C$.*

*Proof.* By symmetry, it suffices to prove the first inequality. We claim that $d(x, z) \geqslant d(x, y) - (x, z)_y$, which implies the result. Expanding the definition of the Gromov product, this inequality holds if and only if

$$\tfrac{1}{2}(d(y, x) + d(y, z) - d(x, z)) + d(x, z) \geqslant d(x, y).$$

This reduces to $d(y, z) + d(x, z) \geqslant d(x, y)$, which is the triangular inequality. □

The next lemma gives estimates for four points, from which results for more points will follow by induction.

**Lemma 3.5.** *Consider $w, x, y, z \in X$, and $C \geqslant 0$. Assume $(w, y)_x \leqslant C$ and $(x, z)_y \leqslant C + \delta$ and $d(x, y) \geqslant 2C + 2\delta + 1$. Then $(w, z)_x \leqslant C + \delta$.*

*Proof.* By definition of the Gromov product, $(x, z)_y + (y, z)_x = d(x, y)$. As $(x, z)_y \leqslant C + \delta$, we get $(y, z)_x \geqslant d(x, y) - C - \delta$. As $d(x, y) \geqslant 2C + 2\delta + 1$, this gives $(y, z)_x \geqslant C + \delta + 1$. Writing down the first condition and the hyperbolicity condition, we get

$$C \geqslant (w, y)_x \geqslant \min((w, z)_x, (z, y)_x) - \delta.$$

If the minimum were realized by $(z, y)_x$, we would get $C \geqslant (C + \delta + 1) - \delta$, a contradiction. Therefore, it is realized by $(w, z)_x$, which gives $(w, z)_x \leqslant C + \delta$. □

**Definition 3.6.** For $C, D \geqslant 0$, a sequence of points $x_0, \dots, x_n$ is a $(C, D)$-chain if one has $(x_{i-1}, x_{i+1})_{x_i} \leqslant C$ for all $0 < i < n$, and $d(x_i, x_{i+1}) \geqslant D$ for all $0 \leqslant i < n$.

**Lemma 3.7.** *Let $x_0, \dots, x_n$ be a $(C, D)$-chain with $D \geqslant 2C + 2\delta + 1$. Then $(x_0, x_n)_{x_1} \leqslant C + \delta$, and*

$$d(x_0, x_n) \geqslant \sum_{i=0}^{n-1} (d(x_i, x_{i+1}) - (2C + 2\delta)) \geqslant n. \tag{3-3}$$

*Proof.* Let us show by decreasing induction on $i$ that $(x_{i-1}, x_n)_{x_i} \leqslant C+\delta$, the result being true for $i = n-1$ by assumption. Assume it holds for $i+1$. Then the points $x_{i-1}, x_i, x_{i+1}, x_n$ satisfy the assumptions of Lemma 3.5, which gives $(x_{i-1}, x_n)_{x_i} \leqslant C+\delta$ as desired.

Let us now show that $d(x_j, x_n) \geqslant \sum_{i=j}^{n-1}(d(x_i, x_{i+1}) - (2C+2\delta))$ by decreasing induction on $j$, the case $j = n$ being trivial and the case $j = 0$ being (3-3). We have

$$\begin{aligned} d(x_j, x_n) &= d(x_j, x_{j+1}) + d(x_{j+1}, x_n) - 2(x_j, x_n)_{x_{j+1}} \\ &\geqslant d(x_j, x_{j+1}) + d(x_{j+1}, x_n) - (2C+2\delta), \end{aligned}$$

which concludes the induction. □

**Lemma 3.8.** *Let $x_0, \dots, x_n$ be a $(C, D)$-chain with $D \geqslant 2C+4\delta+1$. Then for all $i$, one has $(x_0, x_n)_{x_i} \leqslant C+2\delta$.*

*Proof.* Lemma 3.7 applied to the $(C, D)$-chain $x_i, x_{i+1}, \dots, x_n$ allows to show $(x_i, x_n)_{x_{i+1}} \leqslant C+\delta$. The same lemma applied to the $(C, D)$-chain $x_{i+1}, x_i, \dots, x_0$ gives $(x_{i+1}, x_0)_{x_i} \leqslant C+\delta$. Therefore, the points $x_0, x_i, x_{i+1}, x_n$ are $(C+\delta)$-aligned. Let us apply Lemma 3.5 to these points, with $C+\delta$ instead of $C$. It gives $(x_0, x_n)_{x_i} \leqslant C+2\delta$, as claimed. □

We will need to say that a point $z$ belongs to a half-space based at a point $y$ and directed towards a point $y^+$. The usual definition for this is the shadow of $y^+$ seen from $y$, defined as the set $\mathcal{S}_y(y^+; C)$ of points $z$ with $(y, z)_{y^+} \leqslant C$ for some suitable $C$. Unfortunately, this definition is not robust enough for our purposes as we need the property that being in a half-space and walking again from $z$, one stays in the half-space, which is not satisfied by this definition due to the loss of $\delta$ when one applies the hyperbolicity inequality.

A more robust definition can be given in terms of chains. If we have a chain, which goes roughly in a straight direction by the previous lemma, and if we prescribe the direction of its first jump, then we are essentially prescribing the direction of the whole chain. This makes it possible to define another notion that we call chain-shadow, as follows. The choice of the minimal distance $2C+2\delta+1$ between points in the chain in this definition is somewhat arbitrary, it should just be large enough that lemmas on the linear progress of chains apply.

**Definition 3.9.** Let $C \geqslant 0$ and $y, y^+, z \in X$. We say that $z$ belongs to the $C$-chain-shadow of $y^+$ seen from $y$ if there is a $(C, 2C+2\delta+1)$-chain $x_0 = y, x_1, \dots, x_n = z$ satisfying additionally $(x_0, x_1)_{y^+} \leqslant C$. We denote the chain-shadow by $\mathcal{CS}_y(y^+; C)$.

The next lemma shows that this definition of shadow is roughly equivalent to the usual definition in terms of the Gromov product $(y, z)_{y^+}$.

**Lemma 3.10.** *If $z \in \mathcal{CS}_y(y^+; C)$, then $(y, z)_{y_+} \leqslant 2C+\delta$ and $d(y, z) \geqslant d(y, y^+) - 2C - \delta$.*

*Proof.* Let $x_0 = y, x_1, \dots, x_n = z$ be a $(C, 2C + 2\delta + 1)$-chain as in the definition of chain-shadows. We have

$$\begin{aligned} d(y, z) &= d(y, x_1) + d(x_1, z) - 2(y, z)_{x_1} \\ &= d(y, y^+) + d(y^+, x_1) - 2(y, x_1)_{y_+} + d(x_1, z) - 2(y, z)_{x_1}. \end{aligned}$$

Let us bound $(y, x_1)_{y_+}$ with $C$ (by the definition of chain-shadows) and $(y, z)_{x_1}$ by $C + \delta$ (thanks to Lemma 3.7 applied to the chain $x_0, \dots, x_n$). Let us also bound from below $d(y^+, x_1) + d(x_1, z)$ with $d(y^+, z)$. We get

$$d(y, z) \geqslant d(y, y^+) + d(y^+, z) - 4C - 2\delta.$$

Expanding the definition of the Gromov product, this gives $(y, z)_{y^+} \leqslant 2C + \delta$. Then we get $d(y, z) \geqslant d(y, y^+) - 2C - \delta$ by applying Lemma 3.4 to $y, y^+, z$. □

**3C. *Schottky sets.*** To be able to prescribe enough directions at pivotal points, we will use a variation around the notion of Schottky set in [Boulanger et al. 2021]. This is essentially a finite set of isometries such that, for all $x$ and $y$, most of these isometries put $x$ and $sy$ in general position with respect to $o$, i.e., such that $x, o, sy$ are $C$-aligned for some given $C$.

**Definition 3.11.** Let $\eta, C, D \geqslant 0$. A finite set $S$ of isometries of $X$ is $(\eta, C, D)$-Schottky if:

- For all $x, y \in X$, we have $|\{s \in S, (x, sy)_o \leqslant C\}| \geqslant (1 - \eta)|S|$.
- For all $x, y \in X$, we have $|\{s \in S, (x, s^{-1}y)_o \leqslant C\}| \geqslant (1 - \eta)|S|$.
- For all $s \in S$, we have $d(o, so) \geqslant D$.

We could define analogously a notion of an $(\eta, C, D)$-probability measure, where the previous definition would be this property for the uniform measure on $S$.

The next proposition shows that one can find Schottky sets by using powers of two loxodromic isometries.

**Proposition 3.12.** *Fix two loxodromic isometries $u$ and $v$ of $X$, with disjoint sets of fixed points at infinity. For all $\eta > 0$, there exists $C > 0$ such that, for all $D > 0$, there exist $n \in \mathbb{N}$ and an $(\eta, C, D)$-Schottky set in $\{w_1 \cdots w_n : w_i \in \{u, v\}\}$.*

*Proof.* This is essentially a classical application of the ping-pong method. Proposition A.2 in [Boulanger et al. 2021] contains a slightly less precise statement, but their proof also gives our stronger version, as we explain now. Let $S_n = \{w_1 \cdots w_n : w_i \in \{u, v\}\}$.

The ping-pong argument at infinity shows that one can choose $n$ large enough so that for all $m$, the elements $w_1 \cdots w_m$ for $w_i \in \{u^n, v^n\}$ are all different, loxodromic, with disjoint sets of fixed points at infinity. Let us fix such an $n$, and then an $m$ with

$2^{-m} < \eta/2$, and denote these $2^m$ isometries with $g_1, \dots, g_{2^m}$. They all belong to $S_{nm}$. Let $g_i^+$ and $g_i^-$ be their attractive and repulsive fixed points.

Let $K$ be large enough. Define a neighborhood $V(g_i^+) = \{x \in X : (x, g_i^+)_o \geqslant K\}$ and a smaller neighborhood $V'(g_i^+) = \{x \in X : (x, g_i^+)_o \geqslant K + \delta\}$. In the same way, define $V(g_i^-)$ and $V'(g_i^-)$. If $K$ is large enough, then the $2^{m+1}$ sets $(V(g_i^{\pm}))_{i=1,\dots,2^m}$ are disjoint as the fixed points at infinity of the $g_i$ are all different. Moreover, for large enough $p$, then $g_i^p$ maps the complement of $V(g_i^-)$ to $V'(g_i^+)$, and the complement of $V(g_i^+)$ to $V'(g_i^-)$.

We claim that, for all $D$, if $p$ is large enough, then $S = \{g_1^p, \dots, g_{2^m}^p\}$ is an $(\eta, K+\delta, D)$-Schottky set. As all these elements belong to $S_{nmp}$, this will prove the theorem. First, the condition $d(o, so) \geqslant D$ for $s = g_i^p$ is true if $p$ is large enough, as $g_i$ is loxodromic. Let us show that $|\{s \in S, (x, sy)_o \leqslant K + \delta\}| \geqslant (1-\eta)|S|$ for all $x, y$ (the corresponding inequality with $s^{-1}$ is similar). There is at most one $s = g_i$ for which $y \in V(g_i^-)$, as all these sets are disjoint. There is also at most one $s = g_j$ for which $x \in V(g_j^+)$, again by disjointness. If $s = g_k$ is not one of these two, we claim that $(x, sy)_o \leqslant K + \delta$. This will prove the result, since this implies

$$|\{s \in S, (x, sy)_o \leqslant K + \delta\}| \geqslant |S| - 2 = 2^m - 2 = |S|(1 - 2\cdot 2^{-m}) \geqslant (1-\eta)|S|.$$

As $x \notin V(g_k^+)$, we have $(x, g_k^+)_o < K$. As $y \notin V(g_k^-)$, we have $sy = g_k y \in V'(g_k^+)$, i.e., $(sy, g_k^+)_o \geqslant K + \delta$. By hyperbolicity, we obtain

$$K > (x, g_k^+)_o \geqslant \min((x, sy)_o, (sy, g_k^+)_o) - \delta.$$

Note that the hyperbolicity inequality (3-1), initially stated inside the space, remains true for the Gromov product at infinity as we have used inf in its definition (1-1). If the minimum were realized by $(sy, g_k^+)_o \geqslant K + \delta$, we would get $K > (K+\delta) - \delta$, a contradiction. Therefore, the minimum is realized by $(x, sy)_o$, yielding $K > (x, sy)_o - \delta$ as claimed. □

**Corollary 3.13.** *Let $\mu$ be a nonelementary discrete measure on the set of isometries of $X$. For all $\eta > 0$, there exists $C > 0$ such that, for all $D > 0$, there exist $M > 0$ and an $(\eta, C, D)$-Schottky set in the support of $\mu^M$.*

*Proof.* By definition of a nonelementary measure, one can find loxodromic elements $u_0$ and $v_0$ with disjoint fixed points in the support of $\mu^a$ and $\mu^b$ for some $a, b > 0$. Then $u = u_0^b$ and $v = v_0^a$ belong to the support of $\mu^{ab}$ and have disjoint fixed points. Applying Proposition 3.12, we obtain an $(\eta, C, D)$-Schottky set in the support of $\mu^{abn}$ as desired. □

## 4. Linear escape

In this section, we prove Theorem 1.1, in other words, the random walk on $X$ driven by a nonelementary measure escapes linearly towards infinity, with exponential bounds. We copy the proof of Section 2, replacing subtrees with chain-shadows in the definition of pivotal times, and generators with elements of a Schottky set. The reader who would prefer to use shadows instead of chain-shadows may do so for intuition, but should be warned that the argument will then barely fail (at a single place, the backtracking step in the proof of Lemma 4.8).

Like in Section 2, the main technical part is to understand what happens for walks of the form $w_0 s_1 w_1 \cdots w_{n-1} s_n w_n$, where the $w_i$ are fixed, while the $s_i$ are random and drawn from a Schottky set. This will be done in Section 4A, while the application to prove Theorem 1.1 is done in Section 4B.

**4A.** ***A simple model.*** In this section, we fix isometries $w_0, w_1, \dots$ of $X$, a constant $C_0 > 0$, and $S$ a $\left(\frac{1}{100}, C_0, D\right)$-Schottky set of isometries of $X$. We will assume that $D$ is large enough compared to $C_0$; for definiteness, $D \geqslant 20C_0 + 100\delta + 1$ will do. Let $\mu_S$ be the uniform measure on $S$. Let $a_i, b_i$ be i.i.d. random variables distributed like $\mu_S$ and set $s_i = a_i b_i$.

We form a random process on $X$ by composing the $w_i$ and $s_i$ and applying them to the basepoint $o$. Our goal is to understand the behavior of $y_{n+1}^- = w_0 s_1 w_1 \cdots s_n w_n \cdot o$ when $n$ tends to infinity. The main result of this subsection is the following proposition.

**Proposition 4.1.** *There exists a universal constant $\kappa > 0$, not depending on $X$, $S$, $C_0$, $D$, $\delta$, such that, for all $n$,*

$$\mathbb{P}(d(o, y_{n+1}^-) \leqslant \kappa n) \leqslant e^{-\kappa n}.$$

To prove this proposition, we will first describe an inductive construction in a deterministic setting. To a finite or infinite sequence $(w_0, s_1 = a_1 b_1, w_1, s_2 = a_2 b_2, \dots)$, where $a_i, b_i \in S$ and the decomposition of $s_i$ as $a_i b_i$ is part of the data of the sequence, we will associate a set of pivotal times. Then we will obtain estimates on the behavior of this construction in the random setting, where $a_i, b_i$ will be random, while the $w_i$ will still be fixed, as in the setting of Proposition 4.1.

We define

$$y_i^- = w_0 s_1 w_1 \cdots s_{i-1} w_{i-1} \cdot o, \quad y_i = w_0 s_1 w_1 \cdots w_{i-1} a_i \cdot o,$$
$$y_i^+ = w_0 s_1 w_1 \cdots w_{i-1} a_i b_i \cdot o,$$

the three points visited during the transition around $i$. We have $d(y_i^-, y_i) = d(o, a_i \cdot o) \geqslant D$ as $a_i$ belongs to the $\left(\frac{1}{100}, C_0, D\right)$-Schottky set $S$. In the same way, $d(y_i, y_i^+) \geqslant D$. A difficulty that we will need to handle is that $d(y_i^+, y_{i+1}^-)$ may be short, as there is no lower bound on $w_i$, while we need long jumps everywhere to apply the results on chains of Section 3B.

We will define a sequence of pivotal times $P_n \subseteq \{1, \dots, n\}$, evolving with time: when going from $n-1$ to $n$, we will either add a pivotal time at time $n$ (so that

$P_n = P_{n-1} \cup \{n\}$, if the walk is going more towards infinity), or we will remove a few pivotal times at the end because the walk has backtracked (in this case, $P_n = P_{n-1} \cap \{1, \dots, m\}$ for some $m$).

Let us define inductively the pivotal times, starting from $P_0 = \varnothing$. Assume that $P_{n-1}$ is defined, and let us define $P_n$. Let $k = k(n)$ be the last pivotal time before $n$, i.e., $k = \max(P_{n-1})$. If $P_{n-1} = \varnothing$, take $k = 0$ and let $y_k = o$ — we will essentially ignore the minor adjustments to be made in this special case in the forthcoming discussion. Let us say that the local geodesic condition is satisfied at time $n$ if

$$(y_k, y_n)_{y_n^-} \leqslant C_0, \quad (y_n^-, y_n^+)_{y_n} \leqslant C_0 \quad \text{and} \quad (y_n, y_{n+1}^-)_{y_n^+} \leqslant C_0. \tag{4-1}$$

In other words, the points $y_k, y_n^-, y_n, y_n^+, y_{n+1}^-$ follow each other successively, with a $C_0$-alignment condition. As the points are well separated by the definition of Schottky sets, this will guarantee that we have a chain, progressing in a definite direction. We stress that $P_n$ only depends on the walk up to time $n$, that is on $w_0, s_1, \dots, w_{n-1}, s_n, w_n$.

If the local geodesic condition is satisfied at time $n$, then we say that $n$ is a pivotal time, and we set $P_n = P_{n-1} \cup \{n\}$. Otherwise, we backtrack to the largest pivotal time $m \in P_{n-1}$ for which $y_{n+1}^-$ belongs to the $(C_0 + \delta)$-chain-shadow of $y_m^+$ seen from $y_m$. In this case, we erase all later pivotal times, so we set $P_n = P_{n-1} \cap \{1, \dots, m\}$. If there is no such pivotal time $m$, we set $P_n = \varnothing$.

**Lemma 4.2.** *Assume that $P_n$ is nonempty. Let $m$ be its maximum. Then $y_{n+1}^-$ belongs to the $(C_0 + \delta)$-chain-shadow of $y_m^+$ seen from $y_m$.*

*Proof.* If $P_n$ has been defined from $P_{n-1}$ by backtracking, then the conclusion of the lemma is a direct consequence of the definition. Otherwise, the last pivotal time is $n$. In this case, let us show that $y_{n+1}^-$ belongs to the $(C_0 + \delta)$-chain-shadow of $y_n^+$ seen from $y_n$, by considering the chain $y_n, y_{n+1}^-$. By definition of the chain-shadow, we should check that $(y_n, y_{n+1}^-)_{y_n^+} \leqslant C_0 + \delta$ and $d(y_n, y_{n+1}^-) \geqslant 2C_0 + 4\delta + 1$. The first inequality is obvious as $(y_n, y_{n+1}^-)_{y_n^+} \leqslant C_0 \leqslant C_0 + \delta$ by the local geodesic condition (4-1). Moreover, since $(y_n, y_{n+1}^-)_{y_n^+} \leqslant C_0$ by (4-1), Lemma 3.4 gives $d(y_n, y_{n+1}^-) \geqslant d(y_n, y_n^+) - C_0 \geqslant D - C_0$, which is greater than or equal to $2C_0 + 4\delta + 1$ if $D$ is large enough. □

**Lemma 4.3.** *Let $P_n = \{k_1 < \cdots < k_p\}$. The sequence $y_{k_1}^-, y_{k_1}, y_{k_2}^-, y_{k_2}, \dots, y_{k_p}, y_{n+1}^-$ is a $(2C_0 + 3\delta, D - 2C_0 - 3\delta)$-chain.*

*Proof.* Let us first check the condition on Gromov products. We have to show that $(y_{k_{i-1}}, y_{k_i})_{y_{k_i}^-} \leqslant 2C_0 + 3\delta$ and $(y_{k_i}^-, y_{k_{i+1}}^-)_{y_{k_i}} \leqslant 2C_0 + 3\delta$. The first inequality is obvious, as it follows from the first property in the local geodesic condition when introducing the pivotal time $k_i$. Let us show the second one. Lemma 4.2 applied to the time $k_{i+1} - 1$ shows that $y_{k_{i+1}}^-$ belongs to the $(C_0 + \delta)$-chain-shadow of $y_{k_i}^+$ seen from $y_{k_i}$. Lemma 3.10 thus yields $(y_{k_{i+1}}, y_{k_i})_{y_{k_i}^+} \leqslant 2C_0 + 3\delta$. Moreover,

$(y_{k_i}^-, y_{k_i}^+)_{y_{k_i}} \leqslant C_0$ by the local geodesic condition when introducing the pivotal time $k_i$. We apply Lemma 3.5 with the points $y_{k_i}^-, y_{k_i}, y_{k_i}^+, y_{k_{i+1}}^-$, with $C = 2C_0 + 2\delta$. As $d(y_{k_i}, y_{k_i}^+) \geqslant D$ is large enough, this lemma applies and gives $(y_{k_i}^-, y_{k_{i+1}}^-)_{y_{k_i}} \leqslant 2C_0 + 3\delta$. This is the desired inequality.

Let us check the condition on distances. We have to show that $d(y_{k_i}^-, y_{k_i}) \geqslant D - 2C_0 - 3\delta$ and $d(y_{k_i}, y_{k_{i+1}}^-) \geqslant D - 2C_0 - 3\delta$. The first condition is obvious as $d(y_{k_i}^-, y_{k_i}) \geqslant D$. For the second, Lemma 3.10 gives $d(y_{k_i}, y_{k_{i+1}}^-) \geqslant d(y_{k_i}, y_{k_i}^+) - 2C_0 - 3\delta \geqslant D - 2C_0 - 3\delta$. □

The first point in the previous chain can be replaced with $o$:

**Lemma 4.4.** *Let $P_n = \{k_1 < \cdots < k_p\}$. The sequence $o, y_{k_1}, y_{k_2}^-, y_{k_2}, \ldots, y_{k_p}, y_{n+1}^-$ is a $(2C_0 + 4\delta, D - 2C_0 - 3\delta)$-chain.*

*Proof.* We have to control $d(o, y_{k_1})$ and $(o, y_{k_2}^-)_{y_{k_1}}$ as the other quantities are controlled by Lemma 4.3. For this, we will apply Lemma 3.5 to the points $y_{k_2}^-, y_{k_1}, y_{k_1}^-, o$ with $C = 2C_0 + 3\delta$. We have $(y_{k_2}^-, y_{k_1}^-)_{y_{k_1}} \leqslant 2C_0 + 3\delta$ by Lemma 4.3, and $(y_{k_1}, o)_{y_{k_1}^-} \leqslant C_0$, this being the first property in the local geodesic condition when introducing the pivotal time $k_1$, and $d(y_{k_1}, y_{k_1}^-) \geqslant D \geqslant 2C + \delta + 1$. Therefore, Lemma 3.5 gives $(y_{k_2}^-, o)_{y_{k_1}} \leqslant 2C_0 + 4\delta$. Moreover, Lemma 3.4 gives

$$d(y_{k_1}, o) \geqslant d(y_{k_1}, y_{k_1}^-) - (y_{k_1}, o)_{y_{k_1}^-} \geqslant D - C_0 \geqslant D - 2C_0 - 3\delta. \qquad \square$$

**Proposition 4.5.** *We have $d(o, y_{n+1}^-) \geqslant |P_n|$.*

*Proof.* This follows from Lemma 4.4, saying that we have a chain of length at least $|P_n|$ between $o$ and $y_{n+1}^-$, and from Lemma 3.7, saying that the distance grows linearly along a chain. □

This proposition shows that, to obtain the linear escape rate with exponential decay, it suffices to show that there are linearly many pivotal times. We note that, in the above deterministic construction, the set of pivotal times $P_n$ associated to a sequence $(s_i = a_i b_i)_{i \in \mathbb{N}}$ only depends on $s_i$ for $i \leqslant n$. Let us now turn to estimates in the random setting.

**Lemma 4.6.** *Fix $s_1, \ldots, s_n$, and draw $s_{n+1}$ according to $\mu_S^2$. The probability that $|P_{n+1}| = |P_n| + 1$ (i.e., that $n+1$ gets added as a pivotal time) is at least $\frac{9}{10}$.*

*Proof.* In the local geodesic condition (4-1), the last property reads $(g \cdot o, g b_{n+1} w_{n+1} \cdot o)_{g b_{n+1} \cdot o} \leqslant C_0$ for $g = w_0 s_1 \cdots w_n a_{n+1}$. Composing with $b_{n+1}^{-1} g^{-1}$, it becomes $(b_{n+1}^{-1} \cdot o, w_{n+1} \cdot o)_o \leqslant C_0$. By the definition of a Schottky set, this inequality is satisfied with probability at least $1 - \eta = \frac{99}{100}$ when choosing $b_{n+1}$. Once $b_{n+1}$ is fixed, the other two properties in the local geodesic condition only depend on $a_{n+1}$, and each of them is satisfied with probability at least $\frac{99}{100}$, again by the Schottky property. They are satisfied simultaneously with probability at least $\frac{98}{100}$. As $\frac{99}{100} \cdot \frac{98}{100} \geqslant \frac{9}{10}$, this concludes the proof. □

The key point is to control the backtracking length. For this, we will see that for one configuration that backtracks a lot, there are many configurations that do not. Given $\bar{s} = (s_1, \ldots, s_n)$, let us say that another sequence $\bar{s}' = (s_1', \ldots, s_n')$ is pivoted from $\bar{s}$ if they have the same pivotal times, $b_k' = b_k$ for all $k$, and $a_k' = a_k$ when $k$ is not a pivotal time.

**Lemma 4.7.** *Let $i$ be a pivotal time of $\bar{s} = (s_1, \ldots, s_n)$. Replace $s_i = a_i b_i$ with $s_i' = a_i' b_i$ which still satisfies the local geodesic condition (4-1) with $n$ replaced by $i$. Then $(s_1, \ldots, s_i', \ldots, s_n)$ is pivoted from $\bar{s}$.*

*Proof.* We need to show that the pivotal times of $\bar{s}'$ are the same as those of $\bar{s}$. Until time $i$, the sequences are the same, hence they have the same pivotal times: $P_{i-1}(\bar{s}) = P_{i-1}(\bar{s}')$. Then $i$ is added as a pivotal time for both $\bar{s}$ and $\bar{s}'$ by assumption, therefore $P_i(\bar{s}) = P_i(\bar{s}')$. Then the remaining part of the trajectory for $\bar{s}$ never backtracks beyond $i$, as $i$ remains a pivotal time. This backtracking property is defined in terms of the relative position of the trajectory compared to $y_i$ and $y_i^+$, and therefore it depends on $b_i$ but not on the beginning of the trajectory (and in particular it does not depend on $a_i$). Hence, replacing $a_i$ with $a_i'$ does not change the backtrackings, which are the same for $\bar{s}$ and $\bar{s}'$ until time $n$. □

Lemma 4.7 shows that, if a trajectory has $p$ pivotal times, then it has a lot of pivoted trajectories (exponentially many in $p$) as one can change $a_i$ to $a_i'$ at each pivotal time. Denote by $\mathcal{E}_n(\bar{s})$ the set of trajectories which are pivoted from $\bar{s}$. Conditionally on $\mathcal{E}_n(\bar{s})$, the random variables $a_i'$ for $i$ a pivotal time are independent, but not identically distributed, as they are each drawn from a subset of $S$ depending on $i$, of large cardinality.

**Lemma 4.8.** *Let $\bar{s} = (s_1, \ldots, s_n)$ be a trajectory with $q$ pivotal times. We condition on $\mathcal{E}_n(\bar{s})$, and we draw $s_{n+1}$ according to $\mu_S^2$. Then, for all $j \geqslant 0$,*

$$\mathbb{P}(|P_{n+1}| < q - j \mid \mathcal{E}_n(\bar{s})) \leqslant \left(\tfrac{1}{10}\right)^{j+1}.$$

*Proof.* If $q = 0$, then there is nothing to prove. Assume $q > 0$.

First, the probability that $s_{n+1}$ creates a new pivotal time is at least $\frac{9}{10}$, by Lemma 4.6 (and the elements $s_{n+1}$ that create a new pivotal time are the same over the whole equivalence class $\mathcal{E}_n(\bar{s})$ as $q > 0$). Let us now fix a bad $s_{n+1}$, giving rise to backtracking.

We will first show the lemma for $j = 1$. Let $m < k$ be the last two pivotal times. We have to show that

$$\mathbb{P}(|P_{n+1}| < q - 1 \mid \mathcal{E}_n(\bar{s}), s_{n+1}) \leqslant \tfrac{1}{10}, \tag{4-2}$$

i.e., most trajectories do not backtrack beyond $k$: for many choices of $a_k$, then $y_{n+1}^-$ should belong to the $(C_0 + \delta)$-chain-shadow of $y_m^+$ seen from $y_m$. By Lemma 4.2

applied at time $k-1$, we already know that $y_k^-$ belongs to this set. Therefore, there exists a chain $x_0 = y_m, x_1, \dots, x_i = y_k^-$ pointing in the chain-shadow. With a good choice of $a_k$, we will increase the chain by adding $y_{n+1}^-$ at its end.

Let us consider $a'_k$ so that the points $x_{i-1}, y_k^-, y_k, y_{n+1}^-$ are $C_0$-aligned, so that $(x_{i-1}, y_k)_{y_k^-} \leqslant C_0$ and $(y_k^-, y_{n+1}^-)_{y_k} \leqslant C_0$. By the Schottky property, there are at least $\frac{98}{100}|S|$ such $a'_k$. We show that, with this choice, $y_{n+1}^-$ belongs to the chain-shadow of $y_m^+$ seen from $y_m$, and therefore backtracking stops here. For this, it is enough to see that $x_0, \dots, x_{i-1}, y_k^-, y_{n+1}^-$ is a $(C_0+\delta, 2C_0+4\delta+1)$-chain. We have to see that $d(y_k^-, y_{n+1}^-) \geqslant 2C_0+4\delta+1$ and $(x_{i-1}, y_{n+1}^-)_{y_k^-} \leqslant C_0+\delta$. For this, apply Lemma 3.5 to the points $x_{i-1}, y_k^-, y_k, y_{n+1}^-$, which are $C_0$-aligned. As $d(y_k^-, y_k) \geqslant D$ is large enough, this lemma gives $(x_{i-1}, y_{n+1}^-)_{y_k^-} \leqslant C_0+\delta$. Moreover, Lemma 3.4 gives $d(y_k^-, y_{n+1}^-) \geqslant d(y_k^-, y_k) - (y_k^-, y_{n+1}^-)_{y_k} \geqslant D - C_0 \geqslant 2C_0+4\delta+1$, as claimed.

In the equivalence class, the number of possible choices for $a'_k$ when introducing the pivotal time $k$ is at least $\frac{98}{100}|S|$, since most choices satisfy the local geodesic condition (see the proof of Lemma 4.6). The number of choices of $a'_k$ that ensure there is no further backtracking is also bounded below by $\frac{98}{100}|S|$, by the previous discussion, so that the number of bad choices is at most $\bigl(1-\frac{98}{100}\bigr)|S|$. Finally, the proportion of bad choices that lead to further backtracking is at most

$$\frac{\bigl(1-\frac{98}{100}\bigr)|S|}{\frac{98}{100}|S|} < \frac{1}{10}.$$

This proves (4-2) for $j=1$.

To prove the lemma for $j=2$, let us fix $s_{n+1}$ as well as a bad choice of $a'_k$ that gives rise to backtracking beyond $k$, which happens with probability at most $\frac{1}{10}$. We have to show that, once these quantities are fixed, the probability to backtrack past the previous pivotal time is at most $\frac{1}{10}$. This is the same argument as above. The case of general $j$ is proved analogously by induction. □

**Lemma 4.9.** *Let $A_n = |P_n|$ be the number of pivotal times. Then, in distribution, $A_{n+1} \geqslant A_n + U$ where $U$ is a random variable independent from $A_n$ distributed as*

$$\mathbb{P}(U=-j) = 9\bigl(\tfrac{1}{10}\bigr)^{j+1} \text{ for } j>0, \quad \mathbb{P}(U=0)=0 \quad \text{and} \quad \mathbb{P}(U=1) = \tfrac{9}{10}.$$

*In other words, $\mathbb{P}(A_{n+1} \geqslant i) \geqslant \mathbb{P}(A_n + U \geqslant i)$ for all $i$.*

*Proof.* Conditionally on $\mathcal{E}_n(\bar{s})$, this follows from Lemma 4.8, just like in the proof of Proposition 2.6: one shows that

$$\mathbb{P}(A_{n+1} \geqslant i \mid \mathcal{E}_n(\bar{s})) \geqslant \mathbb{P}(A_n + U \geqslant i \mid \mathcal{E}_n(\bar{s})).$$

As the inequality is uniform over the conditioning, the unconditioned version follows. □

**Proposition 4.10.** *There exists a universal constant $\kappa > 0$ such that, for all $n$,*

$$\mathbb{P}(|P_n| \leqslant \kappa n) \leqslant e^{-\kappa n}.$$

*Proof.* Let $U_1, U_2, \dots$ be a sequence of independent copies of the variable $U$ from Lemma 4.9. Iterating this lemma gives

$$\mathbb{P}(|P_n| \geqslant i) \geqslant \mathbb{P}(U_1 + \dots + U_n \geqslant i)$$

for all $i$. In particular, $\mathbb{P}(|P_n| \leqslant \kappa n) \leqslant \mathbb{P}(U_1 + \dots + U_n \leqslant \kappa n)$. As the $U_i$ are real random variables with an exponential moment and positive expectation, $\mathbb{P}(U_1 + \dots + U_n \leqslant \kappa n)$ is exponentially small if $\kappa$ is small enough. □

*Proof of Proposition 4.1.* The linear escape with exponential error term follows from Proposition 4.5 giving $d(o, y_{n+1}^-) \geqslant |P_n|$, and from Proposition 4.10 ensuring that $|P_n|$ grows linearly outside of a set of exponentially small probability. □

**4B.** ***Proof of linear escape and convergence at infinity.*** Let $\mu$ be a nonelementary measure on the set of isometries of the space $X$. In this subsection, we prove Theorem 1.1: the $\mu$-random walk goes to infinity linearly, with an exponential error term. The techniques we develop along the way will also prove convergence of the walk at infinity.

We apply Corollary 3.13 with $\eta = \frac{1}{100}$. Let $C = C_0$ be given by this corollary. Choose $D = D(C_0, \delta)$ large enough, so that the results of Section 4A apply ($D = 20C_0 + 100\delta + 1$ suffices). The corollary gives an $(\eta, C_0, D)$ Schottky set $S$ included in the support of $\mu^M$ for some $M$. For $\alpha > 0$ small enough and $N = 2M$, we may write $\mu^N = \alpha \mu_S^2 + (1-\alpha)\nu$ for some probability measure $\nu$, where $\mu_S$ is the uniform measure on $S$.

As in [Boulanger et al. 2021, Section 6], let us reconstruct in a slightly indirect way the random walk, as follows, on a space $\Omega$ containing Bernoulli random variables $\varepsilon_i$ satisfying $\mathbb{P}(\varepsilon_i = 1) = \alpha$ and $\mathbb{P}(\varepsilon_i = 0) = 1-\alpha$, variables $h_i$ distributed according to $\nu$ and variables $a_i$, $b_i$ distributed according to $\mu_S$, all independent. Define $\gamma_i = s_i := a_i b_i$ if $\varepsilon_i = 1$, and $\gamma_i = h_i$ if $\varepsilon_i = 0$. Then the $\gamma_i$ are independent random variables on $\Omega$, and each of them is distributed as the product of $N$ independent random variables with distribution $\mu$. In particular, $\gamma_0 \cdots \gamma_{n-1}$ is distributed like $Z_{Nn}$. Extending $\Omega$ if necessary, we can also construct on $\Omega$ a sequence of independent random variables $g_0, g_1, \dots$ with distribution $\mu$ such that $\gamma_i = g_{iN} \cdots g_{iN+N-1}$.

Let us give more details on this construction. Let $\Omega'$ be another probability space containing i.i.d. random variables $g_0, g_1, \dots$ distributed according to $\mu$. Let $\gamma_i' = g_{iN} \cdots g_{iN+N-1}$ on $\Omega'$. Then the random variables $(\gamma_i)_{i\in\mathbb{N}}$ on $\Omega$ and $(\gamma_i')_{i\in\mathbb{N}}$ on $\Omega'$ have the same distribution. Then a standard coupling argument (see, e.g., [Berkes and Philipp 1979, Lemma A.1]) ensures that it is possible to realize all the random variables $\varepsilon_i, h_i, a_i, b_i$ and $g_i$ on a common probability space such that $\gamma_i = \gamma_i'$. This entails the equality $\gamma_i = g_{iN} \cdots g_{iN+N-1}$ as claimed.

Let $t_1 < t_2 < \cdots$ be the times where $\varepsilon_i = 1$. Fix $n \in \mathbb{N}$. We let $\tau = \tau(n)$ be the last index $j$ such that $N(t_j + 1) \leqslant n$, so that the interval $[Nt_j, N(t_j + 1))$ is contained in $[0, n)$. We will decompose the product $g_0 \cdots g_{n-1}$ as a product of the elements $s_j' = s_{t_j} = a_{t_j} b_{t_j} = a_j' b_j'$, the product of all $g_i$ for $i \in [Nt_j, N(t_j + 1))$, interspersed with other words that we will consider as fixed, in order to be in the framework of Section 4A. Let $w_j = g_{N(t_j+1)} \cdots g_{Nt_{j+1}-1}$, where by convention $t_0 = -1$, and let $w' = w'(n) = g_{N(t_{\tau(n)}+1)} \cdots g_{n-1}$ be the last missing word, which really depends on $n$, contrary to the previous words that just fill the gaps between blocks corresponding to $\varepsilon_i = 1$. By construction,

$$Z_n \cdot o = w_0 s_1' w_1 \cdots w_{\tau(n)-1} s_{\tau(n)}' w'(n) \cdot o.$$

We can associate to the sequence $(w_0, s_1' = a_1' b_1', w_1, s_2' = a_2' b_2', \dots, w'(n))$ in this decomposition a sequence of pivotal times $P_1^{(n)}, \dots, P_{\tau(n)}^{(n)}$, as in Section 4A, where the exponent $(n)$ is here to emphasize that all this is done for a fixed $n$, and that the objects we introduce may therefore depend on $n$. In fact, the words $w_j$ for $j < \tau(n)$ only depend on $j$ as they are given by $w_j = g_{(N+1)t_j} \cdots g_{Nt_{j+1}-1}$, which does not involve $n$. Hence, the sequence of inductively constructed pivotal times is rather

$$P_1, P_2, \dots, P_{\tau(n)-1}, P_{\tau(n)}^{(n)}. \tag{4-3}$$

The main quantity we will control is

$$u_n := |P_{\tau(n)}^{(n)}|,$$

the final number of pivotal times after $n$ steps of the initial random walk.

**Proposition 4.11.** *There exists $\kappa > 0$ such that $\mathbb{P}(u_n \leqslant \kappa n) \leqslant e^{-\kappa n}$.*

*Proof.* The sequence $t_{j+1} - t_j$ is a sequence of independent random variables with an exponential tail. Therefore, there exist $C > 0$ and $\kappa > 0$ such that

$$\mathbb{P}(t_j - t_0 \geqslant Cj) = \mathbb{P}\left( \sum_{i=1}^{j-1} (t_{i+1} - t_i) \geqslant Cj \right) \leqslant e^{-\kappa j}.$$

Hence, if $\beta > 0$ is small enough, we have $N(t_{\lfloor \beta n \rfloor} + 1) \leqslant n$ outside of a set with exponentially small probability. This gives

$$\mathbb{P}(\tau(n) < \beta n) \leqslant e^{-\kappa n}$$

for some $\kappa > 0$. For any $c > 0$, we get

$$\mathbb{P}(u_n \leqslant cn) \leqslant e^{-\kappa n} + \mathbb{P}(u_n \leqslant cn, \tau \geqslant \beta n).$$

Let us concentrate on the second set. We condition with respect to the $\varepsilon_i$ (which fixes the $t_i$ and $\tau$) and with respect to the $g_i$ outside of the intervals $[Nt_j, N(t_j+1))$ (which fixes the $w_j$ and $w'$). Once these are fixed, we are in the framework of Section 4A. We may therefore apply Proposition 4.10 and deduce that, conditionally on these quantities, we have

$$\mathbb{P}(u_n \leqslant c\tau \mid (\varepsilon_i)_i, (g_i)_{i \notin \bigcup_j [Nt_j, N(t_j+1))}) \leqslant e^{-c\tau},$$

for some $c > 0$. When $\tau \geqslant \beta n$, the left-hand side bounds the conditional probability that $u_n \leqslant c\beta n$, and the right-hand side is bounded by $e^{-c\beta n}$. As this is uniform on the conditioning, this implies that $\mathbb{P}(u_n \leqslant c\beta n, \tau \geqslant \beta n) \leqslant e^{-c\beta n}$, concluding the proof. □

*Proof of Theorem 1.1.* Outside of a set with exponentially small probability, the number of pivotal times at the $n$-th step of the random walk is at least $\kappa n$ for some $\kappa > 0$, by Proposition 4.11. As the distance to the origin is bounded below by the number of pivotal times, by Proposition 4.5, this concludes the proof. □

This argument enables us to recover a theorem of [Maher and Tiozzo 2018], the convergence of the walk at infinity. We even get exponential error terms in the speed of convergence. We start with a lemma ensuring that positions of the random walk stay in a shadow.

**Lemma 4.12.** *Let $n \in \mathbb{N}$ and $C > 0$. Assume that, for all $k \geqslant n$, one has $u_k > C$. Let $x$ be the position of the walk at the $C$-th pivotal time in $P^{(n)}_{\tau(n)}$. Then, for all $k \geqslant n$, the point $Z_k \cdot o$ belongs to the $(2C_0 + 6\delta)$-shadow of $x$ seen from $o$.*

*Proof.* For $k \geqslant n$, the set $P^{(k)}_{\tau(k)}$ has strictly more than $C$ points by assumption. In particular, the $C$-th pivotal time is not introduced at the last step, and the last step does not backtrack beyond this point. The set of pivotal times before the last index does not depend on $k$, as explained before (4-3). It follows that the $C$-th pivotal time in $P^{(k)}_{\tau(k)}$ is independent of $k \geqslant n$. In particular, $x$ is the position of the walk at a pivotal time in $P^{(k)}_{\tau(k)}$, for any $k \geqslant n$.

For $k \geqslant n$, Lemma 4.4 shows that there is a $(2C_0+4\delta, D-2C_0-3\delta)$-chain from $o$ to $Z_k \cdot o$ going through $x$. By Lemma 3.8, we deduce that $(o, Z_k \cdot o)_x \leqslant 2C_0+6\delta$. In other words, all the points $Z_k \cdot o$ remain in the $(2C_0+6\delta)$-shadow of $x$ seen from $o$, as claimed. □

**Proposition 4.13.** *Almost surely, there is a point $Z_\infty \in \partial X$ such that $Z_n \cdot o$ converges to $Z_\infty$. Moreover, there exists $\kappa > 0$ such that for all $n$,*

$$\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant \kappa n) \leqslant e^{-\kappa n}. \tag{4-4}$$

*Proof.* Fix $c > 0$ such that $\mathbb{P}(u_n \leqslant cn) \leqslant e^{-cn}$ for all $n$, by Proposition 4.11. Since $\mathbb{P}(u_n \leqslant cn)$ is exponentially small, Borel–Cantelli ensures that almost surely one has eventually $u_n > cn$. Lemma 4.12 then applies, with $C = \lfloor cn \rfloor - 1$. Let $x_n$ denote the position of the walk at the $(\lfloor cn \rfloor - 1)$-th pivotal time in $P^{(n)}_{\tau(n)}$ for large $n$. By Proposition 4.5, it satisfies

$$d(o, x_n) \geqslant \lfloor cn \rfloor - 1. \tag{4-5}$$

The sequence $Z_k \cdot o$ is eventually trapped in the shadow of $x_n$ seen from $o$ by Lemma 4.12. This implies the convergence at infinity of $Z_k \cdot o$, by Lemma 3.2.

Finally, let us show the quantitative estimate (4-4). Assume that for all $k \geqslant n$, one has $u_k > ck$, which happens with probability at least $1 - Ce^{-cn}$. In this case, all the points $Z_k \cdot o$ for $k \geqslant n$ belong to the $(2C_0 + 6\delta)$-shadow of $x_n$. Therefore, Lemma 3.3 applies and gives

$$(Z_n \cdot o, Z_\infty)_o \geqslant d(o, x_n) - (2C_0 + 6\delta) - 3\delta. \tag{4-6}$$

Together with (4-5), this gives a linear lower bound for the Gromov product, that holds outside of an exponentially small set. □

We will also need the following lemma, that follows from the same techniques.

**Lemma 4.14.** *Let $\mu$ be a nonelementary discrete measure on the set of isometries of a Gromov-hyperbolic space $X$ with basepoint $o$. Let $Z_n = g_0 \cdots g_{n-1}$ where the $g_i$ are i.i.d. with distribution $\mu$. Let $\varepsilon > 0$. There exists $C > 0$ such that, for any isometry $g$,*

$$\mathbb{P}(\forall n, d(o, gZ_n \cdot o) \geqslant d(o, g \cdot o) - C) \geqslant 1 - \varepsilon.$$

The point of the lemma is that the possible loss $C$ is uniform in $g$. Without moment assumptions on $\mu$, it is not possible to get a better bound, contrary to the case of walks with an exponential moment; compare with [Boulanger et al. 2021, Theorem 2.6].

*Proof.* We follow the same construction as at the beginning of this subsection to reconstruct the random walk, but adding the isometry $g$ before the first step of the random walk. Since the estimates of Section 4A are uniform in $w_0$, replacing $w_0$ with $gw_0$ does not change them. Therefore, the number $u_n := |P^{(n)}_{\tau(n)}|$ of pivotal times for the random walk at time $n$ still satisfies the estimate of Proposition 4.11: there exists $\kappa > 0$ (independent of $g$) such that $\mathbb{P}(u_n \leqslant \kappa n) \leqslant e^{-\kappa n}$.

Let us fix $n$ such that $\sum_{i \geqslant n} e^{-\kappa i} < \varepsilon/2$. On a set $A_g$ of probability at least $1 - \varepsilon/2$, which may depend on $g$, one has for all $i \geqslant n$ the inequality $u_i > \kappa i \geqslant \kappa n$. As in the proof of Proposition 4.13, one can then find a point $x_n$ such that, for all $i \geqslant n$, the points $gZ_i \cdot o$ belong to the $(2C_0 + 6\delta)$-shadow of $x_n$ seen from $o$. In particular, by Lemma 3.1,

$$d(gZ_i \cdot o, o) \geqslant d(o, x_n) - 2C_0 - 6\delta.$$

Moreover, $x_n$ is of the form $gZ_k \cdot o$ for some $k \leqslant n$.

By measurability, we can find a set $A$ of measure at least $1-\varepsilon/2$ and a constant $C$ (both independent of $g$) such that, for all $\omega \in A$ and all $k \leqslant n$, holds $d(o, Z_k \cdot o) \leqslant C$.

Consider $\omega \in A_g \cap A$ (this set has measure at least $1-\varepsilon$). Then

$$d(o, x_n) = d(o, gZ_k \cdot o) \geqslant d(o, g \cdot o) - d(g \cdot o, gZ_k \cdot o) = d(o, g \cdot o) - d(o, Z_k \cdot o) \geqslant d(o, g \cdot o) - C.$$

For all $i \geqslant n$, we get $d(gZ_i \cdot o, o) \geqslant d(o, g \cdot o) - C - 2C_0 - 6\delta$. For $i < n$, this estimate also holds as $d(o, Z_i \cdot o) \leqslant C$. This proves the lemma, for the constant $C + 2C_0 + 6\delta$ which is independent of $g$. □

## 5. Precise estimates

**5A.** ***A more complicated model.*** To obtain precise estimates on the rate of convergence to infinity, we will need to compare the distance to the origin with the sum of independent real valued random variables corresponding to the size of jumps of the random walk. This is done in the next proposition.

**Proposition 5.1.** *For $\eta \in \big(0, \frac{1}{100}\big]$, there exists $\kappa = \kappa(\eta) > 0$ with the following property.*

*Let $S$ be an $(\eta, C_0, D)$-Schottky set of isometries of a $\delta$-hyperbolic space $X$ with basepoint $o$, where $D$ is large enough compared to $C_0$; for definiteness, $D \geqslant 20C_0 + 100\delta + 1$ is enough. Let $\rho_1, \rho_2, \dots$ be discrete probability measures on the isometry set of $X$. Let $R$ be a nonnegative real random variable such that for all $i$ and all $M \geqslant 0$, one has*

$$\rho_i\{g : d(o, g \cdot o) \geqslant M\} \geqslant \mathbb{P}(R \geqslant M),$$

*i.e., for all $i$ the distance with respect to the origin for $\rho_i$ dominates stochastically $R$.*

*Let $w_0, w_1, \dots$ be fixed isometries of $X$. Let $s_1, s_2, \dots$ be independent random variables, where $s_i$ is sampled according to $\mu_S^2 * \rho_i * \mu_S^2$. Define $y_{n+1}^- = w_0 s_1 w_1 \cdots s_n w_n \cdot o$. Then for all $M \geqslant 0$,*

$$\mathbb{P}(d(o, y_{n+1}^-) \leqslant M) \leqslant \mathbb{P}(R_1 + \cdots + R_{\lfloor (1-22\eta)n \rfloor} \leqslant M) + e^{-\kappa n},$$

*where $R_1, R_2, \dots$ are independent copies of $R$.*

When all the $\rho_i$ are the Dirac mass at the origin, then the setting of the proposition is essentially the same as the simple model of Section 4A, except that we are sampling the $s_i$ according to $\mu_S^4$ instead of $\mu_S^2$, which does not really make a difference. The conclusion in the general setting of Proposition 5.1 is that the growth rate of the distance to the origin is at least the growth rate of sums of i.i.d. random variables distributed like the $\rho_i$, up to a minor loss that tends to 0 when the proportion $\eta$ of bad elements in the Schottky set tends to 0 and an exponentially small error term. This model will be precise enough to capture the right growth rate of a general random walk, to prove Theorems 1.2 and 1.3 in the next paragraphs, in the same way that we have deduced linear escape with exponential estimates from the results on the simple model of Section 4A. The possibility to have different measures $\rho_i$ at the different jumps will be important in the application of this proposition in Section 5C, but for the proof the reader may pretend for simplicity that they are all equal to a fixed measure $\rho$, and then one can take $R$ to be the distribution of $d(o, g\cdot o)$ with respect to $\rho$.

To prove Proposition 5.1, let us introduce a refined notion of pivotal times, in which we will keep the randomness coming from the $\rho_i$. Write $s_i = a_i b_i r_i c_i d_i$, where $a_i, b_i, c_i, d_i$ are distributed according to $\mu_S$ while $r_i$ is distributed according to $\rho_i$. This gives rise to six successive points at the $i$-th transition:

$$\begin{aligned} y_i^{(0)} &= w_0 s_1 \cdots s_{i-1} w_{i-1}\cdot o = y_i^-, & y_i^{(3)} &= w_0 s_1 \cdots s_{i-1} w_{i-1} a_i b_i r_i \cdot o,\\ y_i^{(1)} &= w_0 s_1 \cdots s_{i-1} w_{i-1} a_i \cdot o, & y_i^{(4)} &= w_0 s_1 \cdots s_{i-1} w_{i-1} a_i b_i r_i c_i \cdot o = y_i,\\ y_i^{(2)} &= w_0 s_1 \cdots s_{i-1} w_{i-1} a_i b_i \cdot o, & y_i^{(5)} &= w_0 s_1 \cdots s_{i-1} w_{i-1} a_i b_i r_i c_i d_i \cdot o = y_i^+. \end{aligned}$$

The distances between two successive points in this list is at least $D$ as it comes from the application of an element of the Schottky set $S$, except for the distance between $y_i^{(2)}$ and $y_i^{(3)}$ for which we have no lower bound as $r_i$ is drawn according to $\rho_i$.

Let us define inductively a set of refined pivotal times associated in a deterministic way to a finite or infinite sequence $(w_0, s_1 = a_1 b_1 r_1 c_1 d_1, w_1, s_2 = a_2 b_2 r_2 c_2 d_2, w_2, \dots)$, that we will denote by $\overline{P}_n$ to differentiate it from the previous unrefined notion. We copy the definition of Section 4A. We start from $\overline{P}_0 = \varnothing$. Assume that $\overline{P}_{n-1}$ is defined, and let us define $\overline{P}_n$. Let $k = k(n)$ be the last pivotal time before $n$, i.e., $k = \max(\overline{P}_{n-1})$. (If $\overline{P}_{n-1} = \varnothing$, take $k = 0$ and let $y_k = o$.) Let us say that the local geodesic condition is satisfied at time $n$ if in the sequence $y_k, y_n^{(0)}, y_n^{(1)}, y_n^{(2)}, y_n^{(3)}, y_n^{(4)}, y_n^{(5)}, y_{n+1}^-$, all successive points are $C_0$-aligned, and moreover $y_n^{(1)}, y_n^{(3)}, y_n^{(4)}$ are $C_0$-aligned, the latter condition being useful to compensate the fact that the jump from $y_n^{(2)}$ to $y_n^{(3)}$ may be small, preventing us to apply the results on chains of Section 3B. If the local geodesic condition is satisfied at time $n$, then we say that $n$ is a refined pivotal time, and we set $\overline{P}_n = \overline{P}_{n-1} \cup \{n\}$. Otherwise, we backtrack to the largest refined pivotal time $m \in \overline{P}_{n-1}$ for which $y_{n+1}^-$ belongs to the $(C_0+\delta)$-chain-shadow of $y_m^+$ seen from $y_m$. In this case, we erase all later pivotal times, so we set $\overline{P}_n = \overline{P}_{n-1} \cap \{1, \dots, m\}$. If there is no such pivotal time $m$, we set $\overline{P}_n = \varnothing$.

For the refined notion, we can prove the analogues of the lemmas of Section 4A.

**Lemma 5.2.** *Assume that $\overline{P}_n$ is nonempty. Let $m$ be its maximum. Then $y_{n+1}^-$ belongs to the $(C_0+\delta)$-chain-shadow of $y_m^+$ seen from $y_m$.*

*Proof.* The proof is exactly the same as for Lemma 4.2: when there is backtracking, this follows from the definition, and when there is no backtracking (the last pivotal time is $n$), then the chain $y_n, y_{n+1}^-$ satisfies all the properties to show that $y_{n+1}^-$ is in the chain-shadow. □

**Lemma 5.3.** *Let $\overline{P}_n = \{k_1 < \cdots < k_p\}$. The sequence $y_{k_1}^-, y_{k_1}, y_{k_2}^-, y_{k_2}, \dots, y_{k_p}, y_{n+1}^-$ is a $(2C_0 + 3\delta, D - 2C_0 - 3\delta)$-chain. Moreover, $d(y_{k_i}^-, y_{k_i}) \geqslant d(o, r_{k_i} \cdot o) + D$ for all $i$.*

*Proof.* This differs a little bit from the proof of Lemma 4.3 as there are more points involved at each pivotal time. It is still basic chain manipulations, with the only difficulty that the jumps corresponding to $r_i$ and $w_i$ may be short, but since they are surrounded by big jumps with controlled alignment conditions this can be circumvented easily.

By definition, the points $y_{k_{i-1}}, y_{k_i}^-, y_{k_i}^{(1)}, y_{k_i}^{(2)}, y_{k_i}^{(3)}, y_{k_i}^{(4)}, y_{k_i}^{(5)}$ are $C_0$-aligned. However, the distances between $y_{k_{i-1}}$ and $y_{k_i}^-$ on the one hand, and between $y_{k_i}^{(2)}$ and $y_{k_i}^{(3)}$ on the other hand, are not obviously bounded below, contrary to the other distances, which are greater than or equal to $D$, so one can not apply the results on chains to these points. However, we can fix this by removing one point: we claim that

$$y_{k_{i-1}}, y_{k_i}^-, y_{k_i}^{(1)}, y_{k_i}^{(3)}, y_{k_i}^{(4)} (= y_{k_i}), y_{k_i}^{(5)} \text{ form a } (C_0 + \delta, D - 2C_0 - 3\delta)\text{-chain.} \quad (5\text{-}1)$$

Let us prove this claim. We may apply Lemma 3.5 to the points $y_{k_i}^-, y_{k_i}^{(1)}, y_{k_i}^{(2)}, y_{k_i}^{(3)}$, with $C = C_0$, to deduce that $(y_{k_i}^-, y_{k_i}^{(3)})_{y_{k_i}^{(1)}} \leqslant C_0 + \delta$. Moreover, Lemma 3.4 gives

$$d(y_{k_i}^{(1)}, y_{k_i}^{(3)}) \geqslant d(y_{k_i}^{(1)}, y_{k_i}^{(2)}) - (y_{k_i}^{(1)}, y_{k_i}^{(3)})_{y_{k_i}^{(2)}} \geqslant D - C_0,$$

and furthermore, $d(y_{k_{i-1}}, y_{k_i}^-) \geqslant D - 2C_0 - 3\delta$ by Lemma 3.10, as $y_{k_i}^-$ is in the $(C_0 + \delta)$-chain-shadow of $y_{k_{i-1}}^+$ seen from $y_{k_{i-1}}$, by Lemma 5.2. Finally, note that $(y_{k_i}^{(1)}, y_{k_i}^{(4)})_{y_{k_i}^{(3)}} \leqslant C_0$ by the last assumption in the local geodesic condition. We have checked all the nontrivial properties in (5-1), completing its proof.

We have in particular $d(y_{k_{i-1}}, y_{k_i}^-) \geqslant D - 2C_0 - 3\delta$, and also by (3-3),

$$\begin{aligned} d(y_{k_i}^-, y_{k_i}) &= d(y_{k_i}^-, y_{k_i}^{(4)}) \\ &\geqslant d(y_{k_i}^-, y_{k_i}^{(1)}) + d(y_{k_i}^{(1)}, y_{k_i}^{(3)}) + d(y_{k_i}^{(3)}, y_{k_i}^{(4)}) - 3(2C_0 + 4\delta). \end{aligned} \quad (5\text{-}2)$$

By Lemma 3.4 applied to $y_{k_i}^{(1)}, y_{k_i}^{(2)}, y_{k_i}^{(3)}$,

$$d(y_{k_i}^{(1)}, y_{k_i}^{(3)}) \geqslant d(y_{k_i}^{(2)}, y_{k_i}^{(3)}) - (y_{k_i}^{(1)}, y_{k_i}^{(3)})_{y_{k_i}^{(2)}} \geqslant d(o, r_{k_i} \cdot o) - C_0.$$

The two other distances in (5-2) are bounded below by $D$. Given that $D \geqslant 3(2C_0 + 4\delta) + C_0$, we obtain

$$d(y_{k_i}^-, y_{k_i}) \geqslant D + d(o, r_{k_i} \cdot o).$$

This proves all the distance conditions in the claim of the lemma.

Let us now check the estimates on the Gromov products. Applying Lemma 3.7 to the chain (5-1), we get $(y_{k_{i-1}}, y_{k_i})_{y^-_{k_i}} \leqslant C_0 + 2\delta \leqslant 2C_0 + 3\delta$, proving one of the desired estimates. The other one is $(y^-_{k_i}, y^-_{k_{i+1}})_{y_{k_i}} \leqslant 2C_0 + 3\delta$. To prove it, let us apply Lemma 3.5 to the points $y^-_{k_i}, y_{k_i}, y^+_{k_i}, y^-_{k_{i+1}}$. The Gromov product of the last three is at most $2C_0 + 3\delta$ by Lemmas 5.2 and 3.10, and the Gromov product of the first three is at most $C_0 + 2\delta$ by applying Lemma 3.7 to the reverse of the chain (5-1). Moreover, the distance $d(y_{k_i}, y^+_{k_i})$ is at least $D$, large enough. Therefore, Lemma 3.5 indeed applies with $C = 2C_0 + 2\delta$, and gives $(y^-_{k_i}, y^-_{k_{i+1}})_{y_{k_i}} \leqslant 2C_0 + 3\delta$ as claimed. □

The first point in the previous chain can be replaced with $o$:

**Lemma 5.4.** *Let $\overline{P}_n = \{k_1 < \cdots < k_p\}$. The sequence $o, y_{k_1}, y^-_{k_2}, y_{k_2}, \dots, y_{k_p}, y^-_{n+1}$ is a $(2C_0{+}4\delta, D{-}2C_0{-}3\delta)$-chain. Moreover, $d(o, y_{k_1}) \geqslant d(o, r_{k_1}\cdot o) + D - C_0 - 3\delta$.*

*Proof.* The only difference compared to the proof of Lemma 4.4 is that we do not have the inequality $(y_{k_1}, o)_{y^-_{k_1}} \leqslant C_0$ due to the more complicated definition of refined pivotal times. If we can prove that $(y_{k_1}, o)_{y^-_{k_1}} \leqslant C_0 + 3\delta$, the proof of Lemma 4.4 goes through. Let us check this inequality.

By (5-1), the points $y^-_{k_1}, y^{(1)}_{k_1}, y^{(3)}_{k_1}, y^{(4)}_{k_1} (= y_{k_1}), y^{(5)}_{k_1}$ form a $(C_0{+}\delta, D{-}2C_0{-}3\delta)$-chain, and thus $(y^-_{k_1}, y_{k_1})_{y^{(1)}_{k_1}} \leqslant C_0{+}2\delta$ by Lemma 3.7. Moreover, $(o, y^{(1)}_{k_1})_{y^-_{k_1}} \leqslant C_0$ by the definition of pivotal times. As $d(y^-_{k_1}, y^{(1)}_{k_1}) \geqslant D$ is large, it follows that Lemma 3.5 applies to the points $o, y^-_{k_1}, y^{(1)}_{k_1}, y_{k_1}$ with $C = C_0{+}2\delta$. It gives $(y_{k_1}, o)_{y^-_{k_1}} \leqslant C_0{+}3\delta$, concluding the proof that we have a chain.

Moreover, Lemma 3.4 together with Lemma 5.3 give

$$d(o, y_{k_1}) \geqslant d(y^-_{k_1}, y_{k_1}) - (o, y_{k_1})_{y^-_{k_1}} \geqslant (d(o, r_{k_1}\cdot o) + D) - (C_0 + 3\delta),$$

proving the last claim. □

**Proposition 5.5.** *Let $\overline{P}_n = \{k_1 < \cdots < k_p\}$. We have $d(o, y^-_{n+1}) \geqslant \sum_i d(o, r_{k_i}\cdot o)$.*

*Proof.* This follows from Lemmas 5.3 and 5.4, saying that we have a chain between $o$ and $y^-_{n+1}$ with jumps of size at least $d(o, r_{k_i}\cdot o) + D - C_0 - 3\delta$, and from Lemma 3.7 saying that the distance grows at least as the size of the jumps along a chain. □

This completes the discussion of the deterministic properties of the refined pivotal times. Let us turn to their behavior in the random setting, to prove Proposition 5.1. We should show that there are many refined pivotal times with overwhelming probability. For this, we follow the same strategy as in Section 4A.

**Lemma 5.6.** *Fix $s_1, \dots, s_n$, and draw $s_{n+1}$ according to $\mu_S^2 * \rho_{n+1} * \mu_S^2$. The probability that $|\overline{P}_{n+1}| = |\overline{P}_n| + 1$ (that $n+1$ gets added as a refined pivotal time) is at least $1 - 7\eta$.*

*Proof.* In the local geodesic condition, there are seven alignment conditions to be satisfied. When drawing $s_{n+1}$ according to $\mu_S^2 * \rho_{n+1} * \mu_S^2$, each of them is satisfied with probability at least $1-\eta$ (for each of them, this can be seen by fixing all variables but one and using that the last one is picked from a Schottky set). Therefore, they are simultaneously satisfied with probability at least $1-7\eta$. □

To control the backtracking, we defined pivoted sequences. Given $\bar{s}=(s_1,\dots,s_n)$, let us say that another sequence $\bar{s}'=(s_1',\dots,s_n')$ is pivoted from $\bar{s}$ if they have the same refined pivotal times, $d_k'=d_k$ at all times, and $a_k'=a_k$, $b_k'=b_k$, $r_k'=r_k$, $c_k'=c_k$ at times which are not a refined pivotal time. In other words, we freeze the last jump $d_k$, but keep the freedom in the other parts of $s_k$ at refined pivotal times only.

The next lemma is proved exactly like Lemma 4.7.

**Lemma 5.7.** *Let $i$ be a refined pivotal time of $\bar{s}=(s_1,\dots,s_n)$. Replace $s_i = a_i b_i r_i c_i d_i$ with $s_i' = a_i' b_i' r_i' c_i' d_i$ which still satisfies the local geodesic condition (with $n$ replaced by $i$). Then $(s_1,\dots,s_i',\dots,s_n)$ is pivoted from $\bar{s}$.*

Denote by $\bar{\mathcal{E}}_n(\bar{s})$ the sequences which are pivoted from $\bar{s}$. Conditionally on $\bar{\mathcal{E}}_n(\bar{s})$, the variables $s_i'$ over pivotal times $i$ are independent, but drawn from distributions that depend on $i$.

**Lemma 5.8.** *Let $\bar{s}=(s_1,\dots,s_n)$ be a trajectory with $q$ refined pivotal times. We condition on $\bar{\mathcal{E}}_n(\bar{s})$, and draw $s_{n+1}$ according to $\mu_S^2 * \rho_{n+1} * \mu_S^2$. Then, for all $j \geqslant 0$,*

$$\mathbb{P}(|\bar{P}_{n+1}| < q-j \mid \bar{\mathcal{E}}_n(\bar{s})) \leqslant (7\eta)^{j+1}.$$

*Proof.* The proof is essentially the same as for Lemma 4.8. Assume that $s_{n+1}$ is fixed and gives rise to some backtracking. Let us show that further backtracking happens with probability at most $7\eta$, from which the estimate follows inductively. Let $m<k$ be the last two refined pivotal times, and let $x_{i-1}$ be the last point in a chain from $y_m$ to $y_k^-$ witnessing that $y_k^- \in \mathcal{CS}_{y_m}(y_m^+; C_0+\delta)$ as guaranteed by Lemma 5.2.

In $s_k'$, let us condition also with respect to $b_k', r_k', c_k'$ compatible with the local geodesic condition. Then the total number of possible values for $a_k'$ that give rise to $s_k'$ satisfying the local geodesic condition is at least $(1-\eta)|S|$, as one should ensure the condition $((a_k')^{-1}\cdot o, b_k'\cdot o)_o \leqslant C_0$ and $S$ is a Schottky set. Among these, the values of $a_k'$ that may give rise to further backtracking are those for which the points $x_{i-1}, y_k^-, y_k^{(1)}, y_{n+1}^-$ are not $C_0$-aligned, because this alignment would imply $y_{n+1}^- \in \mathcal{CS}_{y_m}(y_m^+; C_0+\delta)$ as in the proof of Lemma 4.8 and would block the backtracking. By the Schottky condition applied twice, there are at most $2\eta|S|$ such $a_k'$. Therefore, the probability of further backtracking is at most $2\eta/(1-\eta) \leqslant 7\eta$. □

**Lemma 5.9.** *Let $A_n = |\bar{P}_n|$ be the number of pivotal times. Then, in distribution, $A_{n+1} \geqslant A_n + U$ where $U$ is a random variable independent from $A_n$ distributed as*

$$\mathbb{P}(U=-j)=(1-7\eta)(7\eta)^j \text{ for } j>0, \quad \mathbb{P}(U=0)=0 \quad \text{and} \quad \mathbb{P}(U=1)=1-7\eta.$$

*In other words,* $\mathbb{P}(A_{n+1} \geqslant i) \geqslant \mathbb{P}(A_n + U \geqslant i)$ *for all* $i$.

*Proof.* This is proved exactly like Lemma 4.9 using Lemma 5.8. □

**Proposition 5.10.** *There exists* $\kappa > 0$ *only depending on* $\eta$ *such that for all* $n$,

$$\mathbb{P}(|\bar{P}_n| \leqslant (1-15\eta)n) \leqslant e^{-\kappa n}.$$

*Proof.* Let $U_1, U_2, \dots$ be a sequence of independent copies of the variable $U$ from Lemma 5.9. Iterating this lemma gives

$$\mathbb{P}(|\bar{P}_n| \geqslant i) \geqslant \mathbb{P}(U_1 + \cdots + U_n \geqslant i)$$

for all $i$. In particular, $\mathbb{P}(|\bar{P}_n| \leqslant (1-15\eta)n) \leqslant \mathbb{P}(U_1 + \cdots + U_n \leqslant (1-15\eta)n)$. The $U_i$ are real random variables with an exponential moment, and expectation $1 - 7\eta - 7\eta/(1-7\eta) > 1 - 15\eta$. Large deviations for sums of i.i.d. real random variables ensure that $\mathbb{P}(U_1 + \cdots + U_n \leqslant (1-15\eta)n)$ is exponentially small. □

*Proof of Proposition 5.1.* We wish to bound $\mathbb{P}(d(o, y_{n+1}^-) \leqslant M)$. By Proposition 5.10,

$$\mathbb{P}(d(o, y_{n+1}^-) \leqslant M) \leqslant \mathbb{P}\big(d(o, y_{n+1}^-) \leqslant M, |\bar{P}_n| \geqslant (1-15\eta)n\big) + e^{-\kappa n}. \tag{5-3}$$

Therefore, we may focus on trajectories with $|\bar{P}_n| \geqslant (1-15\eta)n$. Let $\bar{s} = (s_1, \dots, s_n)$ be such a trajectory, and $\bar{\mathcal{E}}_n(\bar{s})$ its equivalence class under the pivotal relation. We will estimate $\mathbb{P}(d(o, y_{n+1}^-) \leqslant M \mid \bar{\mathcal{E}}_n(\bar{s}))$.

Along $\bar{\mathcal{E}}_n(\bar{s})$, we have $d(o, y_{n+1}^-) \geqslant \sum_{i=1}^{p} d(o, r_{k_i} \cdot o)$ where the pivotal times are $k_1 < \cdots < k_p$, by Proposition 5.5. As $p \geqslant (1-15\eta)n$, we obtain in particular

$$d(o, y_{n+1}^-) \geqslant \sum_{i=1}^{\lfloor (1-15\eta)n \rfloor} d(o, r_{k_i} \cdot o). \tag{5-4}$$

Along $\bar{\mathcal{E}}_n(\bar{s})$, the random variables $r_{k_i}$ are independent, as what happens at different pivotal times is independent by construction, but they are not distributed like $\rho_{k_i}$ a priori, since the local geodesic condition may twist its distribution. Denoting by $L_{k_i}$ the set of $(a, b, r, c)$ that satisfy the local geodesic condition, then the distribution of $(a, b, r, c)$ is $(\mu_S^2 * \rho_{k_i} * \mu_S) 1_{L_{k_i}} / (\mu_S^2 * \rho_{k_i} * \mu_S)(L_{k_i})$. In particular, the probability that $r_{k_i}$ equals a given $r$ is

$$\frac{\rho_{k_i}(r) \mu_S^3\{(a,b,c) : (a,b,r,c) \in L_{k_i}\}}{(\mu_S^2 * \rho_{k_i} * \mu_S)(L_{k_i})} \geqslant \rho_{k_i}(r) \mu_S^3\{(a,b,c) : (a,b,r,c) \in L_{k_i}\}.$$

Once $r$ is fixed, there are six alignment relations to be satisfied for $a, b, c$ to make sure that $(a, b, r, c)$ satisfies the local geodesic condition. Each of them is satisfied with probability at least $1 - \eta$, so we get $\mu_S^3\{(a, b, c)$ such that $(a, b, r, c) \in L_{k_i}\} \geqslant 1 - 6\eta$. Finally,

$$\mathbb{P}(r_{k_i} = r \mid \bar{\mathcal{E}}_n(\bar{s})) \geqslant (1-6\eta) \rho_{k_i}(r).$$

As the distance $d(o, r\cdot o)$ for $r$ drawn according to $\rho_{k_i}$ dominates the random variable $R$ in the assumptions of the lemma, it follows that the conditional distribution in $\bar{\mathcal{E}}_n(\bar{s})$ dominates $BR$, where $B$ is a Bernoulli random variable independent of $R$, equal to 1 with probability $1-6\eta$ and to 0 with probability $6\eta$. Conditionally on $\bar{\mathcal{E}}_n(\bar{s})$, it follows from (5-4) that $d(o, y_{n+1}^{-})$ dominates $\sum_{i=1}^{\lfloor(1-15\eta)n\rfloor} B_i R_i$ where the $B_i$ are copies of $B$ and the $R_i$ are copies of $R$, all independent. As this estimate is uniform over the equivalence classes, we get from (5-3) the inequality

$$\mathbb{P}(d(o, y_{n+1}^{-}) \leqslant M) \leqslant \mathbb{P}\Bigg(\sum_{i=1}^{\lfloor(1-15\eta)n\rfloor} B_i R_i \leqslant M\Bigg) + e^{-\kappa n}.$$

Since the $B_i$ have expectation $1-6\eta$, the probability $\mathbb{P}\big(\sum_{i=1}^{n} B_i \leqslant (1-7\eta)n\big)$ is exponentially small. We get

$$\mathbb{P}(d(o, y_{n+1}^{-}) \leqslant M) \leqslant \mathbb{P}\Bigg(\sum_{i=1}^{\lfloor(1-15\eta)n\rfloor} B_i R_i \leqslant M, \sum_{i=1}^{n} B_i \geqslant (1-7\eta)n\Bigg) + e^{-\kappa' n}.$$

To estimate the probability on the right, let us condition with respect to the $B_i$. There are at most $7\eta n$ of them that vanish. Therefore, $\sum B_i R_i$ is a sum of at least $(1-22\eta)n$ independent copies of $R$, and the probability that the sum is at most $M$ is bounded by $\mathbb{P}\big(\sum_{i=1}^{\lfloor(1-22\eta)n\rfloor} R_i \leqslant M\big)$. As this estimate is uniform over the choice of the $B_i$, this concludes the proof. □

**5B.** ***Precise estimates for walks without first moment.*** In this paragraph, we consider a discrete probability measure $\mu$ on the set of isometries of $X$ which has no first moment: $\mathbb{E}(d(o, g\cdot o)) = \infty$ when $g$ is drawn according to $\mu$. We will prove Theorems 1.2 and 1.3 under this assumption. It suffices to prove the latter, as the former follows readily.

Let $r > 0$ be arbitrary. We have to show the existence of $\kappa > 0$ such that

$$\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant rn) \leqslant e^{-\kappa n}.$$

Let $\eta = \frac{1}{100}$. Let $S$ be an $(\eta, C_0, D)$-Schottky set in the support of $\mu^M$ for some $M > 0$, where $D$ is large enough compared to $C_0$, as given by Corollary 3.13. We follow the construction in Section 4B to reconstruct the $\mu$-random walk, except that instead of sampling the specific jumps from $\mu_S^2$, we will sample them from $\mu_S^2 * \mu * \mu_S^2$: for $N = 4M+1$ and some $\alpha > 0$, we may write $\mu^N = \alpha\mu_S^2 * \mu * \mu_S^2 + (1-\alpha)\nu$ for some probability measure $\nu$, where $\mu_S$ is the uniform measure on $S$.

The random walk is reconstructed by starting from Bernoulli random variables $\varepsilon_i$ satisfying $\mathbb{P}(\varepsilon_i = 1) = \alpha$ and $\mathbb{P}(\varepsilon_i = 0) = 1-\alpha$, and sampling from $\mu_S^2 * \mu * \mu_S^2$ when $\varepsilon_i = 1$ and from $\nu$ when $\varepsilon_i = 0$. Conditioning on $(\varepsilon_i)$ and on the jumps when $\varepsilon_i = 0$, we are left with a walk as in Proposition 5.1. For this walk, we define a sequence of refined pivotal times as in Section 5A. Let $t_1, t_2, \dots$ be the successive times where $\varepsilon_i = 1$. Let $\tau = \tau(n)$ be the last index $j$ such that $N(t_j+1) \leqslant n$, so that the interval $[Nt_j, N(t_j+1))$ is contained in $[0, n)$. Then the sequence of refined pivotal times associated to the walk until time $n$ has the form $\overline{P}_1, \overline{P}_2, \dots, \overline{P}_{\tau-1}, \overline{P}_\tau^{(n)}$. Moreover, $u_n := |\overline{P}_{\tau(n)}^{(n)}|$ satisfies

$$\mathbb{P}(u_n \leqslant \kappa n) \leqslant e^{-\kappa n}, \tag{5-5}$$

for some $\kappa > 0$: this is proved as Proposition 4.11, just using Proposition 5.10 instead of Proposition 4.10 inside the proof.

Assume now that the walk converges at infinity, which is true almost everywhere, and that $u_k > \kappa k$ for all $k \geqslant n$, which is true outside of a set of exponentially small measure, by summing the estimates in (5-5). Let $x = x_n$ be the position of the walk at the $(\lfloor \kappa n \rfloor - 1)$-th refined pivotal time in $\overline{P}_{\tau(n)}^{(n)}$. Then for all $k \geqslant n$, the point $Z_k \cdot o$ belongs to the $(2C_0 + 6\delta)$-shadow of $x$ seen from $o$ (this is proved just like Lemma 4.12, using Lemma 5.4). As in (4-6), this implies the inequality

$$(Z_n \cdot o, Z_\infty)_o \geqslant d(o, x_n) - (2C_0 + 9\delta).$$

Finally, we have

$$\mathbb{P}\big((Z_n \cdot o, Z_\infty)_o \leqslant rn\big) \leqslant e^{-\kappa n} + \mathbb{P}\big(u_n \geqslant \kappa n, d(o, x_n) \leqslant rn + (2C_0 + 9\delta)\big).$$

Let us estimate the rightmost probability when $n$ is large enough so that $2C_0 + 9\delta \leqslant n$. We condition on the $(\varepsilon_i)$, which fixes $\tau$, and on the jumps when $\varepsilon_i = 0$, to be in the setting of Section 5A. As $x$ is one of the points $y_{k+1}^-$ for $\frac{1}{2}\kappa n \leqslant k \leqslant n$, we can sum the estimates of Proposition 5.1 (applied to $k$ instead of $n$), to get a bound of the form

$$n\, \mathbb{P}(R_1 + \cdots + R_{\lfloor \frac{1}{2}(1-22\eta)\kappa n \rfloor} \leqslant (r+1)n),$$

where the $R_i$ are independent random variables distributed like $d(o, g \cdot o)$ where $g$ is drawn according to $\mu$. Letting $\beta = \frac{1}{2}(1-22\eta)\kappa > 0$, we get

$$\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant rn) \leqslant e^{-\kappa n} + n\, \mathbb{P}(R_1 + \cdots + R_{\lfloor \beta n \rfloor} \leqslant (r+1)n).$$

Since we are assuming that $\mu$ has no first moment, the nonnegative random variables $R_i$ are not integrable. Applying the usual large deviations estimate to a truncated version of $R$, we deduce that for any $A > 0$, there exists $c(A)$ such that $\mathbb{P}(R_1 + \cdots + R_k \leqslant Ak) \leqslant e^{-c(A)k}$. Together with the previous equation, this gives an exponential bound on $\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant rn)$. This concludes the proof of Theorem 1.3 (and therefore also of Theorem 1.2) when there is no first moment. □

**5C.** ***Precise estimates for walks with a first moment.*** Assume now that $\mu$ is a measure with a finite first moment. Then $\frac{1}{n}\sum_g \mu^n(g) d(o, g \cdot o)$ converges by subadditivity to $\ell$, the escape rate of the walk. Let $r < \ell$. Our goal in this paragraph is to prove Theorem 1.3, and therefore also Theorem 1.2, in this setting: we will show that, for some $\kappa > 0$, we have

$$\mathbb{P}((Z_n \cdot o, Z_\infty)_o \leqslant rn) \leqslant e^{-\kappa n}.$$

To prove this estimate, we will again use the refined model of Section 5A, but we will have to do so in a careful enough way.

Fix $\eta > 0$ small enough depending only on $r$ and $\ell$ (how small will be prescribed at the very end of the proof). By Corollary 3.13, there exists an $(\eta, C_0, D)$-Schottky set $S$ in the support of $\mu^M$ for some $M > 0$, where $D$ is large enough compared to $C_0$. For $N = 2M$, we may write $\mu^N = \alpha\mu_S^2 + (1-\alpha)\nu$ for some probability measure $\nu$. Replacing $\alpha$ with $\alpha/2$ if necessary, we can also assume that $\nu$ is nonelementary.

Let us now fix $A > 0$ very large (how large will be described in the course of the proof, depending on $\eta$, $\alpha$ and $\nu$). Let $\varepsilon_i$ be a sequence of Bernoulli random variables, equal to 1 with probability $\alpha$ and to 0 with probability $1-\alpha$. Define inductively a sequence of times $t_1, t_1', t_2, t_2', \dots$ as follows. First, $t_1$ is the first time with $\varepsilon_{t_1} = 1$. Then $t_1'$ is the smallest time strictly after $t_1 + A$ with $\varepsilon_{t_1'} = 1$. Then $t_2$ is the smallest time strictly after $t_1'$ with $\varepsilon_{t_2} = 1$. And so on, picking the first times where $\varepsilon_i = 1$ but keeping a gap at least $A$ between $t_i$ and $t_i'$. Then, pick $\gamma_n$ distributed according to the following measure: if $n$ is of the form $t_i$ or $t_i'$, use $\mu_S^2$. If $n$ is in $[t_i + 1, t_i + A]$, use $\mu^N$. Otherwise, use $\nu$.

**Claim 5.11.** *With this construction, the $\gamma_i$ are independent random variables distributed according to $\mu^N$. In particular, $\gamma_0 \cdots \gamma_{n-1}$ is distributed like $Z_{Nn}$.*

*Proof.* Conditionally on the $\varepsilon_0, \dots, \varepsilon_{n-1}$ and on $\gamma_0, \dots, \gamma_{n-1}$, we will show that $\gamma_n$ is distributed according to $\mu^N$, from which the result follows. Consider the maximal $t_j$ or $t_j'$ before $n$. If it is a $t_j$ and $n \leqslant t_j + A$, then $\gamma_n$ is picked according to $\mu^N$ by definition, and there is nothing left to prove. Otherwise, the choice of the measure for $\gamma_n$ depends on $\varepsilon_n$: we use $\mu_S^2$ if $\varepsilon_n = 1$ (with probability $\alpha$) or $\nu$ if $\varepsilon_n = 0$ (with probability $1-\alpha$). Altogether, $\gamma_n$ is drawn according to $\alpha\mu_S^2 + (1-\alpha)\nu = \mu^N$, proving the claim. □

With a coupling argument as in Section 4B, extending $\Omega$ if necessary, we can also construct on $\Omega$ a sequence of independent random variables $g_0, g_1, \dots$, with distribution $\mu$ such that $\gamma_i = g_{iN} \cdots g_{iN+N-1}$.

The intuition behind the use of this decomposition is the following. Since $\alpha$ is possibly small, the times with $\varepsilon_i = 1$, which have frequency $1/\alpha$, may be sparse. However, if $A$ is much larger than $1/\alpha$, the waiting time between $t_i + A$ and $t_i'$, or between $t_i'$ and $t_{i+1}$, will be comparatively much shorter. Therefore, the walk will be essentially a concatenation of long jumps corresponding to $\mu^{NA}$, and short jumps coming from the Schottky sets. The long jumps essentially go in independent directions (this is formalized precisely by Proposition 5.1), so the size of the walk at time $NAk$ will be bounded below by the sum of $(1-22\eta)k$ independent random variables distributed like jumps of $\mu^{NA}$, which are of order $NA\ell$. Altogether, the probability to have size smaller than $(1-22\eta)NAk\ell$ at time roughly $NAk$ will be exponentially small, proving Theorem 1.2 in this setting.

To make this precise, we will need to control quantitatively the waiting times. Also, the distribution of the jumps between $t_i$ and $t_i'$ is not $\mu^{NA}$, but $\mu^{NA} * \nu^{t_i'-(t_i+A)}$. We will have to show that the jumps of this family of measures are uniformly controlled from below, to be able to apply Proposition 5.1. Note that this application motivates why we had to formulate this proposition using different measures $\rho_i$ for the different jumps, instead of one single measure $\rho$.

Let us start the proof, adapting the formalism of Section 4B to our current setting. Fix $n \in \mathbb{N}$. We let $\tau = \tau(n)$ be the last index $j$ such that $N(t_j'+1) \leqslant n$, so that the interval $[Nt_j, N(t_j'+1))$ is contained in $[0, n)$. We will decompose the product $g_0 \cdots g_{n-1}$ as a product of the elements $s_j'$, the product of all $g_i$ for $i \in [Nt_j, N(t_j'+1))$, interspersed with other words that we will consider as fixed, to be in the framework of Section 5A. Let $w_j = g_{N(t_j'+1)} \cdots g_{Nt_{j+1}-1}$, where by convention $t_0' = -1$, and let $w' = w'(n) = g_{N(t'_{\tau(n)}+1)} \cdots g_{n-1}$ be the last missing word, which really depends on $n$, contrary to the previous words that just fill the gaps between blocks $[t_j, t_j']$. By construction,

$$Z_n \cdot o = w_0 s_1' w_1 \cdots w_{\tau-1} s_\tau' w'(n) \cdot o,$$

where each $s_j'$ is decomposed as $a_j' b_j' r_j' c_j' d_j'$, with $a_j', b_j', c_j', d_j'$ independent and distributed according to $\mu_S$.

We can associate to the sequence

$$(w_0, s_1' = a_1' b_1' r_1' c_1' d_1', w_1, s_2' = a_2' b_2' r_2' c_2' d_2', \dots, w'(n))$$

in this decomposition a sequence of refined pivotal times $\overline{P}_1^{(n)}, \dots, \overline{P}_{\tau(n)}^{(n)}$, where the exponent $(n)$ is here to emphasize that all this is done for a fixed $n$, and that the objects we introduce may therefore depend on $n$. Note however that the words $w_j$ for $j < \tau(n)$ only depend on $j$ as they are given by $w_j = g_{N(t_j'+1)} \cdots g_{Nt_{j+1}-1}$. Hence, the sequence of refined pivotal times is rather

$$\overline{P}_1, \overline{P}_2, \dots, \overline{P}_{\tau(n)-1}, \overline{P}_{\tau(n)}^{(n)}.$$

If we condition on the $\varepsilon_i$, which fixes the $t_i$ and $t_i'$, and on the $g_i$ for $i$ not belonging to $\bigcup [Nt_j, N(t_j'+1))$, which fixes the $w_i$ and $w'(n)$, then we are in the setting of Proposition 5.1, with $\rho_i = \mu^{NA} * \nu^{t_j'-(t_j+A)}$. To apply this proposition, we need to check that jumps with respect to such a measure are uniformly bounded below.

**Lemma 5.12.** *Assume that $A$ is large enough. Let $R_{NA}$ be the distribution of the size of jumps for $\mu^{NA}$. Let $B$ be a Bernoulli random variable, equal to* 1 *with* *probability* $1-\eta$ *and to* $0$ *with probability* $\eta$, *independent of* $R_{NA}$. *Then, for any* $i \geqslant 0$ *and for any* $M \geqslant 0$,

$$(\mu^{NA} * \nu^i)\{g : d(o, g \cdot o) \geqslant M\} \geqslant \mathbb{P}(BR_{NA} \geqslant M + \eta NA).$$

*In other words, the jumps for* $\mu^{NA} * \nu^i$ *dominate stochastically* $BR_{NA} - \eta NA$, *uniformly in* $i$, *and therefore by nonnegativity they also dominate* $(BR_{NA} - \eta NA) \vee 0$.

*Proof.* We have

$$(\mu^{NA} * \nu^i)\{g : d(o, g \cdot o) \geqslant M\} = \sum_h \mu^{NA}(h) \cdot \nu^i\{g : d(o, hg \cdot o) \geqslant M\}.$$

Let us apply Lemma 4.14 to the nonelementary measure $\nu$ and to $\varepsilon = \eta$. There exists $C > 0$ such that, for any isometry $h$, for any integer $i$, an element $g$ drawn according to $\nu^i$ satisfies $d(o, hg \cdot o) \geqslant d(o, h \cdot o) - C$ with probability at least $1 - \eta$. This gives

$$\nu^i\{g : d(o, hg \cdot o) \geqslant M\} \geqslant (1-\eta) 1_{d(o,h\cdot o) \geqslant M+C}.$$

Therefore,

$$\begin{aligned}(\mu^{NA} * \nu^i)\{g : d(o, g \cdot o) \geqslant M\} &\geqslant \sum_{d(o,h\cdot o) \geqslant M+C} \mu^{NA}(h) \cdot (1-\eta) \\ &= (1-\eta)\mu^{NA}\{h : d(o, h \cdot o) \geqslant M + C\} \\ &= (1-\eta)\, \mathbb{P}(R_{NA} \geqslant M + C) \\ &= \mathbb{P}(BR_{NA} \geqslant M + C).\end{aligned}$$

Taking $A$ large enough so that $\eta NA \geqslant C$, this is bounded from below by $\mathbb{P}(BR_{NA} \geqslant M + \eta NA)$. □

From now on, we will assume that $A$ is large enough, so that Lemma 5.12 holds.

**Lemma 5.13.** *Assume that* $A$ *is large enough. The sequence* $\tau(n)$ *grows like* $n/(NA)$ *with high probability. More precisely, there exists* $c > 0$ *such that*

$$\mathbb{P}\left(\tau(n) \leqslant \frac{(1-\eta)n}{NA}\right) \leqslant e^{-cn}.$$

*Proof.* We have

$$t'_j - t'_0 = Aj + \sum_{i=1}^{j} (t'_i - (t_i + A)) + \sum_{i=1}^{j} (t_i - t'_{i-1}).$$

The random variables $t'_i - (t_i + A)$ and $t_i - t'_{i-1}$ are independent and identically distributed, and have an exponential tail just depending on $\alpha$. Therefore, there exist

$C > 0$ and $c > 0$ not depending on $A$ such that

$$\mathbb{P}\Bigg( \sum_{i=1}^{j} (t_i' - (t_i + A)) + \sum_{i=1}^{j} (t_i - t_{i-1}') \geqslant Cj \Bigg) \leqslant e^{-cj}.$$

Outside of a set $O_j$ with exponentially small probability, we obtain $t_j' + 1 \leqslant Aj + Cj$. Therefore, $N(t_j' + 1) \leqslant N(Aj + Cj)$, which is bounded by $NAj/(1-\eta)$ if $A$ is large enough compared to $C$. Take $j = j(n) = \lfloor (1-\eta)n/(NA) \rfloor$. It satisfies $NAj/(1-\eta) \leqslant n$. On the complement of $O_j$, we have $N(t_j' + 1) \leqslant n$, and therefore $\tau(n) \geqslant j$. Hence, the inequality $\tau(n) \leqslant (1-\eta)n/(NA)$ can only hold on $O_j$, whose probability is exponentially small in terms of $n$. □

Let $u_n := |\overline{P}_{\tau(n)}^{(n)}|$ be the number of refined pivotal times up to time $n$.

**Lemma 5.14.** *There exists $c > 0$ such that $\mathbb{P}(u_n \leqslant (1-16\eta)n/(NA)) \leqslant e^{-cn}$.*

*Proof.* By Lemma 5.13, we have

$$\mathbb{P}\Big(u_n \leqslant \frac{(1-16\eta)n}{NA}\Big) \leqslant e^{-cn} + \mathbb{P}\Big(u_n \leqslant \frac{(1-16\eta)n}{NA}, \tau(n) \geqslant \frac{(1-\eta)n}{NA}\Big).$$

Let us concentrate on the second set. We condition with respect to $\varepsilon_i$, which fixes the $t_i$, the $t_i'$, and $\tau$, and with respect to the $g_i$ outside of the intervals $[Nt_j, N(t_j'+1))$, which fixes the $w_j$ and $w'$. Once these are fixed, we are in the framework of Section 5A. We may therefore apply Proposition 5.10 and deduce that, conditionally on these quantities, we have

$$\mathbb{P}(u_n \leqslant (1-15\eta)\tau \mid (\varepsilon_i)_i, (g_i)_{i \notin \bigcup_j [Nt_j, N(t_j'+1))}) \leqslant e^{-c\tau},$$

for some $c > 0$. When $\tau \geqslant (1-\eta)n/(NA)$, the left-hand side bounds the conditional probability that $u_n \leqslant (1-15\eta)(1-\eta)n/(NA)$, and the right-hand side is bounded by $e^{-c(1-\eta)n/(NA)}$. As $1-16\eta \leqslant (1-\eta)(1-15\eta)$ and the previous bound is uniform on the conditioning, this implies that $\mathbb{P}\big(u_n \leqslant (1-16\eta)n/(NA), \tau(n) \geqslant (1-\eta)n/(NA)\big)$ is exponentially small, concluding the proof of the lemma. □

Let us proceed with the proof of Theorem 1.3 when $\mu$ has a finite first moment. Assume that $Z_k \cdot o$ converges to a point $Z_\infty$ at infinity and moreover, for all $k \geqslant n$, we have $u_k \geqslant (1-16\eta)k/(NA)$, which happens outside of a set of exponentially small probability, by Lemma 5.14. Let $\bar{t} = \bar{t}(n) = \lfloor (1-17\eta)n/(NA) \rfloor < |\overline{P}_{\tau(n)}^{(n)}|$, and let $x = x_n$ be the walk's position at the $\bar{t}$-th refined pivotal time in $\overline{P}_{\tau(n)}^{(n)}$. An adaptation of Lemma 4.12 to this setting (based on Lemma 5.4) shows that, for all $k \geqslant n$, the point $Z_k \cdot o$ belongs to the $(2C_0 + 6\delta)$-shadow of $x$ seen from $o$. In turn, as in (4-6), this implies the inequality

$$(Z_n \cdot o, Z_\infty)_o \geqslant d(o, x_n) - (2C_0 + 9\delta).$$

Finally, we have

$$\mathbb{P}\big((Z_n\cdot o, Z_\infty)_o \leqslant rn\big) \leqslant e^{-cn} + \mathbb{P}\big(d(o, x_n) \leqslant rn + (2C_0+9\delta)\big).$$

For large enough $n$, we have $rn+(2C_0+9\delta) \leqslant (r+\eta)n$. Together with Lemma 5.14, we get

$$\mathbb{P}\big((Z_n\cdot o, Z_\infty)_o \leqslant rn\big) \leqslant e^{-cn} + \mathbb{P}\Big(d(o, x_n) \leqslant (r+\eta)n, u_n \geqslant \frac{(1-16\eta)n}{NA}\Big).$$

for some $c > 0$.

To conclude, it suffices to show that the rightmost probability is exponentially small. Let us condition on the $\varepsilon_i$, which fixes the $t_i$, the $t_i'$ and $\tau$, and on the $g_i$ for $i$ not belonging to $\bigcup [Nt_j, N(t_j'+1))$, to be again in the setting of Section 5A. Note that $\bar{t}$ is not fixed by this conditioning. However, $x_n$ is one of the points $y_{m+1}^- = w_0 s_1' \cdots s_m' w_m$, for some $m \geqslant (1-17\eta)n/(NA)$. We claim that it suffices to show that, for such an $m$, we have

$$\mathbb{P}\big(d(o, y_{m+1}^-) \leqslant (r+\eta)n \mid (\varepsilon_i)_i, (g_i)_{i\notin \bigcup_j [Nt_j, N(t_j'+1))}\big) \leqslant e^{-cm}. \tag{5-6}$$

Indeed, the right-hand side is exponentially small in terms of $n$. Summing over $m \in [(1-17\eta)n/(NA), n/(NA)]$, we get a bound at most $ne^{-c'n}$, which is again exponentially small as desired.

To prove the inequality (5-6), we apply Proposition 5.1 at the time $m$. Lemma 5.12 shows that the stochastic domination assumptions of this proposition are satisfied, for $R = (BR_{NA} - NA\eta) \vee 0$ where $B$ is a $(1-\eta)$-Bernoulli random variable independent of $R_{NA}$. This proposition gives

$$\begin{aligned}\mathbb{P}\big(d(o, y_{m+1}^-) \leqslant (r+\eta)n \mid (\varepsilon_i)_i, (g_i)_{i\notin \bigcup_j [Nt_j, N(t_j'+1))}\big)\\ \leqslant \mathbb{P}\big(R_1+\cdots+R_{\lfloor (1-22\eta)m\rfloor} \leqslant (r+\eta)n\big) + e^{-cm},\end{aligned}$$

where the $R_i$ are independent copies of $R$. The last term is compatible with (5-6). For the first term, we will apply large deviations for sums of i.i.d. real random variables. We have

$$\mathbb{E}(R_i) = \mathbb{E}(R) \geqslant (1-\eta)\mathbb{E}(R_{NA}) - NA\eta \geqslant (1-\eta)NA\ell - \eta NA,$$

as $\mathbb{E}(R_{NA})/(NA)$ is the average drift at time $NA$, which converges to $\ell$ from above by subadditivity. For $z = (1-\eta)NA\ell - 2\eta NA < \mathbb{E}(R)$, large deviations ensure that $\mathbb{P}(R_1+\cdots+R_k \leqslant zk)$ is exponentially small in terms of $k$. Therefore, it is enough to show that $(r+\eta)n \leqslant z(1-22\eta)m$ to conclude. As $m \geqslant (1-17\eta)n/(NA)$, we have

$$\frac{(r+\eta)n}{z(1-22\eta)m} \leqslant \frac{(r+\eta)n}{((1-\eta)NA\ell - 2\eta NA)(1-22\eta)(1-17\eta)n/(NA)} = \frac{r+\eta}{((1-\eta)\ell - 2\eta)(1-22\eta)(1-17\eta)}.$$

When $\eta$ converges to 0, this converges to $r/\ell < 1$. Therefore, for small enough $\eta$, it is smaller or equal to 1 as desired. This concludes the proof of Theorem 1.3 when $\mu$ has a finite first moment. □

**5D.** ***Continuity of the escape rate.*** As an illustration of the power of the tools we have introduced above, we can recover the fact that the escape rate $\ell(\mu)$ depends continuously on the measure $\mu$, a fact that was originally proved in hyperbolic groups by Erschler and Kaimanovich [2013], and which, in the general setting of nonproper hyperbolic spaces, follows from their proof together with the tools of [Maher and Tiozzo 2018].

**Proposition 5.15.** *Consider a discrete nonelementary measure $\mu$ on the space of isometries of a Gromov-hyperbolic space $X$ with a basepoint $o$. Let $r < \ell(\mu)$. There exist $\varepsilon > 0$ and a finite subset $K$ of the support of $\mu$ with the following property. Let $\mu'$ be a probability measure with $\mu'(g) \geqslant \mu(g) - \varepsilon$ for all $g \in K$. Then $\ell(\mu') \geqslant r$.*

*Furthermore, there exists $\kappa > 0$ such that, for any $\mu'$ as above, the corresponding random walk $Z'_n$ satisfies for any $n \in \mathbb{N}$ the inequality*

$$\mathbb{P}(d(o, Z'_n \cdot o) \leqslant rn) \leqslant e^{-\kappa n}. \tag{5-7}$$

Indeed, all the constants in the proofs in Section 5C are completely explicit. Once $K$ is chosen large enough and $\varepsilon$ small enough to ensure that $\mu'^M$ gives a weight bounded from below to all the elements in the Schottky set $S$ chosen at the beginning of this subsection, then all the estimates go through for $\mu'$ just like for $\mu$. In the end, this gives (5-7) with a uniform $\kappa$. This exponential estimate implies $\ell(\mu') \geqslant r$ as $\frac{1}{n} d(o, Z'_n \cdot o)$ converges almost surely to $\ell(\mu')$.

It follows from the proposition that, when $\mu_n$ converges simply to $\mu$, then $\liminf \ell(\mu_n) \geqslant \ell(\mu)$. This is the nontrivial direction to prove that $\ell(\mu_n) \to \ell(\mu)$.

The other direction is easy and classical: if $\mu_n$ converges in $L^1$ norm to $\mu$ (so $\sum_g d(o, g \cdot o)|\mu_n(g) - \mu(g)| \to 0$), then $\limsup \ell(\mu_n) \leqslant \ell(\mu)$. Indeed, for any measure $\mu'$ one has $\ell(\mu') = \mathrm{Inf}_k\big(\frac{1}{k}\,\mathbb{E}(d(o, Z'_k \cdot o))\big)$ by Kingman's theorem, and each of these quantities for fixed $k$ is continuous in $\mu'$ for the $L^1$ topology. The claim follows. Thus, we recover the following corollary.

**Corollary 5.16.** *Consider a discrete nonelementary measure $\mu$ on the space of isometries of a Gromov-hyperbolic space $X$ with a basepoint $o$, and a sequence of* *probability measures $\mu_n$ converging to $\mu$ in the $L^1$-sense, i.e.,*

$$\sum_g d(o, g\cdot o)|\mu_n(g)-\mu(g)| \to 0.$$

*Then $\ell(\mu_n)$ converges to $\ell(\mu)$.*

## Acknowledgements

We thank Çağrı Sert and Axel Péneau for enlightening discussions, and the referee for a careful reading and advice.

## References

[Berkes and Philipp 1979] I. Berkes and W. Philipp, "Approximation theorems for independent and weakly dependent random vectors", *Ann. Probab.* **7**:1 (1979), 29–54. MR Zbl

[Boulanger et al. 2021] A. Boulanger, P. Mathieu, C. Sert, and A. Sisto, "Large deviations for random walks on Gromov-hyperbolic spaces", preprint, 2021. arXiv 2008.02709v2

[Furstenberg 1963] H. Furstenberg, "Noncommuting random products", *Trans. Amer. Math. Soc.* **108** (1963), 377–428. MR Zbl

[Ghys and de la Harpe 1990] É. Ghys and P. de la Harpe (editors), *Sur les groupes hyperboliques d'après Mikhael Gromov* (Bern, 1988), Progress in Mathematics **83**, Birkhäuser, Boston, 1990. MR Zbl

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

[Maher and Tiozzo 2018] J. Maher and G. Tiozzo, "Random walks on weakly hyperbolic groups", *J. Reine Angew. Math.* **742** (2018), 187–239. MR Zbl

[Mathieu and Sisto 2020] P. Mathieu and A. Sisto, "Deviation inequalities for random walks", *Duke Math. J.* **169**:5 (2020), 961–1036. MR Zbl

[Sunderland 2020] M. Sunderland, "Linear progress with exponential decay in weakly hyperbolic groups", *Groups Geom. Dyn.* **14**:2 (2020), 539–566. MR Zbl

Received 15 Feb 2021. Revised 28 Apr 2022.

SÉBASTIEN GOUËZEL:

sebastien.gouezel@univ-rennes1.fr
IRMAR, CNRS UMR 6625, Université de Rennes 1, Rennes, France

# Tunisian Journal of Mathematics

msp.org/tunis





The Tunisian Journal of Mathematics is an international publication published in electronic and print formats by MSP in Berkeley, and organized by the Tunisian Mathematical Society (http://www.tms.rnu.tn).

See inside back cover or msp.org/tunis for submission instructions.

The subscription price for 2022 is US $345/year for the electronic version, and $405/year (+$20, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Tunisian Journal of Mathematics (ISSN 2576-7666 electronic, 2576-7658 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

TJM peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Tunisian Journal of Mathematics
2022 vol. 4 no. 4