Abstract. We show that the displacement and translation distance of non-elementary random walks on isometry groups of hyperbolic spaces satisfy large deviation principles with the same rate function $I$. Roughly, this means that there exists function $I(t)$ which accurately predicts the exponential decay rate of the probability that the translation distance of a random product of length $n$ is $tn$, and similarly for the displacement. This settles a special case of a conjecture concerning the large deviation principle for the spectral radius of random matrix products. In a second part, we give a characterization of the effective support of the rate function only in terms of the deterministic notion of joint stable length. Finally, as a by-product of our techniques, we deduce some further deterministic results on the asymptotics of a bounded set of isometries. Some of the results in this paper were obtained simultaneously and independently by Boulanger–Mathieu as we discuss in the introduction.

1. Introduction

Let $(X,d)$ be a (Gromov) hyperbolic metric space, e.g. a CAT($-1$)-space, and let $\text{Isom}(X)$ be the group of isometries of $X$. Let $\mu$ be a probability measure on $\text{Isom}(X)$ and $Y_1, Y_2, \ldots$ be a sequence of independent random variables with distribution $\mu$. Given $x \in X$ and for $n \in \mathbb{N}$, denoting by $L_n$ the random product $Y_n \cdots Y_1$, we obtain a natural random process on $X$ given by $L_n x$. (We remark that whereas we state our probabilistic results for the left random walk $L_n = Y_n \cdots Y_1$, the same results hold for the right random walk $R_n = Y_1 \cdots Y_n$, since $L_n$ and $R_n$ have the same distributions given by the convolution $\mu^{*n}$.) Fixing a basepoint $x \in X$ and for $g \in \text{Isom}(X)$, denote by $|g|_x$ the distance $d(gx,x)$. The goal of this paper is to study the probabilistic properties of $|L_n|_x$ as well as those of the translation distance $\tau(L_n) := \inf_{y \in X} d(L_n y, y)$ from a large deviations perspective, that is, studying the probability that average $\frac{1}{n} \tau(L_n)$ takes some specific value far from its expectation (see Definition 1.1).

The process $L_n x$ can be thought of as a random walk on the metric space $(X,d)$; indeed, in the special case where $(X,d)$ is the Cayley graph of a hyperbolic group $G$ with respect to a generating set $S$ and $\mu$ is a probability measure supported on $S$, the process $L_n x$ is precisely a random walk on the Cayley graph starting from $x \in G$. The study of random walks on non-commutative groups and the relations of their properties with the geometric and algebraic structure of groups have a long history, we refer the readers to the books [Woe00, Pet17] for a general overview.

Regarding the asymptotic properties of $|L_n|_x$, despite the non-commutativity of the underlying group and mere subadditivity of the displacement functional $|.|_x$, 2010 Mathematics Subject Classification. 20F67,60F10.
under quite general assumptions, the random variables $|L_n|_x$ tend to behave as a sum of independent real random variables. Under some more restrictive assumptions, the same is true for the translation distance $\tau(L_n)$ which do not enjoy subadditivity either. It is a general fact that, with a finite first moment assumption on $\mu$, the sequence $|L_n|_x$ satisfies a law of large numbers, namely $\frac{1}{n}|L_n|_x$ converges almost surely to a limit $\ell_\mu \in \mathbb{R}$, called the drift or rate of escape of the random walk. This follows from Kingman’s subadditive ergodic theorem. In special cases of the general setting that we will consider in this article, other limit theorems for $|L_n|_x$, such as central or local limit theorem and large deviation estimates were studied by several authors, establishing a relatively satisfactory understanding of the random walks on groups with hyperbolic features (see below). Our goal here is to prove the analogue of the classical result of Cramér’s on large deviation principles, for the random variables $|L_n|_x$ and $\tau(L_n)$.

We now describe our setting. Throughout the article, $(X,d)$ is a separable geodesic metric space that is Gromov hyperbolic (see [CDP90, Ch. 1]). The group $\text{Isom}(X)$ is endowed with the topology of pointwise convergence and the corresponding Borel $\sigma$-algebra. For the probability measure $\mu$, we shall always assume that there exists a closed, first countable and separable subgroup $H < \text{Isom}(X)$ such that $\mu(H) = 1$. Such a probability measure on $\text{Isom}(X)$ will be called admissible. We note that if $\mu(G) = 1$ for some countable group $G$, then $\mu$ is admissible and the reader is invited to think of this case in the sequel. We shall say that a subset $S < \text{Isom}(X)$ is non-elementary if the smallest group containing $S$ contains at least two hyperbolic elements with disjoint fixed point sets on the Gromov boundary $\partial X$ of $X$. Note that the group $\text{Isom}(X)$ acts on the Gromov boundary $\partial X$ and an element $g \in \text{Isom}(X)$ is said to be hyperbolic if it has two fixed points on $\partial X$. Accordingly, a probability measure $\mu$ on $\text{Isom}(X)$ is called non-elementary if its support is (see §2.1). These will be the standing assumptions on the distribution $\mu$ of random products $Y_i$ considered in this article. This setting comprises a broad set of actions of non-amenable groups, for some particular cases, see [MT18, page 1].

Some of the first instances where the negative curvature was exploited in the setting of random walks on groups to deduce some limit theorems are in the works of Ledrappier [Led01] (central limit theorem), Gerl [Ger71] and Sawyer [Saw78] (local limit theorems) on free groups (although, earlier works studying various probabilistic phenomenon in relation to negative or non-positive curvature exist: e.g. Dynkin–Malyutov [DM61], Tutubalin [Tut68], Cartier [Car72] and others [Der75, Bou80, Fur63]). Ledrappier’s central limit theorem was later extended, in increasing order of generality, by Björklund [Bjö10], Benoist–Quint [BQ16a] and Mathieu–Sisto [MS20] who proved it for acylindrical actions. The local limit theorem for free groups was later extended by Lalley and Gouëzel [Lal93, GL13, Gou14] for symmetric random walks on hyperbolic groups.

The study of large deviations on hyperbolic-like groups, compared to aforementioned limit theorems, is rather incomplete. By Kingman’s theorem, the probability of large deviations off the mean converges to zero, i.e. for every $\varepsilon > 0$, $P(|\frac{1}{n}|L_n|_x - \ell_\mu| > \varepsilon) \to 0$ as $n \to \infty$. For free groups, using the spectral gap result of Ledrappier [Led01], one can deduce that the decay rate is exponential for every $\varepsilon > 0$, following the strategy of Le Page [LP82] (see also [BL85]). As Gouëzel remarks [Gou17, Page 4], this analytic approach carries over to random walks on
LARGE DEVIATION PRINCIPLES FOR RANDOM WALKS ON HYPERBOLIC GROUPS

hyperbolic groups and can be used to prove an exponential decay result in that setting. These parallel the corresponding large deviation results of Le Page [LP82] for random matrix products which turned out to be a key ingredient in relatively recent works on dynamics on homogeneous spaces [BFLM11, BQ13]. The relation with random matrix products is more than a mere analogy, since rank-one simple linear groups over local fields act isometrically on their symmetric spaces or the associated Bruhat–Tits buildings, which are hyperbolic metric spaces. In Corollary 1.4 we will exploit this relation to settle a case of a conjecture appearing in [Ser19] for the spectral radius of random matrix products.

We note that the aforementioned exponential decay result pertains rather, but not exclusively, to the non-amenable setting. Indeed, for example, for symmetric random walks on countable amenable groups, i.e. random walks driven by a probability measure \( \mu \) satisfying \( \mu(g) = \mu(g^{-1}) \) for every \( g \in \text{Isom}(X) \), by a classical result of Kesten [Kes59b], the decay rate of \( \mu^{*2n}(g) \) is never exponential (see [BC74] for the locally compact case). Finally, we mention the recent work of Corso [Cor20] where a large deviation principle was shown for word-lengths of random walks on free products.

1.1. Main result on displacement. To describe our results on large deviations, we first recall a definition. Let \( Y \) be a topological space and let \( \mathcal{F} \) be a \( \sigma \)-algebra on \( Y \).

**Definition 1.1.** A sequence \( Z_n \) of \( Y \)-valued random variables is said to satisfy a large deviation principle (LDP), if there exists a lower-semicontinuous function (called the rate function) \( I : Y \to [0, \infty] \) such that for every measurable subset \( R \) of \( Y \), we have

\[
- \inf_{\alpha \in \text{int}(R)} I(\alpha) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n \in R) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n \in R) \leq - \inf_{\alpha \in \overline{R}} I(\alpha)
\]

where, \( \text{int}(R) \) denotes the interior and \( \overline{R} \) the closure of \( R \).

Very roughly, the definition says that the probability that \( Z_n \) (to be thought of as \( \frac{1}{n}|L_n|_x \) or \( \frac{1}{n}\tau(L_n) \) in our case) takes value approximately \( \alpha \) is equivalent to \( e^{-I(\alpha)n} \).

With this definition, Cramér’s theorem says that for a sequence of real valued independent and identically distributed (i.i.d.) random variables \( (Y_i) \) with finite exponential moment, the sequence of averages \( \frac{1}{n} \sum_{i=1}^n Y_i \) satisfies a LDP with a proper convex rate function \( I \), given by the convex conjugate (Legendre transform) of the Laplace transform of \( Y_i \)’s.

A probability measure \( \mu \) on \( \text{Isom}(X) \) is said to have a finite exponential moment if for some \( \alpha > 0 \) and \( x \in X \), we have \( \int \exp(\alpha |g|_x) d\mu(g) < \infty \). This condition is obviously independent of the choice of \( x \).

We can now state our first result.

**Theorem A** (Large deviation principle displacement functionals). Let \((X,d)\) be a separable geodesic hyperbolic metric space and \( \mu \) a non-elementary and admissible probability measure on \( \text{Isom}(X) \). Suppose that \( \mu \) has a finite exponential moment. Then, for every \( x \in X \), the sequence \( \frac{1}{n}|L_n|_x \) of random variables satisfies a large deviation principle with a proper convex rate function \( I : [0, \infty) \to [0, \infty] \).
Remark 1.2. 1. We also prove the existence of a weak large deviation principle without any moment assumption (see Theorem 2.2).

2. Under a stronger moment condition, by exploiting the convexity of $I$, we identify the rate function $I$ with the convex conjugate of a limit log-Laplace transform of the random variables $\frac{1}{n} |L_n|_x$.

It is easy to see that the rate function $I$ given by Theorem A does not depend on the choice of the base point $x \in X$. Below, we list some more remarks on this result:

Remark 1.3. 1. By convexity of $I$, the set $D_I = \{ \alpha \in [0, \infty) \mid I(\alpha) < \infty \}$, called the effective support of $I$, is an interval and $I$ is continuous on $D_I$. By Theorem A, this in turn implies that for every subset $J$ of $D_I$ satisfying $\text{int}(J) = J$ (e.g. any interval with non-empty interior), the limit $\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n} |L_n|_x \in J)$ exists and is equal to $\max_{\alpha \in J} -I(\alpha)$ (see Theorem C for more on $D_I$).

2. It is not hard to see that under mild assumptions on the space $(X, d)$ and the action (see §2.5), for any symmetric finitely supported probability measure $\mu$, we have $I(0) = -\log(r_\mu)$, where $r_\mu$ is the spectral radius of the random walk, i.e. the spectral radius of the self-adjoint operator associated to $\mu$ through the regular representation on $\ell^2(G_\mu)$, which, by Kesten’s formula, is given by $\lim_{n \to \infty} \mu^* e^{2n(e)^2}$. Here $G_\mu$ denotes the group generated by $\mu$.

3. It is clear that the drift $\ell_\mu$ is a zero of the rate function $I$. The aforementioned property of exponential decay of probabilities of large deviations off the drift, which can be deduced from a spectral gap result [Gou17, page 4] in the particular case of random walks on hyperbolic groups, is equivalent to $\ell_\mu$ being the unique zero of $I$.

We were informed that an ongoing work of Boulanger–Mathieu proves this decay result in our setting (see Remark 1.9). We do not deal with this point in this article.

It is not very difficult to pinpoint the explicit expression of the rate function for the standard random walk on the free group $F_q$ of rank $q \geq 1$. It is given by the following

$$I(\alpha) = \begin{cases} \frac{1+\alpha}{2} \log(1+\alpha) + \frac{1-\alpha}{2} \log(1-\alpha) + \log(q) - \frac{1+\alpha}{2} \log(2q-1) & \alpha \in [0, 1] \\ \infty & \text{otherwise} \end{cases}$$

We remark that, among others, this function satisfies the following properties:

1) it is analytic and strictly convex on its effective support $D_I := \{ x \in [0, \infty) \mid I(x) < \infty \},$

2) $I(0) = -\log \frac{2^{q-1}}{q}$ where $\frac{2^{q-1}}{q}$ is the spectral radius of the standard random walk on $F_q$ calculated by Kesten [Kes59b],

3) the drift $\frac{q-1}{q}$ is the unique zero of $I,$

4) if $\Lambda(\lambda)$ denotes the Legendre transform of $I$ given by $\Lambda(\lambda) = \sup_{\alpha \in \mathbb{R}} (\lambda \alpha - I(\alpha)),$ then $\Lambda''(0) - (\frac{q-1}{q})^2$ is the asymptotic variance appearing in the central limit theorem for the standard random walk on the free group (this fact can be deduced either directly or as in [BL85, Lemma 5.2]).

Whereas finding an explicit expression for the rate function $I$ does not seem to
be feasible in general, pinning down some of its general properties, paralleling the
above ones, is a more tractable challenge. For example, as mentioned previously,
the properties 2) and 3) above are known to hold under general assumptions and
the properties 1) and 4) naturally suggest the corresponding open problems. We
mention only a few of them:

Questions: Is the rate function appearing in Theorem A strictly convex? Analy-
tic? Do these properties depend on generating set or probability measure?

Regarding the proof of Theorem A we use a general criterion provided by Theo-
rem 2.7 implementing it in a similar way as in Ser19. This requires being able to
find, in quantitative and uniform fashion, local perturbations of the random walk
such that the displacement functionals $|.|_x$ of the corresponding random walk on X
will have an almost-additive property. Such perturbations are found thanks to the
use the negatively curved geometric structure in a form which manifests itself in the
unfolding lemma (Lemma 2.5). In our setting, this lemma yields the analogous tools
provided by the results of Abels–Margulis–Soifer AMS95 and Benoist Ben96 on
the proximality and almost-norm-multiplicativity of linear transformations.

1.2. Main result on translation distance. Another important notion to measure
the size of an isometry $g$ of $(X, d)$ is its translation length $\tau(g) = \inf_{x \in X} |g|_x$. It
has the advantage of being more intrinsic, in particular it does not depend on the
choice of a basepoint $x \in X$ and it is a conjugacy invariant. On the other hand, it
displays a more chaotic behaviour compared to $|g|_x$. For example, the useful
subadditivity property $|gh|_x \leq |g|_x + |h|_x$ fails for $\tau(.)$. The following theorem
proves the existence of a LDP for the sequence $\frac{1}{n} \tau(L_n)$. Moreover, we note that,
under a moment assumption, it can be shown that for a non-elementary random
walk, on a set of large probability, the behaviour of $\frac{1}{n} \tau(L_n)$ is similar to that of
$\frac{1}{n} |L_n|_x$ (for example, they converge almost surely to the same constant $\ell_{\mu}$). The
following result also says that, at an exponential scale, the probabilistic behaviours
of these two different notions of size, coincide not only for events of large probability
but also for the rare (exponentially small) ones:

Theorem B (Large deviation principle for translation distances). Let $(X, d)$ be
a separable geodesic hyperbolic metric space and $\mu$ be a boundedly supported, non-
 elemental and admissible probability measure on the group of isometries of $(X, d)$.
The sequence $\frac{1}{n} \tau(L_n)$ of random variables satisfies a large deviation principle with
the same rate function $I : [0, \infty) \to [0, \infty]$ given by Theorem A.

Here, a probability measure $\mu$ on $\text{Isom}(X)$ is said to be bounded if $\sup_{s \in \text{supp} \mu} |s|_x$ is finite for some $x$ (equivalently for all $x \in X$).

A common and sometimes more convenient way to express a notion of translation
length of $g \in \text{Isom}(X)$ is given by that of asymptotic translation length or stable
length defined as $\ell(g) = \lim_{n \to \infty} \frac{1}{n} |g^n|_x$. The limit exists by subbadditivity and does
not depend on $x$. The difference $|\ell(.) - \tau(.)|$ is uniformly bounded on $\text{Isom}(X, d)$ (see
CDP90, Ch.10, Prop. 6.4). Consequently, the previous theorem applies equally to
the random variables $\frac{1}{n} \ell(L_n)$ with the same conclusion.
1.3. Consequences for rank-one linear groups. Let us explain a consequence of this result that confirms a conjecture stated in [Ser19] in a particular case. A simple linear algebraic group $H$ of rank 1 over a local field $k$ (e.g. $SL_2(\mathbb{R})$ or $SL_2(\mathbb{Q}_p)$), has a natural, up to finite index, faithful action by isometries on its symmetric space or the associated Bruhat–Tits tree $(X,d)$. The metric space $(X,d)$ is a Gromov hyperbolic space and for some finite dimensional representation of $H$ on a $k$-vector space $V$ endowed with a Hermitian (if $k$ is archimedean) or ultrametric (if $k$ is non-archimedean) norm, and for $x \in X$ and $h \in H$, the displacement functional $|h|_x$ is given by the logarithm of the associated operator norm $||.||$ (see e.g. [BQ16b] Chapter 6.8 and [Qm02] §6). Moreover, the translation distance $\tau(h)$ corresponds to the logarithm of the spectral radius $\rho(h)$ of $h$, defined by the spectral radius formula $\rho(h) = \lim_{n \to \infty} ||h^n||^{1/n}$. In this case, under the assumptions of Theorem A the existence of a convex rate function for $\frac{1}{n}|L_n|_x$ follows from the main result of [Ser19] (as well as, from Theorem A). It was conjectured [Ser19] Conjecture 6.2 (see also [BSIS §5.15]) that, if the support of the probability measure $\mu$ on $H$ generates a Zariski-dense semigroup (equivalently, if $\mu$ is non-elementary), then the sequence $\frac{1}{n} \log \rho(L_n)$ satisfies a LDP and the rate function coincides with the rate function of the sequence $\frac{1}{n} \log ||L_n||$. Now one sees that for a probability measure $\mu$ with a compact support, this conjecture follows from Theorem B.

**Corollary 1.4.** Let $H$ be a simple linear algebraic group of rank one over a local field $k$ endowed with an absolute value $|.|$. Let $\mu$ be a compactly supported probability measure on $H$ whose support generates a Zariski dense semigroup in $H$. Let $||.||$ be an operator norm on a finite dimensional representation $V$ of $H$ as above and $I:[0,\infty) \to [0,\infty]$ be the rate function of the LDP of $\frac{1}{n} \log ||L_n||$. Then, the sequence $\frac{1}{n} \log \rho(L_n)$ of random variables satisfies a LDP with rate function $I$.

1.4. On the proof of Theorem B. Theorem B cannot be proven with methods similar to those that work for Theorem A due to the more chaotic behavior of $\tau(.)$. We use genuinely different ideas for its proof, in particular cyclic permutations of trajectories. In fact, the proof of Theorem B uses Theorem A and consists of several parts. In a first part, using the unfolding lemma, we show that, given a prescribed speed $\alpha \geq 0$, for large $n$, the event $\frac{1}{n} \tau(L_n) \sim \alpha$ is, at an exponential scale of the probability, more likely than $\frac{1}{n}|L_n|_x \sim \alpha$. In the second part, for all prescribed speeds $\alpha > 0$, we show that, up to a linear loss (i.e. up to multiplying one of the probabilities by $n$), the converse inequality holds for probabilities of these events, which implies the desired inequality of exponential rates. This step relies on Lemma 3.5 that uses an argument that finds, among the cyclic permutations of a given trajectory, a word whose displacement $|.|_x$ is uniformly close to the translation distance, which is invariant by cyclic permutation. Finally, the control of exponential decay rates for sublinear speed, requires an upgrade of the previous combinatorial lemma in order to carry out a contradiction argument using Theorem A.

1.5. Properties of the rate function. A natural question motivated by the previous results concerns the understanding of the effective support $D_I = \{\alpha \in [0,\infty) | I(\alpha) < \infty\}$ of the rate function $I$. By convexity of the rate function, the effective support $D_I$ is an interval in $[0,\infty)$. The following result gives a geometric characterization of this interval only in terms of the support of the probability
measure $\mu$ and it relates the effective support with the recently introduced notion of asymptotic joint displacement of a bounded set of isometries of a metric space.

To state this result, we need some terminology. A set $S$ of isometries of a metric space $(X,d)$ is said to be non-arithmetic if there exist $n \in \mathbb{N}$ and $g, g' \in S^n$ such that $\ell(g) \neq \ell(g')$. As in [BQ16a], we shall also call a probability measure $\mu$ non-arithmetic if its support is. A set $S$ is said to be bounded if the quantity $L(S,x) := \sup_{s \in S} |s|_x$, called the maximal displacement of $S$ at $x \in X$, is bounded for some $x$ (equivalently for all $x \in X$). The joint minimal displacement of $S$ is defined to be $L(S) := \inf_{x \in X} L(S,x)$. We shall denote the limit $\lim_{n \to \infty} \frac{1}{n} L(S^n)$, which exists by subadditivity, by $\ell(S)$; it is called the asymptotic joint displacement [BF18] or joint stable length [OR18] of $S$. Similarly, let $\ell_{\text{sub}}(S)$ denote the limit $\lim_{n \to \infty} \inf_{g \in S^n} \frac{1}{n}|g|_x$ that we call lower asymptotic joint displacement of $S$. The previous limit exists by subadditivity and does not depend on $x$.

**Theorem C** (Effective support of the rate function). Let $(X,d)$ be a separable geodesic hyperbolic metric space and $\mu$ be a non-elementary and admissible probability measure on the group of isometries of $(X,d)$. Let $I$ be the rate function given by Theorem A (see also Theorem 2.2).

1) The probability measure $\mu$ is non-arithmetic if and only if the effective support $D_I$ of $I$ is an interval with non-empty interior.

2) If the support $S$ of $\mu$ is a bounded set, we have $(\ell_{\text{sub}}(S), \ell(S)) \subseteq D_I \subseteq \overline{D_I} \subseteq [\ell_{\text{sub}}(S), \ell(S)]$.

3) If $S$ is finite, then $D_I = [\ell_{\text{sub}}(S), \ell(S)]$.

**Remark 1.5.** Regarding the second statement in the previous theorem, in §4.1, we provide examples of probability measures $\mu$ with bounded support $S$ for which the rate function $I$ explodes on the boundary points of $D_I$.

The notion of asymptotic joint displacement is analogous to the classical notion of joint spectral radius from linear algebra. In this geometric setting, it was recently studied by Oregón-Reyes [OR18] and Breuillard–Fujiwara [BF18] who proved the geometric analogues of some of the main results on joint spectral radius. The previous result parallels [Ser19, Theorem 1.7] where the effective support of the rate function of the norms of random matrix products was related to joint spectral radii.

**1.6. Deterministic results.** Here we record some deterministic (as opposed to probabilistic) consequences of our results and the ingredients we develop to prove them.

**1.6.1. Hausdorff convergence.** The following is a consequence of the combination of Theorems A, B and C.

**Proposition 1.6.** Given a bounded non-elementary subset $S$ of Isom$(X)$, the sequences of subsets $\frac{1}{n}|S^n|_x$ and $\frac{1}{n}\tau(S^n)$ of $\mathbb{R}$ converge to $[\ell_{\text{sub}}(S), \ell(S)]$ with respect to the Hausdorff metric.

In other words, the sequences $\frac{1}{n}|S^n|_x$ and $\frac{1}{n}\tau(S^n)$ become more and more dense in the interval $[\ell_{\text{sub}}(S), \ell(S)]$ as $n$ grows. In fact, Theorems A and B can be seen as quantitative refinements of this convergence.
This result parallels the convergence result proven in [BS18, Theorem 1.3] for the vectors of singular values and moduli of eigenvalues of powers of a set of matrices. Indeed, for the particular case of rank-one symmetric spaces of non-compact type, this convergence follows from [BS18]. The interval \([\ell_{\text{sub}}(S), \ell(S)]\) corresponds to what is called the joint spectrum of \(S\) in that article.

In passing, we also note that a direct deterministic consequence of Theorem [\(\text{C}\)] (in fact, rather, of its proof, see Remark [\(\text{A.1}\)]) regarding the asymptotic joint displacements/joint stable length is the following:

**Corollary 1.7.** Given a non-elementary and bounded subset \(S\) of the isometry group of a hyperbolic space \((X, d)\), the set \(S\) is non-arithmetic if and only if \(\ell_{\text{sub}}(S) \neq \ell(S)\).

1.6.2. **Berger–Wang equality.** We finally mention a deterministic consequence of the geometrical ingredients that we develop to prove Theorem [\(\text{A}\)] on large deviations. For a subset \(S\) of \(\text{Isom}(X, d)\), we denote by \(\ell_{\text{max}}(S)\) the maximal stable length \(\max_{g \in S} \ell(g)\) and by \(\ell_{\infty}(S) = \sup_{k \in \mathbb{N}} \frac{1}{k} \ell_{\text{max}}(S^k)\). Clearly \(\ell_{\text{max}}(S) \leq \ell_{\infty}(S) \leq \ell(S)\) (see also [BF18, Lemma 1.1]). Using the unfolding lemma, we can easily deduce

**Proposition 1.8** (Geometric Berger–Wang equality, [OR18, BF18]). For a bounded non-elementary subset \(S\) of \(\text{Isom}(X, d)\), we have \(\ell_{\infty}(S) = \ell(S)\).

This result is a geometric analogue of a result due to Berger–Wang which tells that for a bounded set \(S\) of matrices in \(\text{Mat}(d, \mathbb{C})\), one has \(\sup_{k \in \mathbb{N}} \sup_{g \in S^k} \frac{1}{k} \log \rho(g) = \lim_{n \to \infty} \sup_{g \in S^n} \frac{1}{n} \log \|g\|\), where \(\rho(.)\) denotes the spectral radius and \(\|.,\|\) is any operator norm on the algebra \(\text{Mat}(d, \mathbb{C})\). We note that a stronger (without the non-elementary assumption) form of Proposition [LS] as well as more general Bochi-type inequalities were recently proven by Oregón-Reyes [OR18] and by Breuillard–Fujiwara [BF18].

We note that using the Berger–Wang inequality, one can show that the joint stable length/asymptotic joint displacement \(\ell(.)\) is continuous on the set of compact subsets of \(\text{Isom}(X)\) endowed with the Vietoris topology (see [ORIN]). On the other hand, in the particular case of \(X = \mathbb{H}^2\), in [BM15, Corollary 1.9] Bochi–Morris proves also the continuity of the lower asymptotic joint displacement \(\ell_{\text{sub}}(.)\). This leaves us wonder whether the interval \([\ell_{\text{sub}}(.), \ell(.)]\) is continuous in general (see also [OR18, §6]).

1.7. **Outline.** The paper is organized as follows. In Section [\(\text{B}\)], we introduce some preliminaries of large deviation theory and we prove Theorem [\(\text{B.2}\)] which is a more general version of Theorem [\(\text{A}\)]. In Section [\(\text{C}\)] after introducing the necessary geometric ingredients, we prove Theorem [\(\text{B}\)] in three subsection corresponding to three parts of the proof. Finally, in Section [\(\text{D}\)], we prove Theorem [\(\text{C}\)] as well as Propositions [\(\text{1.6}\)] and [\(\text{1.8}\)].

**Remark 1.9.** Before the completion of this article, we were informed that Adrien Boulanger and Pierre Mathieu were also studying the large deviations phenomena in our setting. In particular, for countably supported probability measures, Theorem [\(\text{A}\)] also follows from their work. Our proofs for that result are somewhat similar. Boulanger–Mathieu then concentrate on the positivity properties of the rate function and they show that in our setting, for countable groups, the rate function has a unique zero. Unlike the analytic approach mentioned above to tackle such problems,
their work, together with an observation of Hamana [Ham01], makes use of various nice elaborations of deviation inequalities developed in [MS20], thereby exploiting directly the coarse negative curvature to tackle this problem. However, they do not treat large deviations of translation distance (Theorem B) and the characterization of the support of the rate function (Theorem C). Our works can be seen to be complementary to each other.

Acknowledgements. We thank Pierre Mathieu for remarks on a draft version of this article. C.S. is supported by SNF grant 182089.

2. Large deviation principle for displacement functionals

In §2.1, we introduce the notion of weak LDP and state the more precise version of Theorem A. We prove the existence of weak LDP for displacement functionals in §2.1 and in §2.3, we show the convexity of the rate function. Finally, in §2.4, we complete the proof of Theorem 2.2, and hence that of Theorem A.

2.1. Existence of Weak LDP with a convex rate function. Recall that by definition of an admissible probability measure $\mu$, there exists a closed, first countable and separable group $H$ such that $\mu(H) = 1$. By Birkhoff-Kakutani theorem, such a group is metrizable and by [Par05, Theorem 2.1], there exists a unique minimal closed subset $S$ of $H$ with $\mu(S) = 1$. The subset $S$ is called the support of $\mu$. It has the property that for every $g \in S$ and any neighborhood $U$ of $g$, $\mu(U) > 0$. The subgroup of $H$ generated by $S$ will be denoted by $G$. To avoid any measurability issue, we complete the Borel $\sigma$-algebra of $H$ with respect to the convolutions $\mu_n$, which are defined as push-forwards of the product measures $\mu_n$ on $G_n$ to $G$ by the multiplication map. Finally, given such a probability measure $\mu$ on Isom($X$), we fix some probability space $(\Omega, F, \mathbb{P})$ on which the independent random walk increments $Y_i$ for $i \geq 1$ are defined and have distribution $\mu$.

In Definition 1.1, a LDP with a rate function $I$ for a sequence of random variables $Z_n$ (in our case, to be thought of as $\frac{1}{n}|L_n|$ or $\frac{1}{n}\tau(L_n)$) with values in a topological space $Y$ (in our case $Y = [0, \infty)$), can be reformulated as saying

1. (Upper bound) For any closed set $F \subset Y$, $\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n \in F) \leq -\inf_{\alpha \in F} I(\alpha)$.
2. (Lower bound) For any open set $O \subset Y$, $\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n \in O) \geq -\inf_{\alpha \in O} I(\alpha)$.

The definition of a weak LDP is a slight weakening of the upper bound in the previous reformulation:

**Definition 2.1.** A sequence of $Y$-valued random variables $Z_n$ is said to satisfy a weak LDP with a rate function $I : Y \to [0, \infty]$ if the upper bound 1. (above) holds for all compact sets and the lower bound 2. also holds, for all open sets in $Y$.

We note in passing that if $Y$ is locally compact or a polish space and a sequence of random variables $Z_n$ on $Y$ satisfies a weak LDP with a rate function $I$, then $I$ is unique.

Recall that the **limit Laplace transform** of the sequence $(|L_n|)_n$ is the function $\Lambda : \mathbb{R} \to \mathbb{R}$ defined as

$$\Lambda(\lambda) = \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{\lambda|L_n|}]$$
Nice analytic properties (e.g. differentiability, steepness) of this function have direct implications for the LDP (see e.g. Gärtner-Ellis theorem [DZ02, §4]). Finally, we say that the probability measure $\mu$ has a strong exponential moment if $\int \exp(\alpha |g|_x) d\mu(g) < \infty$ for every $\alpha > 0$.

We can now state the more precise version of Theorem A.

Theorem 2.2 (Weak and full LDP for displacement functionals). Let $(X, d)$ be a separable geodesic hyperbolic metric space and $\mu$ be a non-elementary and admissible probability measure on $\text{Isom}(X)$. Then, for every $x \in X$, the sequence $\frac{1}{n} |L_n|_x$ of random variables satisfies a weak large deviation principle with a convex rate function $I : [0, \infty) \to [0, \infty]$. If $\mu$ has a finite exponential moment, the (full) LDP holds and the rate function $I$ is proper. If, furthermore, $\mu$ has strong exponential moment, then $I$ is given by the Legendre transform of $\Lambda$, i.e. $I(\alpha) = \sup_{\lambda \in \mathbb{R}} (\alpha \lambda - \Lambda(\lambda))$.

The rest of this section is devoted to the proof of this theorem. We fix a basepoint $x \in X$ for the sequel.

2.2. Existence of weak LDP for displacement functionals. The aim of this part is to prove the existence of weak LDP for displacement functionals without any moment assumption on the probability measure. This is done in the following Proposition 2.3. Under the additional exponential moment assumption, the existence of a proper rate function as claimed in Theorem 2.2 will follow by standard techniques in large deviation theory; this will be shown in §2.4.

Proposition 2.3 (Weak LDP). Keep the assumptions of Theorem 2.2. Then, the sequence $\frac{1}{n} |L_n|_x$ of random variables satisfies a large deviation principle with a rate function $I : [0, \infty) \to [0, \infty]$.

The following technical result contains the crucial stability property that will be key in the proof of Theorem A together with the Lipschitz property of the displacement functionals. It says that if the averaged displacement functionals visit a certain neighborhood of a speed $\alpha$ with an exponential decay rate $\beta$ along some subsequence of times, then they will visit a slightly larger neighborhood of $\alpha$ with at most a slightly larger decay rate along periodic times. To ease the notation, we denote by $B(\alpha, r)$ the intersection of the interval $[\alpha - r, \alpha + r]$ with $[0, \infty)$.

Lemma 2.4 (Stability of exponential decay rate). Let $\alpha \in [0, \infty)$. For every $\epsilon, \delta > 0$ and neighborhood $B(\alpha, r)$ of $\alpha$ in $[0, \infty)$, there exists $m_0 \in \mathbb{N}$ such that for every $k \geq 1$, we have

$$- \frac{1}{km_0} \log \mathbb{P}\left(\frac{1}{km_0} |L_{km_0}|_x \in B(\alpha, r + \epsilon)\right) - \delta \leq - \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} |L_n|_x \in B(\alpha, r)\right)$$

(2.1)

For the proof of this lemma, we shall require the following key geometric ingredient.

The reader interested in discrete supports only does not need to worry about neighborhoods, and just take $V_f = \{f\}$.

Lemma 2.5 (Unfolding Lemma). Given a non-elementary set $S$, there exist $k_S \in \mathbb{N}$, a constant $L > 0$, a finite subset $F \subseteq S^{k_S}$ and bounded neighborhoods $\{V_f\}_{f \in F}$ of
the elements of $\mathcal{F}$ so that the following holds. For every $n \in \mathbb{N}$ and elements $g_1, \ldots, g_n \in \text{Isom}(X)$ with $\min_{i=1,\ldots,n} |g_i|_x \geq L$, there exist elements $f_i \in \mathcal{F}$ with the property that

$$\tau(f'_n g_n f'_{n-1} g_{n-1} \cdots f'_1 g_1) \geq \sum_{i=1}^n |g_i|_x.$$ 

for every $f'_i \in V_{f_i}$.

Regarding the neighborhoods, recall that we call a subset of $G$ bounded if it moves a given point of $X$ a bounded amount.

This lemma parallels an observation of Benoist [Ben96, Proposition 6.4] concerning the coarse multiplicity of spectral radius and norms of perturbations of products matrices based on Abels–Margulis–Soifer’s [AMS95] main result. In our geometric setting, for statements of somewhat similar flavor, see [Gou14, Lemma 2.4], [MS20, Lemma 9.5] and [DS18, Lemma 6.3].

**Proof.** The bound on the translation length will follow from the following well-known fact, several variations of which have appeared in the literature, see e.g. [Gro87, 7.2.C], [GdlH90, Théorème 5.16], [MS19, Proposition 14]. There exists $C = C(\delta) > 0$ so that, given $h_i \in G$, $i = 1, \ldots, n$, satisfying

$$|h_i|_x \geq (h_{i-1}^{-1} x |h_i x)_x + (h_i^{-1} x |h_i x)_x + C,$$

where we set $h_0 = h_n$ and $h_{n+1} = h_1$, we have

$$\tau(h_1 \ldots h_n) \geq \sum_{i=1}^n |h_i|_x - \sum_{i=1}^n (h_i^{-1} x |h_i+1 x)_x - nC.$$

It is possible to give a more constructive argument for the proof of the lemma, but we chose a probabilistic argument instead. First, notice that there is a finite subset $S'$ of $S$ so that there are independent loxodromic elements in the semigroup generated by $S'$. We consider the uniform probability distribution $\nu$ on $S'$, and the random walk driven by $\nu$ the convolution powers of $\nu$. Also, we denote the Gromov product at $z$ by $(\cdot|\cdot)_z$.

**Claim.** There exists $k$ so that for each $g \in G$ the following hold:

- $\nu^k(\{f : |f|_x \geq 10C \} > .9,$
- $\nu^k(\{f : (f x|gx)_x, (f^{-1} x|gx)_x \leq |f|_x / 2 - C - 1) > .9,$

**Proof.** This is a combination of the facts that

- the random walk driven by $\nu$ makes linear progress in $X$, meaning that with probability going to 1 as $k$ goes to infinity it has a linear lower bound on $|\cdot|_x$ ([MT18, Theorem 1.2]), and
- the exponential decay of shadows estimate given by [Mah12, Lemma 2.10] (which, as observed [MT18, Page 220], holds in our setting as well).

We fix $k_S = k$ as in the claim and set $\mathcal{F} = (S')^k$. We can choose neighborhoods $V_f$, for $f \in \mathcal{F}$, so that $d(f x, f' x) \leq 1$ and $d(f^{-1} x, f'^{-1} x) \leq 1$ for all $f' \in \mathcal{F}$. In particular, if $|f|_x \geq 10C + 1$ then $|f'|_x \geq 10C + 1$ for all $f \in V_f$, and similarly Gromov products vary by at most 1 when replacing $f x$ with $f' x$ or $f^{-1} x$ with $f'^{-1} x$. Finally, we choose $L = 10k_S + 10C$. 

In view of the claim, for each $i$ there is $f_i \in \mathcal{F}$ so that for any $f_i' \in V_{f_i}$ we have
\[(f_i'^{-1}x|g_i|x)_x \leq \min\{|g_i|x/2, |f_i'|x/2\} - C,\]
and
\[(g_i+1|f'_i|x)_x \leq \min\{|g_i+1|x/2, |f'_i|x/2\} - C,\]
where we set $g_{n+1} = g_1$.

The required bound on $\tau(f'_n g_n \ldots f'_1 g_1)$ now follows from the estimate given at the beginning of the proof (notice that the positive contributions of the $|f'_i|x$ outweigh the negative contributions of the Gromov products and the $"-nC")$. \hfill \Box

**Proof of Lemma 2.4.** When $\alpha = 0$, (2.4) follows by subadditivity of the displacement functional and i.i.d. property of the increments $Y_i$'s. So, we can suppose that $\alpha > 0$. It is also easy to see that it suffices to prove the statement for every $r > 0$ small enough, so we suppose $0 < r < \alpha$ and fix $\varepsilon > 0$ with $\alpha > r + \varepsilon$.

Consider the finite set $\mathcal{F}$, the neighborhoods $(V_f)_{f \in \mathcal{F}}$, and the constants $k_S, L > 0$ as given by Lemma 2.5. Let $\sup_{f \in \mathcal{F}}|f'|_x := M_F \in \mathbb{R}$, where $\mathcal{V} = \cup_{f \in \mathcal{F}} V_f$. By the property of $k_S \in \mathbb{N}$, the finite set $\mathcal{F}$ is in the support of $\mu^{k_S}$ and we have $\mu^{k_S} (V_f) > 0$ for every $f \in \mathcal{F}$; set $\alpha_F := \min_{f \in \mathcal{F}} \mu^{k_S} (V_f)$. We have $\alpha_F > 0$. For any $k \in \mathbb{N}$ and $k$-tuple $(g_k, \ldots, g_1) \in C^k$ with $|g_i|x \geq L$ as in Lemma 2.5 using that lemma, choose elements $f(g_i, g_{i-1}) \in \mathcal{F}$ so that we have
\[|g_k f'(g_k, g_{k-1})g_{k-1} \ldots g_2 f'(g_2, g_1)|_x \geq \sum_{i=1}^k |g_i|x \tag{2.2}\]
for every $f'(g_i, g_{i-1}) \in V_{f(g_i, g_{i-1})}$.

Setting $\beta := -\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n}|L_n|x \in B(\alpha, r))$, fix $n_0 \in \mathbb{N}$ large enough so that
\[
\begin{align*}
(1) \quad & \mathbb{P}\left(\frac{1}{n_0}|L_{n_0}|x \in B(\alpha, r)\right) \geq e^{-n_0(\beta + \delta/2)}, \\
(2) \quad & n_0(\alpha - r) \geq L, \\
(3) \quad & n_0 \geq \frac{1}{\varepsilon} \max\{k_S \alpha_F, M_F\}, \\
(4) \quad & n_0 \geq -\frac{2}{\delta} \log \alpha_F.
\end{align*} \tag{2.3}
\]

Denote by $E_0 := \{g \in S^{n_0} | |g|x \in B(n_0 \alpha, n_0 r)\}$. Note that the item (1) above is equivalent to $\mathbb{P}(L_{n_0} \in E_0) \geq e^{-n_0(\beta + \delta/2)}$. To make the notation lighter, for $n \geq m$, denote $L^{n}_m = Y_n \ldots Y_m$. Now for every $k \geq 1$ and for $1 \leq i \leq k$, we set
\[
\Omega_k := \left\{ \frac{1}{k n_0 + (k - 1) k_S} |L_{k n_0 + (k-1) k_S}|x \in B(\alpha, r + \varepsilon) \right\},
\]
\[
\hat{E}_i := \{L_{i(n_0 + k_S)} \in E_0\},
\]
\[
\hat{F}_i := \{L_{i(n_0 + k_S)} \in V_f, \text{ where, } f = f(L_{i(n_0 + k_S)}, L_{i(n_0 + (i-1) k_S)}(i-1) k_S)\}.
\]

We described these events because when all events $\hat{E}_i$ and $\hat{F}_i$ occur, then Lemma 2.5 applies. The following inclusion of events holds true in view of subadditivity of
displacement functionals, (2.2) and the choice of \( n_0 \in \mathbb{N} \) (namely items (2) and (3) above):
\[
\Omega_k \supseteq \bigcap_{i=0}^{k-1} \hat{E}_i \cap \bigcap_{i=1}^{k-1} \hat{F}_i
\]
As a consequence denoting by \( \mathcal{F}_E \) the \( \sigma \)-algebra generated by the random variables used in the events \( \hat{E}_i \)'s, we obtain \( \mathbb{P} \)-a.s.
\[
\mathbb{P}(\Omega_k) \geq \mathbb{E}[\mathbb{E}[\Pi_{i=0}^{k-1} \hat{E}_i \Pi_{i=1}^{k-1} \hat{F}_i | \mathcal{F}_E]] = \mathbb{E}[\Pi_{i=0}^{k-1} \hat{E}_i] \mathbb{E}[\Pi_{i=1}^{k-1} \hat{F}_i | \mathcal{F}_E]] \quad (2.4)
\]
On the other hand, by definition of the events \( \hat{F}_i \)'s and \( \alpha_F > 0 \), we readily have \( \mathbb{P} \)-a.s. \( \mathbb{E}[\Pi_{i=1}^{k-1} \hat{F}_i | \mathcal{F}_E] \geq \alpha_F^{-1} \). Using this lower bound and the fact that \( Y_i \)'s are i.i.d. we deduce from (2.4) that we have
\[
\mathbb{P}(\Omega_k) \geq \mathbb{P}(\hat{E}_0)^k \alpha_F^k
\]
Now using items (1) and (4) above, we obtain that
\[
\frac{1}{kn_0 + (k-1)k_S} \log \mathbb{P}(\Omega_k) \geq -\beta - \delta \quad (2.5)
\]
The statement follows by setting \( m_0 = n_0 + k_S \).
\( \square \)

The following lemma crucially relies on the Lipschitz property of displacement functionals and it says that the exponential rate of decay of probabilities of events concerning the asymptotic behaviour of displacement functionals along periodic times matches the decay rate considered along all times.

**Lemma 2.6** (Filling the gaps). For every \( m \geq 1, \alpha, r \geq 0, \varepsilon > 0 \), we have
\[
- \liminf_{k \to \infty} \frac{1}{mk} \log \mathbb{P}\left( \frac{1}{mk} |L_{mk}|_x \in B(\alpha, r) \right) \geq - \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left( \frac{1}{n} |L_n|_x \in B(\alpha, r + \varepsilon) \right).
\]

**Proof.** Fix a a subset \( K \) of \( G \) such that for every \( i = 1, \ldots, m \), we have \( \mu^{\tilde{e}_i}(K) \geq \frac{1}{4} \) and set \( \max_{g \in K} |g|_x =: C_K \in \mathbb{R} \). This is possible since the sequence of increasing closed sets \( \{ g \in G \mid |g|_x \leq n \} \) covers \( G \). Setting \( \beta := - \liminf_{k \to \infty} \frac{1}{mk} \log \mathbb{P}(\frac{1}{mk} |L_{mk}|_x \in B(\alpha, r)) \), for \( \delta > 0 \), choose \( k_0 \) large enough so that for every \( k \geq k_0 \), we have
\[
- \frac{1}{mk} \log \mathbb{P}(\frac{1}{mk} |L_{mk}|_x \in B(\alpha, r)) \geq \beta - \delta.
\]
The following inclusion of events follows from the Lipschitz property of displacement functionals and the choice of \( k_0 \): for every \( n \geq mk_0 \), denoting by \( k_n \) the integer satisfying \( m(k_n + 1) > n \geq mk_n \), we have
\[
\left\{ \frac{1}{n} |L_n|_x \in B(\alpha, r + \varepsilon) \right\} \supseteq \left\{ \frac{1}{mk_n} |L_{mk_n}|_x \in B(\alpha, r) \right\} \cap \{ Y_n \ldots Y_{mk_n+1} \in K \}.
\]
Evaluating the probabilities and using the independence of \( Y_i \)'s, we get
\[
\mathbb{P}\left( \frac{1}{n} |L_n|_x \in B(\alpha, r + \varepsilon) \right) \geq \mathbb{P}\left( \frac{1}{mk_n} |L_{mk_n}|_x \in B(\alpha, r) \right) \mathbb{P}(Y_n \ldots Y_{mk_n+1} \in K) \quad (2.6)
\]
Since \( Y_i \)'s are identically distributed, we have \( \mathbb{P}(Y_n \ldots Y_{mk_n+1} \in K) \geq \frac{1}{2} \). Taking the logarithm, dividing by \( n \) and passing to \( \liminf \) on both sides of (2.6), the claim follows since \( \delta > 0 \) is arbitrary. \( \square \)
In the proof of Theorem [A] we shall make use of the following general criterion (see also the subsequent remark) for the existence of a LDP:

**Theorem 2.7** (see Theorem 4.1.11 in [DZ02]). Let \( Y \) be a topological space endowed with its Borel \( \sigma \)-algebra \( \beta Y \), and \( Z_n \) be a sequence of \( Y \)-valued random variables. Denote by \( \mu_n \) the distribution of \( Z_n \). Let \( A \) be a base of open sets for the topology of \( Y \). For each \( \alpha \in Y \), define:

\[
I_l(\alpha) := \sup_{A \in A} \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(A) \quad \text{and} \quad I_r(\alpha) := \sup_{A \in A} \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(A)
\]

Suppose that for all \( \alpha \in Y \), we have \( I_l(\alpha) = I_r(\alpha) \). Then, the sequence \( Z_n \) satisfies a LDP with rate function \( I \) given by \( I(\alpha) := I_l(\alpha) = I_r(\alpha) \).

**Remark 2.8.** In our setting where \( Y = [0, \infty) \), the hypothesis of the previous theorem is actually equivalent to the existence of a weak LDP (see [DZ02]). In particular, one readily sees that a rate function of a LDP is lower semi-continuous.

We are now ready to elucidate

**Proof of Proposition 2.3.** To prove the statement, taking \( Y = [0, \infty) \), \( Z_n \) to be the random variables \( \frac{1}{n} |L_n|_x \) in Theorem 2.7, we need (see Remark 2.8) to show that \( I_l(\alpha) = I_r(\alpha) \) for every \( \alpha \in [0, \infty) \). Suppose this is not the case for some \( \alpha \in [0, \infty) \).

This means that there exists \( r_1 > 0 \) such that for every \( r_1 > r_0 > 0 \), we have

\[
\beta_1 := -\liminf_{n \to \infty} \frac{1}{n} \log P \left( \frac{1}{n} |L_n|_x \in B(\alpha, r_1) \right) > -\limsup_{n \to \infty} \frac{1}{n} \log P \left( \frac{1}{n} |L_n|_x \in B(\alpha, r_0) \right) =: \beta_2.
\]

Fix \( r_0 \in (0, r_1) \). Applying Lemma 2.4 with \( r = r_0 \), some \( \varepsilon \in (0, \frac{r_1 - r_0}{2}) \) and \( \delta \in (0, \beta_1 - \beta_2) \), we deduce that for some \( m_0 \in \mathbb{N} \) and for every \( k \geq 1 \), we have

\[
\beta_2 > \frac{1}{km_0} \log P \left( \frac{1}{km_0} |L_{km_0}|_x \in B(\alpha, r_0 + \varepsilon) \right) - \delta.
\]

Now applying Lemma 2.6 with \( m = m_0 \) and \( r = r_0 + \varepsilon \), since \( r_0 + 2\varepsilon < r_1 \), we deduce that

\[
\beta_2 \geq \frac{1}{n} \log P \left( \frac{1}{n} |L_n|_x \in B(\alpha, r_0 + 2\varepsilon) \right) - \delta \geq \beta_1 - \delta,
\]

in contradiction with the choice of \( \delta \). This proves the proposition. \( \square \)

### 2.3. Convexity of the rate function

Here we prove

**Proposition 2.9.** Keep the assumptions of Theorem 2.2. The rate function \( I : [0, \infty) \to [0, \infty] \) given by Proposition 2.3 is convex.

Recall that it readily follows from the expression of the rate function given by Theorem 2.7 that \( I \) is lower semi-continuous (see Remark 2.8). The previous result shows in particular that \( I \) is continuous in the interior of its effective support.

The proof uses the expression of the rate function \( I \) as in Theorem 2.7 (see also Remark 2.8) and the Unfolding Lemma (Lemma 2.5) in a crucial way.
Proof. Since the rate function \( I : [0, \infty) \to [0, \infty] \) is lower semi-continuous, to show that it is convex, it suffices to show that for every \( \alpha_1, \alpha_2 \in [0, \infty) \) and \( \varepsilon, \delta > 0 \), we have

\[
I \left( \frac{\alpha_1 + \alpha_2}{2} \right) - \delta \leq \left( \frac{1}{2} + \varepsilon \right) (I(\alpha_1) + I(\alpha_2)).
\]

(2.7)

So, let \( \alpha_2 > \alpha_1 \in [0, \infty) \), \( \delta > 0 \) be given. We will suppose that \( \alpha_1 \neq 0 \); this case can be treated similarly as below.

Fix some \( 0 < r < \min\{\alpha_1, \alpha_2\} \) and some positive constant \( \varepsilon < \min\{\frac{\varepsilon}{\alpha_2 - \alpha_1}, 1\} \). Let \( F \) be the finite set, \( (V_f)_{f \in F} \) be the neighborhoods and \( k_S, L > 0 \) be the constants given by Lemma 2.5. Set \( \max_{f \in F} |f'|_x := M_F \in \mathbb{R} \), where \( V = \bigcup_{f \in F} V_f \). By Lemma 2.5, we have \( \mu^{k_S} V_f > 0 \) for every \( f \in F \); set \( \alpha_F := \min_{f \in F} \mu^{k_S} V_f \). We have \( \alpha_F > 0 \). Given any \( k \)-tuple \( (g_k, \ldots, g_1) \in G^k \) with \( |g_i|_x \geq L \) as in Lemma 2.5, using that lemma, choose elements \( f(g_i, g_{i-1}) \in F \) so that (2.2) holds.

In view of Remark 2.8, by Proposition 2.3 and Theorem 2.7, there exists \( n_0 \in \mathbb{N} \) large enough so that for \( j = 1, 2 \)

\[
e^{-n_0\alpha_j + \delta/2} \leq \mathbb{P} \left( \frac{1}{n_0} |L_{n_0}|_x \in B(\alpha_j, r) \right),
\]

(2.8)

and

1. \( n_0(\alpha_1 - r) \geq L \),
2. \( n_0 \geq 2\frac{1}{r} \max\{M_F, k_S(\alpha_1 + \alpha_2)\} \),
3. \( n_0 \geq \frac{2}{3} \log \alpha_F \).

For \( j = 1, 2 \), let us denote by \( E_j \) the subset of \( G \) given by \( \{g \in S^{n_0} | \frac{1}{n_0} |g|_x \in B(\alpha_j, r)\} \). For \( i \) an integer, denote \( j(i) = 1 \) if \( i \) is odd and \( j(i) = 2 \) if \( i \) is even. For \( n \geq m \), using the same notation \( L_n^m = Y_n \ldots Y_m \), as before, consider the following events for every \( k \geq 1 \) and \( 1 \leq i \leq k \):

\[
\Omega_k := \left\{ \frac{1}{k n_0 + (k - 1) k_S} |L_{k n_0 + (k - 1) k_S}|_x \in B \left( \frac{\alpha_1 + \alpha_2}{2}, 2r \right) \right\},
\]

\( \hat{E}_i := \{L_{(i+1)n_0+kS}^{(i)n_0+kS} \in E_{j(i)}\} \),

\( \hat{F}_i := \{L_{(i)n_0+(i-1)kS+1}^{(i)n_0+(i-1)kS} \in V_f \) where \( f = f(L_{(i+1)n_0+i^2}^{(i)n_0+(i-1)kS}) \).

Using subadditivity of displacement functionals, (2.2), it is not hard to verify that the following inclusion of events holds true for every \( k \geq 1 \), thanks to the choice of \( n_0 \) (i.e. satisfying (1) and (2) above):

\[
\Omega_k \supseteq \bigcap_{i=1}^{k-1} \hat{F}_i \cap \bigcap_{i=1}^k \hat{E}_i.
\]

(2.9)

Now let us denote by \( \mathcal{F}_E \) the \( \sigma \)-algebra generated by the random variables \( Y_i \)'s for \( i = 1, \ldots, k n_0 + (k - 1) k_S \) except those \( i \in \bigcup_{t=1}^{k-1} \{t(n_0 + k_S), \ldots, t n_0 + (t - 1) k_S + 1\} \). Using independence of \( Y_i \)'s, evaluating the probabilities on both sides of (2.9) we obtain \( \mathbb{P} \)-a.s.

\[
\mathbb{P}(\Omega_k) \geq \prod_{i=1}^k \mathbb{E}[1_{\hat{E}_i} \mathbb{E}[\Pi_{i=1}^{k-1} 1_{\hat{F}_i} | \mathcal{F}_E]].
\]

(2.10)
As in the proof of Proposition 2.3, using the \( \mathbb{P} \)-a.s. bound \( \mathbb{E}[\prod_{i=1}^{k-1} \tilde{F}_i | \mathcal{F}_E] \geq \alpha_F^{k-1}, \) (2.10) reads

\[
\mathbb{P}(\Omega_k) \geq \alpha_F^{k-1} \prod_{i=1}^{k} \mathbb{P}(\tilde{E}_i) \tag{2.11}
\]

Using the fact that \( Y_i \)'s are identically distributed, (2.8) implies that for every \( i \), we have \( \mathbb{P}(\tilde{E}_i) = \mathbb{P}(L_n \in E_{j(i)}) \geq e^{-n_0(I(\alpha_i) + \delta/2)}, \) so that (2.11) reads

\[
\mathbb{P}(\Omega_k) \geq \alpha_F^{k-1} e^{-n_0 |j^{-1}(1)| (I(\alpha_1) + \delta/2)} e^{-n_0 |j^{-1}(2)| (I(\alpha_2) + \delta/2)}. \]

Note that \( |j^{-1}(i)| \leq k/2 + 1 \). By (3) in the choice of \( n_0 \) above, we deduce that for every \( k \) large enough, we have

\[
\frac{1}{kn_0 + (k-1)k_S} \log \mathbb{P}(\Omega_k) \geq -\frac{1}{2} (I(\alpha_1) + I(\alpha_2)) - \varepsilon (I(\alpha_1) + I(\alpha_2)).
\]

Now applying Lemma 2.12 with \( m = n_0 + k_S, \alpha = \frac{\alpha_1 + \alpha_2}{2} \) and 2\( r \), and using the expression of \( I \) given by Theorem 2.7, we deduce

\[
I \left( \frac{\alpha_1 + \alpha_2}{2} \right) - \delta \leq \left( \frac{1}{2} + \varepsilon \right) (I(\alpha_1) + I(\alpha_2)).
\]

This shows (2.7) and proves the proposition.

\( \square \)

2.4. Proof of Theorem 2.2. The rest of this section is devoted to completing the proof of Theorem 2.2. It remains to show that the (full) LDP holds under a finite exponential moment condition and that we can give an alternative expression for the rate function under a strong exponential moment condition. These will follow using rather standard techniques in large deviation theory. We spell out the details for reader’s convenience.

2.4.1. Existence of (full) LDP under exponential moment condition. The following classical notion of large deviations theory allows one to formulate a sufficient condition (see Lemma 2.11) to strengthen a weak LDP to a LDP with proper rate function:

**Definition 2.10.** A sequence of random variables \( Z_n \) on a topological space \( Y \) is said to be exponentially tight, if for all \( \alpha \in \mathbb{R} \), there exists a compact set \( K_\alpha \subset Y \) such that \( \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n \in K_\alpha) < -\alpha \).

With this definition, we have

**Lemma 2.11 ([DZ02]).** If an exponentially tight sequence of random variables on \( X \) satisfies a weak LDP with a rate function \( I \), then it satisfies a (full) LDP with a proper rate function \( I \).

Therefore, by Proposition 2.3, to prove the full LDP and properness of \( I \) in Theorem 2.2, we only need to show that a finite exponential moment condition on \( \mu \) implies that the sequence \( \frac{1}{n} |L_n|_x \) of random variables is exponentially tight.

**Lemma 2.12.** If \( \mu \) has a finite exponential moment, then the sequence random variables \( \frac{1}{n} |L_n|_x \) is exponentially tight.
Proof. It suffices to show that
\[
\lim_{t \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} |L_n|_x \geq t \right) = -\infty
\]
By Chebyshev inequality, for every \( s \geq 0 \), we have
\[
\mathbb{P}(|L_n|_x \geq tn) \leq \mathbb{E}[e^{s|L_n|_x}] e^{-stn}
\]
In this inequality, taking log, dividing by \( n \) and specializing to some \( s_0 \in \mathbb{R} \) such that \( c \geq s_0 > 0 \), we get
\[
\frac{1}{n} \log \mathbb{P}(|L_n|_x \geq tn) \leq -(s_0 t - \frac{1}{n} \log \mathbb{E}[e^{s_0|L_n|_x}])
\]
On the other hand, it follows by the independence of random walk increments and the subadditivity of the displacement functional \(|.|_x\) that for all \( n \geq 1 \), we have
\[
\frac{1}{n} \log \mathbb{E}[e^{s_0|L_n|_x}] \leq \log \mathbb{E}[e^{s_0|Y_1|_x}].
\]
Therefore, we obtain
\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} |L_n|_x \geq t \right) \leq -(s_0 t - \mathbb{E}[e^{s_0|Y_1|_x}])
\]
Since \( \mathbb{E}[e^{s_0|Y_1|_x}] \) is finite by the exponential moment condition, the result follows by taking limit in both sides as \( t \) goes to \( +\infty \). \( \square \)

2.4.2. Identification of the rate function. Using Fenchel–Moreau duality and Varadhan’s integral lemma, we now give an alternative expression for the rate function \( I \) as the Legendre transform of a limit Laplace transform of the distributions of \(|L_n|_x\).

We now complete the

Proof of Theorem 2.2 (Identification of the rate function). It is easy to see that if \( \mu \) has a strong exponential moment, then we have \( \Lambda(\lambda) < \infty \) for all \( \lambda \in \mathbb{R} \). Hence, it follows from Varadhan’s integral lemma (see [DZ02] section 4.3) that in fact for all \( \lambda \in \mathbb{R} \), one has
\[
\Lambda(\lambda) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{\lambda|L_n|_x}] = \sup_{\alpha \in \mathbb{R}} (\lambda \alpha - I(\alpha))
\]
where \( I \) is the proper rate function of the LDP.

Now, for a function \( f \) on \( \mathbb{R} \), denote its convex conjugate (Legendre transform) by \( f^*(.) \), where \( f^*(\lambda) := \sup_{\alpha \in \mathbb{R}} (\lambda \alpha - f(\alpha)) \). The above conclusion of Varadhan’s integral lemma hence reads as \( \Lambda(\lambda) = I^*(\lambda) \). Now, since \( I \) is a convex rate function, Fenchel-Moreau duality tells us that \( I(\alpha) = I^{**}(\alpha) = \Lambda^*(\alpha) \), identifying \( I(\alpha) \) with \( \Lambda^*(\alpha) \) and completing the proof. \( \square \)

2.5. The value of the rate function at zero. To end this section, here we briefly spell out the details of 3. of Remark [1.3] We shall say that a proper metric space \((X,d)\) has at most exponential doubling property if there exists a constant \( \alpha > 1 \) such that for every \( R > 0 \) large enough, any ball of radius \( 2R \) can be covered with at most \( \alpha^R \) balls of radius \( R \). This condition is clearly satisfied for every finite dimensional Riemannian manifold with pinched negative curvature.

**Lemma 2.13.** Let \((X,d)\) be a hyperbolic space with at most exponential doubling property. Let \( \mu \) be a finitely supported symmetric non-elementary probability measure on \( \text{Isom}(X) \) and suppose that the group \( G \) generated by the support of \( \mu \) acts properly
discontinuously on $X$. Let $I$ be the rate function given by Theorem [A]. Then, we have $I(0) = -\log r_\mu$, where $r_\mu$ is the spectral radius of the random walk.

Note since $\mu$ is non-elementary, its support generates a non-amenable countable subgroup of $\text{Isom}(X)$ and by Kesten’s criterion [Kes59a], we have $I(0) > 0$.

Proof. By applying at most exponential doubling property, we find that for some $R > 0$ and $x \in X$, we have $|\{g \in G \mid gx \in B(x, nR)\}| = O(\alpha^{nR})$. Denote by $\lambda(\mu)$ the self-adjoint operator on $\ell^2(G)$ defined by $\lambda(\mu)f(x) = \sum_\gamma f(g^{-1}x)\mu(g)$. We have $\lambda(\mu^\ast n) = \lambda(\mu)^n$ and by Cauchy-Schwarz inequality, $\mu^{2n}(g) \equiv \langle \lambda(\mu)^{2n}\delta_g, \delta_\epsilon \rangle \leq ||\lambda(\mu)\delta_g|| ||\lambda(\mu)\delta_\epsilon|| = \mu^{2n}(\epsilon)$. Now using this and Kesten’s formula, we deduce that for every $\varepsilon > 0$, $\log r_\mu \leq \limsup_{n \to \infty} \frac{1}{2n} \log \mathbb{P}(|L_{2n}|x \leq 2nR\varepsilon) \leq \log r_\mu + R\varepsilon$, which shows that for the rate function $I$ given by Theorem [A] we have $I(0) = -\log r_\mu$. \qed

It is easy to see that this argument also gives a lower bound to the drift of the random walk. Such relations between drift, spectral radius and entropy of the random walk, and growth rate of the group are well-known [Ave76, Gui80].

3. Large Deviation Principle for Translation Distance

This section is devoted to the proof of Theorem [B]. The proof will be carried out in three parts consecutively in §3.2 §3.3 and §3.4. In the first part (§3.1), we start with developing the necessary geometric ingredients.

3.1. Geometric ingredients. Let $(X, d)$ be a hyperbolic space, with a fixed base-point $x \in X$, and let $S$ be an arbitrary bounded subset of $\text{Isom}(X)$.

Some of the results in this subsection might be known to experts, but we provide detailed proofs for completeness. Here, the key results that will be used for the probabilistic arguments are Lemmas 3.5 and 3.6. We start with some preliminary results to prepare the proofs of those lemmas.

Remark 3.1. For convenience, in the proofs below we will assume that $d(x, sx) < 1$ for each $s \in S$. This can be achieved by rescaling $X$, and it is readily seen that all the statements hold for $X$ if and only if they hold for a rescaling of $X$, up to changing the constants.

Lemma 3.2. For every $\varepsilon > 0$ there exist $D_0, N \geq 1$ so that the following holds. Let $s_1, \ldots, s_n \in S$, for some $n \geq N$, and let $g_i = s_1 \ldots s_i$. Let $\gamma$ be a subpath of length $\geq \varepsilon n$ of a geodesic from $x$ to $g_n x$. Then there exists $i$ with $d(g_i x, \gamma) \leq D_0$.

Proof. This can be deduced from [HS17, Claim 2 within Lemma 2.6], which in our setting says the following. There exist $\epsilon_0 > 0$ and $D' > 0$ (independent of $x$ and $g_n$) so that, given disjoint balls $B_1, \ldots, B_k$ of radius $D \geq D'$ centered on $\gamma$, any path $\alpha$ from $x$ to $g_n x$ that avoids all $B_i$ satisfies $l(\alpha) \geq k(1 + \epsilon_0)^D$. Choose $D_0 \geq D' + 1$ so that $(1 + \epsilon_0)^{D_0 - 1} > 3D_0/\varepsilon$. Also, we let $N \geq 6D_0/\varepsilon$, and check that these choices work. In the setting of the statement, suppose by contradiction that we have $d(g_i x, \gamma) > D_0$ for all $i$. Then we can find at least $k \geq \varepsilon n/(2D_0) - 1 \geq \varepsilon n/(3D_0)$ disjoint balls $B_i$ of radius $D_0 - 1$ centered on $\gamma$ so that the path $\alpha$ in $X$ obtained concatenating geodesics from $s_i x$ to $s_{i+1} x$ avoids all $B_i$. The length of $\alpha$ is at most $n$, so we obtain:

$$n \geq \frac{\varepsilon n}{3D_0}(1 + \epsilon_0)^{D_0 - 1} > n.$$
Let $\delta \geq 1$ be a hyperbolicity constant for $X$. For $g \in G$, define

$$
\text{Min}(g) = \{z \in X : d(z, gz) \leq \tau(g) + 4\delta\}.
$$

Also, for $z \in Z$, denote by $\pi^g(z)$ a point in $\text{Min}(g)$ so that $d(z, \pi^g(z)) \leq d(z, \text{Min}(g)) + 1$. (That is, $\pi^g$ is coarsely the closest-point projection to $\text{Min}(g)$.) We can and will assume that $g\pi^g(z) = \pi^g(gz)$ holds for all $g$ and $z$.

It is known that $\text{Min}(g)$ is quasiconvex (see e.g. [DG08, Proposition 2.3.3] and [Cou14, Proposition 2.28]), but we will only need the following special case of quasiconvexity, which has a very short proof:

**Lemma 3.3.** If $y \in \text{Min}(g)$, then any point on any geodesic from $y$ to $gy$ is also contained in $\text{Min}(g)$.

**Proof.** First, observe that given $y \in X$ and a geodesic $[y, gy]$, any $z \in [y, gy]$ has

$$
d(z, gz) \leq d(z, gy) + d(gy, gz) = d(z, gy) + d(y, z) = d(y, gy).
$$

The desired statement easily follows. \qed

We now show that geodesics from $z$ to $gz$ pass close to the projection points of the endpoints onto $\text{Min}(g)$.

**Lemma 3.4.** There exists $D_1 \geq 0$ so that the following holds. For every $g \in G$ and $z \in X$, we have that any geodesic $\gamma$ from $z$ to $gz$ passes $D_1$-close to $\pi^g(z)$ and $\pi^g(gz)$. Moreover, we have

$$
d(z, gz) \geq 2d(z, \pi^g(z)) + \tau(g) - D_1.
$$

**Proof.** Consider any geodesic $\gamma$ from $z$ to $gz$. We will show that $\gamma$ passes $(4\delta + 2)$-close to $\pi^g(z)$, the argument for $\pi^g(gz)$ being similar. We will use $2\delta$-thinness of a quadrangle with vertices $z, \pi^g(z), \pi^g(gz), gz$.

Suppose by contradiction that $\gamma$ does not pass $(4\delta + 2)$-close to $\pi^g(z)$. Consider the point $z'$ on a geodesic from $z$ to $\pi^g(z)$ at distance $2\delta + 2$ from $\pi^g(z)$. We observe that $z'$ cannot be $2\delta$-close to any geodesic $[\pi^g(z), \pi^g(gz)]$, for otherwise there would be a point $q$ on said geodesic, whence on $\text{Min}(g)$ by Lemma 3.3, which satisfies $d(z, q) < d(z, \pi^g(z)) - 1$, contradicting the defining property of $\pi^g$. Also, $z'$ cannot be $2\delta$-close to $\gamma$ by hypothesis, so $z'$ is $2\delta$-close to $g[z, \pi^g(z)]$. But then it must be $4\delta$-close to the point on that geodesic at distance $4\delta + 2$ from $\pi^g(gz)$, this point being $gz'$. We just showed $d(z', gz') \leq 4\delta$, which implies $z' \in \text{Min}(g)$. But $d(z, z') < d(z, \pi^g(z)) - 1$, contradicting the defining property of $\pi^g(z)$.

Now, the fact that $\gamma$ passes $(4\delta + 2)$-close to $\pi^g(z)$ and $\pi^g(gz)$ implies the following inequality:

$$
d(z, gz) \geq d(z, \pi^g(z)) + d(\pi^g(z), \pi^g(gz)) + d(\pi^g(gz), gz) - 4(4\delta + 2).
$$

The first and third terms on the right-hand side are both equal to $d(z, \pi^g(z))$, while the second term is at least $\tau(g)$. Therefore, we can conclude by setting $D_1 = 4(4\delta + 2)$. \qed

We are now ready to prove the following key lemma which will be instrumental in showing that for a fixed speed $\alpha > 0$, up to multiplying its probability by the linear factor $n$, the event $|L_n|_x \sim \alpha n$ is at least as likely as the event $\tau(L_n) \sim \alpha n$. \hfill \Box
Lemma 3.5. For each \( \epsilon > 0 \) there exist \( D, N \geq 1 \) so that the following holds. Let \( s_1, \ldots, s_n \in S \), for some \( n \geq N \), and let \( g_i = s_1 \ldots s_i \) and \( r_i = s_{i+1} \ldots s_n \). Suppose that \( \tau(g_n) \geq \epsilon n \). Then there exists \( i \) so that \( |r_i g_i| - \tau(g_n)| \leq D \).

Proof. Assuming that \( \epsilon n \) is sufficiently large compared to the constant \( D_1 \) of Lemma 3.3, we see that any geodesic from \( x \) to \( g_n x \) contains a subgeodesic \( \gamma \) of length \( \geq \epsilon n / 2 \) with endpoints \( D_1 \)-close to \( \operatorname{Min}(g_n) \). In view of Lemma 3.3 and \( 2\delta \)-thinness of quadrangles, any point \( p \) on \( \gamma \) is \( (D_1 + 2\delta) \)-close to \( \operatorname{Min}(g_n) \), and in particular \( p \) satisfies
\[
d(p, g_n p) \leq \tau(g_n) + 4\delta + 2(D_1 + 2\delta).
\]
In view of Lemma 3.2 (applied with \( \epsilon / 2 \) replacing \( \epsilon \), and denoting \( D_0 \) the corresponding constant), we have \( d(g_i x, p) \leq D_0 \) for some \( p \in \gamma \), provided that \( n \) is large enough. We now have:
\[
|r_i g_i|_x = d(x, r_i g_i x) = d(g_i x, g_n g_i x) \leq d(p, g_n p) + 2D_0 \leq \tau(g_n) + 2(D_1 + D_0) + 8\delta.
\]
On the other hand, \( |r_i g_i|_x \geq \tau(r_i g_i) = \tau(g_n) \), where the equality holds since \( r_i g_i \) and \( g_n \) are conjugate. This concludes the proof. \( \square \)

The following useful lemma will be used to show that the exponential rate of decay of the event that \( |L_n| \) stays sublinear is comparable to the event that \( \tau(L_n) \) stays sublinear.

Lemma 3.6. For each \( \epsilon > 0 \) there exists \( N \geq 1 \) so that the following holds. Let \( s_1, \ldots, s_n \in S \), for some \( n \geq N \), and let \( g_i = s_1 \ldots s_i \) and \( r_i = s_{i+1} \ldots s_n \). Then for every \( r \in [\tau(g_n), |g_n|_x] \) there exists \( i \) so that \( |r_i g_i|_x - r| \leq \epsilon n \).

Proof. First, notice that for \( r \geq |g_n|_x - \epsilon n \) we can just choose \( i = n \), so in the arguments below we assume \( r \leq |g_n|_x - \epsilon n \).

Set \( d = (r - \tau(g_n))/2 \), and assume that \( n \) is larger than the \( N \) from Lemma 3.2 with \( \epsilon / 4 \) replacing \( \epsilon \). We impose further constraints on \( n \) later.

Claim: If \( n \) is sufficiently large, then we can find a subgeodesic \( \gamma' \) of length \( \epsilon n / 4 \) of a geodesic \( \gamma' \) from \( x \) to \( g_n x \) so that any \( p \in \gamma \) has

- \( d(p, q) \leq D_1 + \delta \) for some \( q \) on a geodesic from \( x \) to \( \pi^{g_n}(x) \),
- \( d(p, \pi^{g_n}(x)) - d \leq \epsilon n / 3 \).

Proof. We let \( p_0 \) be the point along \( \gamma' \) so that
\[
d(x, p_0) = d(x, \pi^{g_n}(x)) - d - \epsilon n / 4 = \hat{d},
\]
and we let \( \gamma \) be the subgeodesic of \( \gamma' \) of length \( \epsilon n / 4 \) with starting point \( p_0 \). We now check that, for \( n \) large enough, this is all well-defined, and that \( \gamma \) has the required property.

Let us make the preliminary observation that
\[
|g_n|_x \leq d(x, \pi^{g_n}(x)) + d(\pi^{g_n}(x), \pi^{g_n}(g_n x)) + d(\pi^{g_n}(g_n x), g_n x) \leq 2d(x, \pi^{g_n}(x)) + \tau(g_n) + 4\delta.
\]
Observe now that we have
\[
d \leq (|g_n|_x - \epsilon n - \tau(g_n))/2 \leq d(x, \pi^{g_n}(x)) + 4\delta - \epsilon n / 2,
\]
implying \( \hat{d} \geq d - 4\delta + \epsilon n / 4 \), which is a positive quantity if \( n \) is sufficiently large.

Also, again by Lemma 3.3, there exists \( p' \in \gamma' \) so that \( d(p', \pi^{g_n}(x)) \leq D_1 \); denote by \( \gamma'' \) the initial subgeodesic of \( \gamma' \) with terminal point \( p' \).
Notice that for $n$ large enough we have

$$d(x, p') \geq d(x, \pi^{g_n}(x)) - D_1 = d + \hat{d} + \epsilon n/4 - D_1 \geq \hat{d} + \epsilon n/4.$$ 

The inequalities we just showed imply that $p_0$ and $\gamma$ are well-defined and, furthermore, that $\gamma$ is a subgeodesic of $\gamma''$.

Considering a triangle with vertices $x, p', \pi^{g_n}(x)$ and containing $\gamma''$, we see that any point on $\gamma''$, whence any point on $\gamma$, is $(D_1 + \delta)$-close to a point on a geodesic from $x$ to $\pi^{g_n}(x)$. In particular, for any $p \in \gamma$ we have

$$|d(p, \pi^{g_n}(x)) + d(x, p) - d(x, \pi^{g_n}(x)| \leq 2D_1 + 2\delta,$$

and hence

$$|d(p, \pi^{g_n}(x)) - d|(x, \pi^{g_n}(x)) - d(x, p) - d| + 2D_1 + 2\delta \leq |d(x, \pi^{g_n}(x)) - d(x, p_0) - d - d(p_0, p)| + 2D_1 + 2\delta \leq \epsilon n/4 + 2D_1 + 2\delta.$$

Provided that $n$ is large enough, this concludes the proof of the claim. □

By Lemma 3.2, there exists $i$ so that we have $d(g_i, p) \leq D_0$ for some $p \in \gamma$, and by the claim we have $d(p, q) \leq D_0 + 2D_1 + 2\delta$, whence $d(g_i, q) \leq D_0 + D_1 + \delta$, for some $q$ on a geodesic from $x$ to $\pi^{g_n}(x)$. Notice that we can assume that $\pi^{g_n}(q) = \pi^{g_n}(x)$ since

$$d(q, \pi^{g_n}(x)) = d(z, \pi^{g_n}(x)) - d(q, x) \leq d(x, Min(g_n)) + 1 - d(q, x) \leq d(q, Min(g_n)) + 1.$$

In particular, in view of Lemma 3.4, we have

$$d(q, g_nq) \geq 2d(q, \pi^{g_n}(x)) + \tau(g_n) - D_1.$$

We can now compute

$$|r_i g_i|_x = d(x, r_i g_i x) = d(g_i x, g_n g_i x) \geq d(q, g_n q) - 2(D_0 + D_1 + \delta) \geq 2d(q, \pi^{g_n}(x)) + \tau(g_n) - 3(D_0 + D_1 + \delta).$$

Hence,

$$|r_i g_i|_x \geq 2d + \tau(g_n) - 2\epsilon n/3 - 5(D_0 + D_1 + \delta) = r - 2\epsilon n/3 - 5(D_0 + D_1 + \delta).$$

For $n$ sufficiently large, this last quantity is $\geq r - \epsilon n$.

On the other hand, we also have

$$|r_i g_i|_x = d(g_i x, g_n g_i x) \leq d(p, \pi^{g_n}(x)) + d(\pi^{g_n}(x), \pi^{g_n}(g_n x)) + d(g_n p, \pi^{g_n}(g_n x)) + 2D_0 \leq 2d + 2\epsilon n/3 + 2D_1 + \tau(g_n) + 4\delta + 2D_0 \leq r + 2\epsilon n/3 + 2D_0 + 2D_1 + 4\delta.$$

For $n$ sufficiently large, this last quantity is $\leq r + \epsilon n$, and we are done. □

We now proceed with the proof of Theorem 13.

Following Theorem 2.7 for $\alpha \in [0, \infty)$, we set

$$J_{\text{ls}}(\alpha) := \sup_{r \to 0} - \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \tau(L_n) \in B(\alpha, r) \right)$$

and

$$J_{\text{ls}}(\alpha) := \sup_{r \to 0} - \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \tau(L_n) \in B(\alpha, r) \right).$$

Let $I : [0, \infty) \to [0, \infty]$ be the convex rate function given by Theorem 2.2.

In §3.2 we show that for every $\alpha > 0$, $I(\alpha) \geq J_{\text{li}}(\alpha)$. In §3.3 we show that for every $\alpha > 0$, $J_{\text{ls}}(\alpha) \geq I(\alpha)$. And in §3.4, we show that $I(0) = J_{\text{ls}}(0) = J_{\text{li}}(0)$ and this concludes the proof of Theorem 13 in view of Theorem 2.7.
3.2. One direction: translation distance at least as likely as displacement functional (at the exponential scale). The following is a straightforward consequence of Lemma 2.5.

**Lemma 3.7.** There exist a constant \( k_S \in \mathbb{N} \), a finite set \( \mathcal{F} \subset S^{k_S} \), for every \( f \in \mathcal{F} \) a neighborhood \( V_f \), and a constant \( L > 0 \) such that for every \( g \in \text{Isom}(X) \), there exists \( f \in \mathcal{F} \) with \( \tau(f'|g) \geq |g|_x - L \) for every \( f' \in V_f \). \( \square \)

Note that here, the constant \( L \) accounts for the lack of hypothesis on the size of \( |g|_x \) as in Lemma 2.5. Indeed, one can take \( L \) to be the same constant given by that lemma.

This geometric ingredient is sufficient to show that the exponential rates of decay of probabilities of rare events for the lengths of translation distances of random walks are bounded above by those of displacement functionals:

**Proof of “\( I(\alpha) \geq J_{\ell}^{(\alpha)} \) for every \( \alpha \geq 0 \” \).** Let \( \alpha \in [0, \infty) \) and \( \delta > 0 \) be given. As before, set \( M_F := \max_{f' \in \mathcal{F}} |f'|_x \), where \( \hat{V} = \cup_{f \in \mathcal{F}} V_f \) and \( \alpha_F := \min_{f \in \mathcal{F}} \mu^{k_S}(V_f) > 0 \).

By definition of the function \( J_{\ell} : [0, \infty) \to [0, \infty] \), for every small enough \( \eta_0 > 0 \), there exists an increasing sequence \( n_\ell \) of positive integers such that for every \( \ell \geq 1 \), we have

\[
\mathbb{P}\left( \frac{1}{n_\ell} \tau(L_{n_\ell}) \in B(\alpha, \eta_0) \right) = \mathbb{P}\left( \frac{1}{n_\ell} \tau(R_{n_\ell}) \in B(\alpha, \eta_0) \right) \leq e^{-n_\ell(j_{\ell}(\alpha) - \delta)}, \tag{3.1}
\]

where the first equality follows since the left random walk \( L_n \) and the right random walk \( R_n \) have the same distribution. On the other hand, for \( n \geq m \) noting \( R^n_\ell = Y^\ell_{m+1} \ldots Y^\ell_n \), an application of Theorem 2.2 and using fact that \( Y^\ell_i \)’s are i.i.d, yield that for every \( \ell \) large enough, we have

\[
\mathbb{P}(\frac{1}{n_\ell - k_S} |R^{n_\ell}_{k_\ell}|_x \in B(\alpha, \eta_0/2)) \geq e^{-(n_\ell - k_S)(I(\alpha) + \delta)}. \tag{3.2}
\]

Using Lemma 3.7, for every \( g \in G \), let us fix an element \( f(g) \in \mathcal{F} \) such that \( \tau(f'(g)) \geq |g|_x - L \) for every \( f' \in V_f \). Now, taking \( n_\ell \geq 1 \) \( \max\{4\alpha k_S, M_F, L\} \), it follows by this lemma that we have the following inclusion of events

\[
\{R_{k_\ell} = f(R^{n_\ell}_{k_\ell})\} \cap \{\frac{1}{n_\ell - k_S} |R^{n_\ell}_{k_\ell}|_x \in B(\alpha, \eta_0/2)\} \subseteq \frac{1}{n_\ell} \tau(R_{n_\ell}) \in B(\alpha, \eta_0) \quad (3.3)
\]

Denote the events \( \frac{1}{n_\ell} \tau(R_{n_\ell}) \in B(\alpha, \eta_0) \), \( \frac{1}{n_\ell - k_S} |R^{n_\ell}_{k_\ell}|_x \in B(\alpha, \eta_0/2) \} \) and \( \{R_{k_\ell} \in V_f \text{ where } f = f(R^{n_\ell}_{k_\ell})\} \), respectively, by \( \Omega_\ell \), \( \mathcal{E}_\ell \) and \( \mathcal{F}_\ell \). Furthermore, for every \( \ell \geq 1 \), denote by \( \mathcal{E}_\ell \) the \( \sigma \)-algebra generated by \( Y_{k_\ell + 1}, \ldots, Y_{n_\ell} \). Taking expectation of the inequality \( 1_{\mathcal{E}_\ell} 1_{\mathcal{F}_\ell} \leq 1_{\Omega_\ell} \) given by (3.3), we have \( \mathbb{P} \)-a.s.

\[
\mathbb{P}(\Omega_\ell) \geq \mathbb{E}[1_{\mathcal{F}_\ell} 1_{\mathcal{E}_\ell}] = \mathbb{E}[1_{\mathcal{E}_\ell} \mathbb{E}[1_{\mathcal{F}_\ell} | \mathcal{E}_\ell]] \geq \mathbb{P}(\mathcal{E}_\ell) \alpha_F,
\]

where the last inequality follows since by construction \( \mathbb{P} \)-a.s. \( \mathbb{E}[1_{\mathcal{F}_\ell} | \mathcal{E}_\ell] \geq \alpha_F \). Combining the previous equation with (3.2), we deduce \( \mathbb{P}(\Omega_\ell) \geq e^{-(n_\ell - k_S)(I(\alpha) + \delta)} \alpha_F \).

This yields that for every \( n_\ell \) large enough,

\[
\mathbb{P}(\frac{1}{n_\ell} \tau(R_{n_\ell}) \in B(\alpha, \eta_0)) \geq e^{-n_\ell((1 + \delta)I(\alpha) + \delta)}
\]
Taking logarithm, dividing by \( n \), we deduce from (3.1) that we have

\[(1 + \delta) I(\alpha) + 2\delta \geq J_\ell(\alpha),\]

which shows that \( I(\alpha) \geq J_\ell(\alpha) \) since \( \delta > 0 \) is arbitrary. \( \square \)

3.3. The other direction: displacement functional at least as likely as translation distance (at the exponential scale). Here we show that for every \( \alpha > 0 \), \( J_\ell(\alpha) \geq I(\alpha) \). Let \( \alpha \in (0, \infty) \) be given. Fix some \( \delta > 0 \). Using Theorem 2.2 there exists \( r_0 > 0 \) such for every positive \( r < \min\{r_0, \alpha\} \), there exists \( n_0 \in \mathbb{N} \) such that

\[
\begin{align*}
(1) & \quad \frac{\log n_0}{n_0} \leq \frac{\delta}{3}, \\
(2) & \quad n_0 \geq \max\{N, \frac{D}{\alpha r}\} \quad \text{where } N \text{ and } D \text{ are as given by Lemma 3.5 with } \epsilon = \frac{1}{2}(\alpha - r), \\
(3) & \quad \mathbb{P}\left( \frac{1}{n_0} \tau(L_{n_0}) \in B(\alpha, r) \right) \geq e^{-n_0(J_\ell(\alpha)+\delta/3)}, \\
(4) & \quad \mathbb{P}\left( \frac{1}{n_0} |L_{n_0}| \in B(\alpha, 2r) \right) \leq e^{-n_0(I(\alpha)-\delta/3)}.
\end{align*}
\]

Let \( E_0 \) be the subset of the Cartesian product \( S^{n_0} \) given by

\[ E_0 = \{(s_1, \ldots, s_{n_0}) \mid \frac{1}{n_0} \tau(s_1 \ldots s_{n_0}) \in B(\alpha, r)\}. \]

By choice of \( \epsilon = \frac{1}{2}(\alpha - r) \) and \( n_0 \) with (2) above, applying Lemma 3.5 we get that for some \( D > 0 \) and for every \((s_1, \ldots, s_{n_0}) \in E_0\), there exists an integer \( i \in [1, n_0] \) such that we have

\[ ||s_{i+1} \ldots s_{n_0}s_1 \ldots s_i|| = \tau(s_1 \ldots s_n) \leq D \quad (3.4) \]

Recall that the asymptotic translation distance \( \ell(g) \) of an element \( g \in \text{Isom}(X) \) is continuous [OR13, Theorem 1.9] and the difference \( |\ell(g) - \tau(g)| \) is uniformly bounded (depending only on the hyperbolicity constant of \( X \)) [CDP90, Ch.10, Prop. 6.4]. Using this fact and the continuity of \( ||| \cdot ||| \), for each \((s_1, \ldots, s_{n_0}) \in E_0\), up to increasing \( D \) by a bounded amount (depending only on the hyperbolicity constant of \( X \)), we can fix such a choice of \( i \in [1, n_0] \) (as in (3.4)) so that the corresponding function \( p : \tilde{E}_0 \to [1, n_0] \) is measurable. Now for every integer \( k \in [1, n_0] \), let

\[ E_{0,k} = \{(s_1, \ldots, s_{n_0}) \in E_0 \mid p(s_1, \ldots, s_{n_0}) = k\}. \]

By construction, the \( E_{0,k} \) give a partition of \( E_0 \), so we have

\[ \sum_{k=1}^{n_0} \mathbb{P}( (Y_1, \ldots, Y_{n_0}) \in E_{0,k}) \geq \mathbb{P}\left( \frac{1}{n_0} \tau(R_{n_0}) \in B(\alpha, r) \right) \geq e^{-n_0(J_\ell(\alpha)+\delta/3)}. \]

Consequently, there exists some integer \( k_0 \in [1, n_0] \) such that we have

\[ \mathbb{P}( (Y_1, \ldots, Y_{n_0}) \in E_{0,k_0}) \geq \frac{1}{n_0} e^{-n_0(J_\ell(\alpha)+\delta/3)} \geq e^{-n_0(J_\ell(\alpha)+\delta)}, \quad (3.5) \]

where the last inequality follows by (1) in the choice of \( n_0 \).

On the other hand, by (3.4), for every \((s_1, \ldots, s_{n_0}) \in E_{0,k_0}\), we have

\[ ||s_{k_0+1} \ldots s_{n_0}s_1 \ldots s_{k_0}|| = \tau(s_1 \ldots s_{n_0}) \leq D, \]

so that we have the inclusion of events

\[ \{Y_1 \ldots Y_{n_0} \in B(\alpha n_0, rn_0 + D)\} \supseteq \{(Y_{k_0+1}, \ldots, Y_{n_0}, Y_1, \ldots, Y_{k_0}) \in E_{0,k_0}\}. \quad (3.6) \]
Moreover, using the fact that $Y_i$'s are i.i.d, we also have
\[ \mathbb{P}(Y_{k_0+1}, \ldots, Y_n, Y_1, \ldots, Y_{k_0} \in E_{0,k_0}) = \mathbb{P}(Y_1, \ldots, Y_n \in E_{0,k_0}). \tag{3.7} \]
Combining (3.5), (3.6) and (3.7), and since $D \leq r n_0$, we deduce
\[ \mathbb{P}\left( \frac{1}{n_0} |R_{n_0}| x \in B(\alpha, 2r) \right) \geq e^{-n_0(J_{ls}(\alpha)+2\delta/3)}. \]
Now taking logarithm, dividing by $n_0$, plugging in (4) above, we deduce
\[ I(\alpha) \leq J_{ls}(\alpha) + \delta. \]
Since $\delta > 0$ is arbitrary, we are done with showing $J_{ls}(\alpha) \geq I(\alpha)$.

### 3.4. Dealing with the decay rates at zero.

We recall that up to now, we showed that $I(0) \geq J_{li}(0)$ and for every $\alpha > 0$, $I(\alpha) = J_{li}(\alpha) = J_{ls}(\alpha)$. It only remains to show $J_{ls}(0) \geq I(0)$.

To do this, we now suppose that $J_{ls}(0) < I(0)$ and we shall deduce a contradiction using our large deviation Theorem 2.2 for the displacement functionals and geometric Lemma 3.6.

Since the function $J_{ls}$ is non-negative, the function $I : [0, \infty) \to [0, \infty]$ attains zero at some $\alpha_0$ strictly positive. Since $I$ is lower semi-continuous at $0$ and convex on $[0, \infty)$, it follows that there exists $\epsilon_0 > 0$ and $\delta > 0$ such that for every $\alpha \in [0, 5\epsilon_0]$, we have $I(\alpha) > J_{ls}(0) + 5\delta$.

Up to reducing $\epsilon_0 > 0$, it follows by Theorem 2.2 and the definition of $J_{ls}$ that there exists a sequence $n_\ell$ of positive integers such that
\[ \mathbb{P}(\tau(R_{n_\ell}) \leq n_\ell \epsilon_0) \geq e^{-n_\ell(I(0)-3\delta)} \]
and
\[ \mathbb{P}(|R_{n_\ell}| x \leq 4n_\ell \epsilon_0) \leq e^{-n_\ell(I(0)+\delta)} \tag{3.8} \]
It follows that for every $\ell \geq 1$ large enough, we have
\[ \mathbb{P}(\{\tau(R_{n_\ell}) \leq n_\ell \epsilon_0\} \cap \{|R_{n_\ell}| \geq 4n_\ell \epsilon_0\}) \geq e^{-n_\ell(I(0)-2\delta)}. \tag{3.9} \]
Denote by $E_\ell$ the subset of the Cartesian product $S^{n_\ell}$ given by
\[ \{(s_1, \ldots, s_{n_\ell}) \in S^{n_\ell} | \tau(s_1 \ldots s_{n_\ell}) \leq n_\ell \epsilon_0 \text{ and } |s_1 \ldots s_{n_\ell}| x \geq 4n_\ell \epsilon_0\}, \]
so that by (3.9), for every $\ell \geq 1$ large enough, we have
\[ \mathbb{P}(Y_1, \ldots, Y_{n_\ell}) \in E_\ell) \geq e^{-n_\ell(I(0)-2\delta)}. \tag{3.10} \]
By Lemma 3.6 there exists $N \in \mathbb{N}$ such that for every $\ell$ large enough (so that, we also have $n_\ell \geq N$) and for every $(s_1, \ldots, s_{n_\ell}) \in E_{n_\ell}$, there exists an integer $i \in [1, \ldots, n_\ell]$ such that
\[ |s_{i+1} \ldots s_{n_\ell}s_1 \ldots s_i| x \in [2\epsilon_0 n_\ell, 3\epsilon_0 n_\ell]. \tag{3.11} \]
For every large enough $\ell$ and $(s_1, \ldots, s_{n_\ell}) \in E_{n_\ell}$, fix such a choice of $i$ and denote the corresponding mapping by $p$, i.e. $p(s_1, \ldots, s_{n_\ell}) = i$ (as in 3.3 we can make sure that $p$ is measurable). Furthermore, for every large enough $\ell$ and $i = 1, \ldots, n_\ell$, set
\[ E_{n_\ell,i} = \{(s_1, \ldots, s_{n_\ell}) \in E_{n_\ell} | p(s_1, \ldots, s_{n_\ell}) = i\}, \]
so that $(E_{n_\ell,i})_{i=1,\ldots,n_\ell}$ gives a partition of $E_{n_\ell}$. 
As in (3.3), it follows that for every $\ell$ large enough, there exists an integer $k_\ell \in [1, n_\ell]$ such that
\[
P(\{Y_1, \ldots, Y_{n_\ell} \in E_{n_\ell, k_\ell} \}) \geq \frac{1}{n_\ell} e^{-n_\ell(I(0) - 2\delta)} \geq e^{-n_\ell(I(0) - \delta)}.
\] (3.12)

On the other hand, by (3.11), we have the following inclusion of events
\[
\{Y_{k_\ell + 1} \ldots Y_{n_\ell} Y_1 \ldots Y_{k_\ell} \mid x \in [2n_\ell \epsilon_0, 3n_\ell \epsilon_0] \} \supset \{Y_1, \ldots, Y_{n_\ell} \in E_{n_\ell, k_\ell} \}.
\] (3.13)

Evaluating probabilities on both sides of (3.13), plugging in (3.12) and using the fact that $Y_i$’s are i.i.d, we get
\[
P\left(\frac{1}{n_\ell} |R_{n_\ell}| x \in [2\epsilon_0, 3\epsilon_0] \right) = P\left(\left\{ \frac{1}{n_\ell} |Y_{k_\ell + 1} \ldots Y_{n_\ell} Y_1 \ldots Y_{k_\ell}| \mid x \in [2\epsilon_0, 3\epsilon_0] \right\} \right) \geq e^{-n_\ell(I(0) - \delta)},
\]
contradicting (3.8) and showing that $J_{1a}(0) \geq I(0)$.

As indicated earlier, we have thus shown that $\frac{1}{n} \tau(L_n)$ satisfies a weak LDP with the same rate function $I : [0, \infty) \to [0, \infty]$ given by Theorem 2.2. To see that the full LDP holds under the exponential moment assumption, one just observes that since for every $g \in G$, $\tau(g) \leq |g|_x$, exponential tightness of $\frac{1}{n} L_n$ (Lemma 2.12) implies that of $\frac{1}{n} \tau(L_n)$. This finishes the proof of Theorem 13. \( \square \)

4. Support of the rate function

This section is devoted to the proofs of Theorem C and Propositions L6, L8.

4.1. Support of the rate function.

Proof of Theorem C. 1) Suppose that the interval given by effective support of $I$ has non-empty interior. Taking $\alpha < \beta$ inside the interior of $D_I$, we see that for some $\delta < \frac{1}{2}(\beta - \alpha)$ and for every $t \in \mathbb{N}$ large enough, we have $P(\frac{1}{t} |g|_x \geq \beta - \delta) \geq e^{-t(\beta - \delta)} > 0$ and $P(\frac{1}{t} |g|_x \leq \alpha + \delta) \geq e^{-t(\alpha)} > 0$. So, for every $t$ large enough, there exist elements $g_1(t), g_2(t)$ in $S^t$ satisfying $|g_1(t)|_x \leq t(\alpha + \delta)$ and $|g_2(t)|_x \geq t(\beta - \delta)$. Applying Lemma 4.7 to $g_2(t)$, we obtain an element $g_2'(t) \in S^{k_\ell + t}$ with $\tau(g_2'(t)) \geq t(\beta - \alpha) - L$. Similarly, multiplying $g_1(t)$ with the $k_\ell$-power of a fixed element $h$ in $S$, we obtain $h g_1(t) \in S^{k_\ell + t}$ with $\tau(h g_1(t)) \leq |h g_1(t)|_x \leq t(\alpha + \delta)$. Since $2\delta < \beta - \alpha$ and the difference $|\tau(\cdot) - \ell(\cdot)|$ is uniformly bounded, non-arithmeticity follows provided that $t$ is large enough.

Remark 4.1 (Non-arithmeticity and the equality $\ell_{\text{sub}}(S) = \ell(S)$). The same proof above also shows that for a bounded set $S$, if $\ell_{\text{sub}}(S) \neq \ell(S)$ then $S$ is non-arithmetic. The converse implication being obvious, it follows that $S$ is non-arithmetic if and only if $\ell_{\text{sub}}(S) \neq \ell(S)$.

For the other direction, since $\mu$ is non-elementary and non-arithmetic, it is easy to see that there exists $n_0 \in \mathbb{N}$ such that $S^{n_0}$ contains two hyperbolic elements $h, h'$ with disjoint sets of fixed points on $\partial X$ as well as two elements $g, g'$ with distinct asymptotic translation distances $\ell(g) \neq \ell(g')$. Since the action of $G$ on $\partial X$ as well as the function $\ell$ is continuous (for the latter see [ORT8, Theorem 1.9]), we can choose some disjoint neighborhoods $U_h, U'_h, U_g, U'_g$ such that the corresponding properties (i.e. disjoint sets of fixed points and distinct stable lengths) holds for every pair of elements in these neighborhoods. Let $\nu$ be the probability measure given by the
restriction of \( \mu^{x_0} \) to \( U_h \cup U_{h'} \cup U_g \cup U_{g'} \). It follows that \( \nu \) is a non-elementary probability measure and satisfies \( \nu \leq c \mu^{x_0} \) where \( \frac{1}{n} = \mu^{x_0}(U_h \cup U_{h'} \cup U_g \cup U_{g'}) > 0 \).

On the other hand, it is readily observed (by definitions) that, if we denote by \( I_\mu \) and \( I_\nu \), the rate functions given by Theorem A of the random walks, respectively, driven by \( \mu \) and \( \nu \), we have \( I_\nu(\alpha) \geq n_0 I_\mu(\frac{\alpha}{n_0}) - \log e \) for every \( \alpha \geq 0 \). Now, as \( \nu \) has bounded support, the claim will follow from \ref{lemma_2.4} applied to \( D_{I_\nu} \), since for the support \( S_\nu \) of \( \nu \), we clearly have \( \ell_{\sub}(S_\nu) \leq \ell(S_\nu) \).

2) The fact that \( \overline{D_I} \subseteq [\ell_{\sub}(S), \ell(S)] \) is readily seen from the definitions of \( \ell_{\sub}(S) \) and \( \ell(S) \). So we only need to prove \( D_I \supseteq (\ell_{\sub}(S), \ell(S)) \). There is nothing to prove if \( \ell_{\sub}(S) = \ell(S) \), so suppose \( \ell(S) > \ell_{\sub}(S) \). First, suppose that \( \lambda(\mu) \neq \ell_{\sub}(S) \) and let \( \alpha \in (\ell_{\sub}(S), \ell(S)) \) be such that \( \alpha \leq \lambda(\mu) \). Then, by definition of \( \ell(S) \), there exists \( n_0 \in \mathbb{N} \) and \( g \in S^{x_0} \) such that \( \frac{1}{n_0} |g|_x < \alpha \). Let \( U_g \) be a neighborhood of \( g \) such that \( \mu^{x_0}(U) > 0 \) and \( \frac{1}{n_0} |g'g|_x < \alpha \) for every \( g' \in U \). Using Theorem A we can now write,

\[
I(\alpha) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left( \frac{1}{n} |L_n|_x < \alpha \right) \leq - \liminf_{t \to \infty} \frac{1}{tn_0} \log \mathbb{P}\left( \frac{1}{tn_0} |L_{tn_0}|_x < \alpha \right) \\
\leq - \liminf_{t \to \infty} \frac{1}{tn_0} \log \mathbb{P}(L_{tn_0} \in U^t) \leq - \lim_{t \to \infty} \frac{1}{tn_0} \log \mathbb{P}(L_{tn_0} \in U) \\
= - \frac{1}{n_0} \log \mu^{x_0}(U) < \infty,
\]

where in the first equality we used the convexity of \( I \) together with the fact that \( \alpha \leq \lambda(\mu) \), in the second inequality we used the subadditivity of the displacement functional \( |.|_x \) and in the third inequality we used the fact that the random walk increments are i.i.d.

The proof of the other case is very similar to the proof of Lemma \ref{lemma_2.4} to avoid repetition, we only indicate how to adapt that proof. So, suppose that \( \lambda(\mu) \neq \ell(S) \) and let \( \alpha \in (\ell_{\sub}(S), \ell(S)) \) be such that \( \alpha \geq \lambda(\mu) \). Fix \( \eta > r > 0 \) and \( \varepsilon > 0 \) with \( r + \varepsilon < \eta \) and \( \alpha + 2\eta \leq \ell(S) \). By definition of \( \ell(S) \), for every \( n \in \mathbb{N} \), there exists \( g_n \in S^n \) such that \( \frac{1}{n} |g_n|_x \geq \alpha + 2\eta \). By continuity, for every \( n \in \mathbb{N} \) there exists a neighborhood \( U_n \) of \( g_n \) such that for every \( g'_n \in U_n \), we have \( \frac{1}{n} |g'_n|_x \geq \alpha + \eta \) and \( \mu^n(U_n) > 0 \). Now replace \( \alpha \) in the proof of Lemma \ref{lemma_2.4} by \( \alpha + \eta \), fix \( \delta > 0 \), then choose \( n_0 \in \mathbb{N} \) satisfying (2),(3) and (4) of \ref{lemma_2.3}. Let \( \beta > 0 \) be a constant so that we have \( -\frac{1}{n_0} \log \mathbb{P}(\frac{1}{n_0} |L_n|_x \in B(\alpha + \eta, r)) \leq \beta + \delta/2 \). Note that the choice of \( \beta \) to be finite is guaranteed by the fact that \( \mu^{x_0}(U_n) > 0 \). Now the rest of the proof of Lemma \ref{lemma_2.4} (from \ref{lemma_2.3} onward) goes precisely the same and yield \ref{lemma_2.5}, i.e. for every \( k \geq 1 \),

\[
-I(\alpha) \geq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left( \frac{1}{n} |L_n|_x \geq \alpha \right) \\
\geq \frac{1}{kn_0 + (k-1)kS} \log \mathbb{P}\left( \frac{1}{kn_0 + (k-1)kS} |L_{kn_0 + (k-1)kS}|_x \in B(\alpha, r + \varepsilon) \right) \\
\geq -\beta - \delta.
\]
where we used Theorem A in the first inequality. This shows $I(\alpha) < \infty$ and completes the proof.

3) If $\ell_{\text{sub}}(S) = \ell(S)$, then $\lambda(\mu) = \ell(S)$ and $I(\ell(S)) = 0$ so that by 2), $D_I = \{\ell(S)\}$. Therefore we suppose that $\ell_{\text{sub}}(S) < \ell(S)$ and we only need to show that on $(\ell_{\text{sub}}(S), \ell(S))$, the rate function $I$ is bounded above by $-\min_{g \in S} \log \mu(g) < \infty$; lower semi-continuity of $I$ then entails that $I$ is bounded above by $-\min_{g \in S} \log \mu(g)$ on $[\ell_{\text{sub}}(S), \ell(S)]$ proving the claim. So let $\alpha \in (\ell_{\text{sub}}(S), \ell(S))$. By 2), $\alpha \in D_I$ and by Theorem A for any $\delta > 0$, for every small enough $r > 0$ and large enough $n \in \mathbb{N}$, we have

$$0 < e^{-n(I(\alpha)+\delta)} \leq \mathbb{P}\left(\frac{1}{n}|L_n|_x \in B(\alpha, r)\right) \leq e^{-n(I(\alpha)-\delta)}.$$ (4.1)

It follows that for every such $n \in \mathbb{N}$, there exists $g_n \in \text{supp}(\mu^n)$ with $\frac{1}{n}|g_n|_x \in B(\alpha, r)$. By the i.i.d. property of random walk increments, writing $g_n$ as a product $s_n \ldots s_1$ with $s_i \in \text{supp}(\mu)$, it follows that $\mathbb{P}(L_n = g_n) \geq (\min_{s \in S} \mu(s))^n$. Plugging this in (4.1), since $\delta > 0$ is arbitrary, we deduce that $I(\alpha) \leq -\min_{s \in S} \log \mu(s)$, as required. $\square$

We now construct some examples illustrating in the setting of Theorem C that when the support of the probability measure is not finite, the rate function of LDP can explode on $\ell_{\text{sub}}(S)$ or $\ell(S)$ or both. In fact, by considering the action of $\text{SL}_2(\mathbb{R})$ on the Poincaré disc $\mathbb{D}$ and using the relation $\frac{1}{2} \log ||g|| = |g|_0$ where $g \in G$, $0$ denotes the origin in $\mathbb{D}$ and $||.||$ is the operator norm induced by the Euclidean norm on $\mathbb{R}^2$, [Ser19, Example 5.5] already provides an example of a rate function that explodes on $\ell(S)$. Below, we shall give more examples where rate function explodes on any subset of $[\ell_{\text{sub}}(S), \ell(S)]$.

**Example 4.2 (Rate function that explodes on the boundary).**

Consider $G = \text{SL}_2(\mathbb{R})$ acting isometrically on the Poincaré disc $\mathbb{D}$ endowed with the usual hyperbolic metric $d$. Let $c_a$ and $c_b$ be two geodesics in $\mathbb{D}$ that are of distance $d = 1$ to the origin and denoting their endpoints on $\partial \mathbb{D}$, respectively, by $\{x^+_a, x^-_a\}$ and $\{x^+_b, x^-_b\}$, suppose that these are ordered as $(x^-_b, x^+_b, x^+_a, x^-_a)$. For $k \geq 1$, $t \in \{a, b\}$, $t_k$ be hyperbolic elements of $G$ with translation axis $c_t$, attracting/repelling fixed points $x^+_t$ and $x^-_t$ and translation distance $10 - \frac{1}{k}$ for $t = b$ and $\frac{1}{k}$ for $t = a$. Let $S = \{a_k, b_k | k \geq 1\}$ and let $Y_i$’s be the coordinate functions on $S^\mathbb{N}$. It is easy to see that the semigroup $\Gamma$ generated by $S$ consists of hyperbolic elements whose translation axes is contained in the connected region bounded by $c_a$ and $c_b$. It follows, e.g. by [BS18, Lemma 6.3], that denoting by $D > 0$ twice the distance between $c_a$ and $c_b$, for any $g, h \in \Gamma$, we have

$$\tau(g) + \tau(h) \leq \tau(gh) \leq \tau(g) + \tau(h) + D.$$ (4.2)

Now for $t \in \{a, b\}$, $n \geq k \geq 1$ and $(s_n, \ldots, s_1) \in S^n$, denote by $\ell_{t, k}$ the number of $t_i$’s in $(s_n, \ldots, s_1)$ with $i \geq k$. It is readily seen by (4.2) that we have the following inclusion of events for every $n \geq 1$ and $1 \leq k \leq n$:

$$\left\{\frac{1}{n}\tau(Y_n \ldots Y_1) > 10 - \frac{1}{2k}\right\} \subset \left\{\ell_{b, k} \leq \frac{n}{2}\right\}.$$ (4.3)
and
\[ \left\{ \frac{1}{n} \tau(Y_n \ldots Y_1) < \frac{1}{2k} \right\} \subseteq \left\{ \hat{a}_{n,k} \geq \frac{n}{2} \right\}. \] (4.4)

Notice also that by elementary plane hyperbolic geometry, for every $g \in \Gamma$, we have
\[ 0 \leq |g|_0 - \tau(g) \leq D + 4. \] (4.5)

For any probability measure $\mu$ on $S$, the random variables $\frac{1}{n}|L_n|_o$ satisfies a LDP with some rate function $I$; this follows from Theorem [A] if the support of $\mu$ contains $a_i$'s and $b_i$'s (so that $\mu$ is non-elementary) and from classical theorem of Cramér (which can be readily deduced from Theorem [2.7]) if the support contains only $a_i$'s or $b_i$'s. The inequality (4.5) entails by Theorem 2.7 (or using [DZ 02, Theorem 4.2.13]) that $\frac{1}{n}\tau(L_n)$ satisfies a LDP with rate function $I$ too. Now let $\alpha_k > 0$ be such that $\sum_{k \geq 1} \alpha_k = \frac{1}{2}$ and consider $\mu_1 = \frac{1}{2}\delta_{a_1} + \sum_{k \geq 1} \alpha_k \delta_{b_k}$ supported on $S_1$ and $\mu_2 = \frac{1}{2}\delta_{b_1} + \sum_{k \geq 1} \alpha_k \delta_{a_k}$ supported on $S_2$ and $\mu = \frac{1}{2}(\mu_1 + \mu_2)$ supported on $S$. Denote by $I_1$, $I_2$ and $I$ the rate functions of the LDP of $\frac{1}{n}|L_n|_o$ when $S^n$ is endowed with, respectively, $\mu_1 \otimes \mu_2 \otimes \mu_{N} \otimes \mu_N$. Using Theorem [C] and Stirling's formula, it is not hard to deduce from (4.3) that $\ell(S_1) \notin D_{I_1}$, and from (4.4) that $\ell_{\sub}(S_2) \notin D_{I_2}$, and finally, that we have $D_I = (\ell_{\sub}(S), \ell(S))$. Moreover, using (4.2), one sees that $D_{I_1} = (\ell_{\sub}(S_1), \ell(S_1))$ and $D_{I_2} = (\ell_{\sub}(S_2), \ell(S_2))$.

4.2. Proofs of deterministic consequences.

4.2.1. Hausdorff convergence. The unfolding Lemma [2.5] and Lemmas [3.5] and [3.6] can be used to give a direct proof of Proposition [1.6]. Here, we give a short proof based on our large deviations results.

Proof of Proposition [1.6]. It follows from the definitions that for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$ and $g \in S^n$, $\ell_{\sub}(S) \leq \frac{1}{n}|g|_x \leq \ell(S) + \varepsilon$. This already implies the statement if $\ell_{\sub}(S) = \ell(S)$, so let $\ell_{\sub}(S) < \ell(S)$. Then, for every $\varepsilon \in (0, (\ell(S) - \ell_{\sub}(S))/2)$, the set $S$ contains a finite subset $S'$ such that $\ell(S') > \ell(S) - \varepsilon$ and $\ell_{\sub}(S') \leq \ell(S) - \varepsilon$. This follows from the definitions of $\ell(S)$ and $\ell_{\sub}(S)$. Therefore, in view of the previous paragraph, we can suppose that $S$ is finite. Now endow $S$ with the uniform probability measure and consider the corresponding random walk on $\text{Isom}(X)$. Now given an interval $J$ of non-empty interior in $[\ell_{\sub}(S), \ell(S)]$, by Theorem [A] and 3. of Theorem [C],

\[ \liminf_{n \to \infty} \frac{1}{n} \log P\left( \frac{1}{n}|L_n|_x \in J \right) > \infty, \]

which in particular says that for every $n \in \mathbb{N}$, large enough $J \cap \frac{1}{n}|L_n|_x \neq \emptyset$. Together with the first paragraph above, this shows the Hausdorff convergence of $\frac{1}{n}|S^n|_x$. The convergence of $\frac{1}{n}\tau(S^n)$ is deduced similarly using Theorem [B].

4.2.2. Berger–Wang equality for isometries of hyperbolic spaces.

Proof of Proposition [1.8]. Assume for a contradiction that we have $\ell_{\infty}(S) < \ell(S)$. It follows that there exists $\delta > 0$ such that for every $n \geq 1$, there exists $g_n \in S^n$ with $\frac{1}{n}|g_n|_x \geq \ell_{\infty}(S) + \delta$. By Lemma [2.5] and the fact that $|\tau(.) - \ell(.)|$ is uniformly bounded, we deduce that for every $n$ large enough, there exists $f_n \in \mathcal{F} \subset S^3$ such
that \( \frac{1}{n} \ell(f_n g_n) \geq \ell_\infty(S) + \delta \). Since \( f_n g_n \in S^{k_S + n} \), for every \( n > \frac{2k_S \ell_\infty(S)}{\delta} + k_S \), we get
\[
\frac{\ell(f_n g_n)}{k_S + n} \geq \ell_\infty(S) + \frac{\delta}{2}
\]
yielding a contradiction since \( \ell(g^n) = n \ell(g) \) for every \( n \in \mathbb{N} \) and \( g \in \text{Isom}(X) \). \( \square \)

References

[AMS95] H. Abels, G. A. Margulis, and G. A. Soifer. Semigroups containing proximal linear maps. Israel J. Math., 91(1-3):1–30, 1995.

[Ave76] André Avez. Croissance des groupes de type fini et fonctions harmoniques. In Théorie ergodique (Actes Journées Érgodiques, Rennes, 1973/1974), pages 35–49. Lecture Notes in Math., Vol. 532, 1976.

[BC74] Christian Berg and Jens Peter Reus Christensen. Sur la norme des opérateurs de convolution. Inventiones mathematicae, 23(2):173–178, 1974.

[Ben96] Yves Benoist. Actions propres sur les espaces homogènes réductifs. Ann. of Math. (2), 144(2):315–347, 1996.

[BF18] E. Breuillard and K. Fujiwara. On the joint spectral radius for isometries of non-positively curved spaces and uniform growth. arXiv:1804.00748, 2018.

[BFLM11] Jean Bourgain, Alex Furman, Elon Lindenstrauss, and Shahar Mozes. Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. J. Amer. Math. Soc., 24(1):231–280, 2011.

[Bjö10] M. Björklund. Central limit theorems for Gromov hyperbolic groups. J. Theoret. Probab., 23(3):871–887, 2010.

[BL85] Philippe Bougerol and Jean Lacroix. Products of random matrices with applications to Schrödinger operators, volume 8 of Progress in Probability and Statistics. Birkhäuser Boston, Inc., Boston, MA, 1985.

[BM15] Jairo Bochi and Ian D Morris. Continuity properties of the lower spectral radius. Proceedings of the London Mathematical Society, 110(2):477–509, 2015.

[Bou80] Philippe Bougerol. Comportement asymptotique des puissances de convolution d’une probabilité sur un espace symétrique. In Conference on Random Walks (Kleebach, 1979) (French), volume 74 of Astérisque, pages 29–45, 3. Soc. Math. France, Paris, 1980.

[BQ13] Yves Benoist and Jean-François Quint. Stationary measures and invariant subsets of homogeneous spaces (III). Ann. of Math. (2), 178(3):1017–1059, 2013.

[BQ16a] Yves Benoist and Jean-François Quint. Central limit theorem on hyperbolic groups. Izv. Ross. Akad. Nauk Ser. Mat., 80(1):5–26, 2016.

[BQ16b] Yves Benoist and Jean-François Quint. Random walks on reductive groups, volume 62 of Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Springer, Cham, 2016.

[BS18] E. Breuillard and C. Sert. The joint spectrum. arXiv:1809.02404, 2018.

[Car72] P. Cartier. Fonctions harmoniques sur un arbre. In Symposia Mathematica, Vol. IX (Convegno di Calcolo delle Probabilità, INDAM, Rome, 1971), pages 203–270. 1972.

[CDP90] M. Coornaert, T. Delzant, and A. Papadopoulos. Géométrie et théorie des groupes, volume 1441 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1990. Les groupes hyperboliques de Gromov [Gromov hyperbolic groups]. With an English summary.

[Cor20] E. Corso. Large deviations for random walks on free products of finitely generated groups. arXiv:2004.02291, 2020.

[Cou14] Rémi Coulon. On the geometry of Burnside quotients of torsion free hyperbolic groups. Internat. J. Algebra Comput., 24(3):251–345, 2014.

[Der75] Yves Derriennic. Marche aléatoire sur le groupe libre et frontière de Martin. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 32(4):261–276, 1975.

[DG08] Thomas Delzant and Misha Gromov. Courbure méososcopique et théorie de la toute petite simplification. J. Topol., 1(4):804–836, 2008.
[DM61] E. B. Dynkin and M. B. Maljutov. Random walk on groups with a finite number of generators. *Dokl. Akad. Nauk SSSR*, 137:1042–1045, 1961.

[DS18] Thomas Delzant and Markus Steenbock. Product set growth in groups and hyperbolic geometry. *arXiv preprint arXiv:1804.01867*, 2018.

[DZ02] Amir Dembo and Ofer Zeitouni. Large deviations and applications. In *Handbook of stochastic analysis and applications*, volume 163 of *Statist. Textbooks Monogr.*, pages 361–416. Dekker, New York, 2002.

[Fur63] Harry Furstenberg. A Poisson formula for semi-simple Lie groups. *Ann. of Math. (2)*, 77:335–386, 1963.

[GdlH90] É. Ghys and P. de la Harpe, editors. *Sur les groupes hyperboliques d’aprè s Mikhael Gromov*, volume 83 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1990. Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988.

[Ger71] Peter Gerl. Über die Anzahl der Darstellungen von Worten. *Monatsh. Math.*, 75:205–214, 1971.

[GL13] Sébastien Gouëzel and Steven P. Lalley. Random walks on co-compact Fuchsian groups. *Ann. Sci. Éc. Norm. Supér. (4)*, 46(1):129–173 (2013), 2013.

[Gou14] Sébastien Gouëzel. Local limit theorem for symmetric random walks in Gromov-hyperbolic groups. *J. Amer. Math. Soc.*, 27(3):893–928, 2014.

[Gou17] S. Gouëzel. Analyticity of the entropy and the escape rate of random walks in hyperbolic groups. *Discrete Anal.*, pages Paper No. 7, 37, 2017.

[Gro87] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.* pages 75–263. Springer, New York, 1987.

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

[Ham01] Yuji Hamana. Asymptotics of the moment generating function for the range of random walks. *Journal of Theoretical Probability*, 14(1):189–197, 2001.

[HS17] David Hume and Alessandro Sisto. Groups with no coarse embeddings into hyperbolic groups. *New York J. Math.*, 23:1657–1670, 2017.

[Kes59a] Harry Kesten. Full Banach mean values on countable groups. *Math. Scand.*, 7:146–156, 1959.

[Kes59b] Harry Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.

[Lal93] Steven P. Lalley. Finite range random walk on free groups and homogeneous trees. *Ann. Probab.*, 21(4):2087–2130, 1993.

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

[LP82] Émile Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 258–303. Springer, Berlin-New York, 1982.

[Mah12] Joseph Maher. Exponential decay in the mapping class group. *J. Lond. Math. Soc. (2)*, 86(2):366–386, 2012.

[MS19] Joseph Maher and Alessandro Sisto. Random subgroups of acylindrically hyperbolic groups and hyperbolic embeddings. *Int. Math. Res. Not. IMRN*, (13):3941–3980, 2019.

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

[MT18] Joseph Maher and Giulio Tiozzo. Random walks on weakly hyperbolic groups. *J. Reine Angew. Math.*, 742:187–239, 2018.

[OR18] Eduardo Oregón-Reyes. Properties of sets of isometries of Gromov hyperbolic spaces. *Groups Geom. Dyn.*, 12(3):889–910, 2018.

[Par05] Kalyanapuram R Parthasarathy. *Probability measures on metric spaces*, volume 352. American Mathematical Soc., 2005.

[Pet17] G. Pete. Probability and geometry on groups, lecture notes for a graduate course. 2017.
[Qui02] JF Quint. Cônes limites des sous-groupes discrets des groupes réductifs sur un corps local. *Transformation groups*, 7(3):247–266, 2002.

[Saw78] Stanley Sawyer. Isotropic random walks in a tree. *Z. Wahrsch. Verw. Gebiete*, 42(4):279–292, 1978.

[Ser19] Cagri Sert. Large deviation principle for random matrix products. *Ann. Probab.*, 47(3):1335–1377, 2019.

[Tut68] VN Tutubalin. On random walk in the lobachevsky plane. *Theory of Probability & Its Applications*, 13(3):487–490, 1968.

[Woe00] W. Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000.

**Institut für Mathematik, Universität Zürich, 190, Winterthurerstrasse, 8057 Zürich, Switzerland**

E-mail address: cagri.sert@math.uzh.ch

**Departement Mathematik, ETH Zürich, 101, Rämistrasse, 8092 Zürich, Switzerland**

E-mail address: sisto@math.ethz.ch