LARGE DEVIATIONS FOR RANDOM WALKS ON FREE PRODUCTS OF FINITELY GENERATED GROUPS

EMILIO CORSO

ABSTRACT. We prove existence of the large deviation principle, with a proper convex rate function, for the distribution of the renormalized distance from the origin of a random walk on a free product of finitely generated groups. As a consequence, we derive the same principle for nearest-neighbour random walks on regular trees.

CONTENTS

1. INTRODUCTION AND MAIN RESULT

The study of random walks on algebraic and geometric structures, most notably graphs and groups, has attracted considerable attention over the last four decades. Initiated by Polya's celebrated results on recurrence and transience of symmetric simple random walks on integer lattices ([29]), the subject rose to prominence in the sixties, starting with Kesten's foundational work in the context of groups ([18]). It was later repopularised, mainly owing to pioneering contributions due to Kaimanovich, R. Lyons, Varopoulos, Vershik, to name but a few; several directions of investigation gradually emerged, alongside new connections with various branches of pure and applied mathematics. For further details, we refer the reader to Woess' monograph [36] and the extensive bibliography therein.

In this article, we confine ourselves to the study of random walks on a class of finitely generated groups, and specifically to the investigation of the asymptotic properties of the distribution of the renormalized distance from the origin. Prior to stating our main result, we provide a brief overview of the context within which it can be inscribed.

Let G be a finitely generated group, endowed with the discrete topology, and μ a probability measure on G. The measure μ defines a right random walk $(Y_n)_{n\in\mathbb{N}}$ started at $Y_0 = e$, the identity element of G, given by $Y_n = X_1 \cdots X_n$ for every $n \geq 1$, where the X_n's are independent G-valued random variables identically distributed according to μ (see Section 2 for precise definitions). Select a subset $S \subset G$ generating the group G. It determines a length function ℓ on G, measuring the size of its elements with respect to S; more precisely, for every $g \in G$, $\ell(g)$ is the minimal number of elements from the set $S \cup S^{-1}$ which are needed to obtain g

by multiplying them together. This corresponds to the path distance from the identity on the Cayley graph of G with respect to the generating set S. To simplify the discussion, and in accordance with the cases of utmost interest, we shall always assume that S is finite, though this is not necessary for the validity of Theorem 1.4, which represents the main contribution of the article.

The following well-known result provides an analogue, in a possibly non-commutative setting, of the strong law of large numbers for sums of independent real random variables.

Theorem 1.1. *Assume that μ has finite first moment with respect to the length function ℓ, that is $\int_G \ell(g)\,d\mu(g) < \infty$. Then, there exists a non-negative real number λ such that*

$$\lim_{n \to \infty} \frac{1}{n}\,\ell(Y_n) = \lambda \quad \mathbb{P}\text{-almost surely.}$$

Theorem 1.1 is a consequence of Kingman's subadditive ergodic theorem ([19]); for a proof, we refer to the original article of Guivarc'h [16].

The constant λ appearing in Theorem 1.1 is called the *escape rate* (or *speed*) of the random walk; it clearly depends on μ and on the length function ℓ.

Once almost-sure convergence of the sequence $\left(\frac{1}{n}\ell(Y_n)\right)_{n \geq 1}$ is established, it is natural to enquire about the asymptotic behaviour of the deviations from the mean $\ell(Y_n) - n\lambda$. In this spirit, a central limit theorem was first established in [31] for the case of free groups; a second, more geometric proof of the same result was later provided by Ledrappier in [23]. Subsequently, Bjorklund ([5]) transposed Ledrappier's argument to the setting of Gromov-hyperbolic groups (cf. [15, 13]), proving a central limit theorem for the Green metric on the group G. The rationale behind the introduction of such a metric is of geometric nature: with respect to the Green metric, the horofunction boundary of G is G-equivariantly homeomorphic to the Gromov boundary, a technical assumption which is instrumental in Bjorklund's approach. Thereafter, Benoist and Quint ([4]) extended the result to distance functions defined by word lengths, by adapting the method introduced in [3].

Theorem 1.2 ([4, Thm. 1.1]). *Let G be a Gromov-hyperbolic group, and suppose that μ is a non-elementary and non-arithmetic probability measure on G with finite second moment, that is $\int_G \ell(g)^2\,d\mu(g) < \infty$. Then the sequence of renormalized random variables*

$$\frac{1}{\sqrt{n}}(\ell(Y_n) - n\lambda)\,,\ n \geq 1,$$

converges in distribution to a non-degenerate Gaussian law.

For an explanation of the assumptions on the measure μ appearing in Theorem 1.2, we refer the reader to [4]. It is worth noticing that all earlier works on the central limit theorem in this context rely on the stronger assumption of finiteness of some exponential moment for μ. A recent paper by Mathieu and Sisto ([26]), in which Theorem 1.2 is established for the yet broader class of acylindrically hyperbolic groups, also deserves mention.

In light of Theorem 1.1, it is clear that

$$\mathbb{P}(|\ell(Y_n) - n\lambda| \geq \delta n) \overset{n \to \infty}{\longrightarrow} 0 \text{ for any } \delta > 0. \tag{1.1}$$

We are interested in the decay rate of the probability of such rare events. Properly speaking, we ask whether the sequence of random variables $\left(\frac{1}{n}\ell(Y_n)\right)_{n \geq 1}$ satisfies the large deviation principle (see Section 3); loosely, it amounts to asking if there is a well-defined exponential decay rate for the probability of events of the type appearing in (1.1).

It is natural to expect the large deviation principle to hold for a large class of finitely generated groups, in particular for Gromov-hyperbolic groups; we expand slightly more on possible

extensions of our approach[1] in this direction in Section 6. The applicability of the same strategy to such extensions, as well as to analogous questions in random matrix products, is already mentioned in [34].

Our main result establishes the existence of the large deviation principle, with a proper convex rate function, for the collection of non-trivial free products of finitely generated groups, under a non-degeneracy assumption on the semigroup Γ generated by the support of the driving measure μ. Specifically, we require that Γ is *pattern-avoiding*: there exists a positive integer $D > 0$ such that, for any reduced word $\omega = y_1 \cdots y_D$ of *type size* D in the free product, there is an element $g \in \Gamma \setminus \{e\}$ which neither *starts with* ω nor *ends with* ω^{-1}. For a precise definition, we refer to Section 2.2, while the relevance of this condition to the purposes of the proof is explained in Section 1.1. For the sake of illustration, we hasten to observe that the pattern-avoidance condition is fulfilled, for instance, if Γ intersects two distinct factors of the free product non-trivially (see Example 2.4).

Expanding upon the latter observation, we precede the statement of the main result, Theorem 1.4, with a simpler and more concise version which already singles out a broad class of admissible driving measures.

Proposition 1.3. *Let $r \geq 2$ be an integer, $G_1, \ldots, G_r$ non-trivial finitely generated groups, $G = G_1 * \cdots * G_r$ their free product, S_i a finite generating set of G_i for $i = 1, \ldots, r$, $S = \bigcup_{i=1}^{r} S_i$, ℓ the length function on G determined by S. Let μ be a probability measure on G, and assume its support generates a semigroup Γ with the property that, for any $i \in \{1, \ldots, r\}$, there is an element $g \in \Gamma$ which neither starts nor ends in the factor G_i. If $(Y_n)_{n \geq 0}$ is a right random walk on G with increments distributed according to μ, then the sequence of random variables $\left(\frac{1}{n} \ell(Y_n) \right)_{n \geq 1}$ satisfies the weak large deviation principle with a convex rate function.*

Observe that any semigroup Γ fulfilling the assumptions of Proposition 1.3 avoids patterns of type size $D = 1$ (the converse clearly fails, as shown in Example 2.4). In order to deal with more general pattern-avoiding semigroups, our method compels us to impose an additional constraint on the size of the factors $G_1, \ldots, G_r$.

The complete formulation of our results reads as follows:

Theorem 1.4. *Let $r \geq 2$ be an integer, $G_1, \ldots, G_r$ non-trivial finitely generated groups of subexponential growth, $G = G_1 * \cdots * G_r$ their free product, S_i a finite generating set of G_i for $i = 1, \ldots, r$, $S = \bigcup_{i=1}^{r} S_i$, ℓ the length function on G determined by S. Suppose that μ is a probability measure on G whose support generates a pattern-avoiding semigroup, and let $(Y_n)_{n \geq 0}$ be a right random walk on G with increments distributed according to μ.*

(1) The sequence of random variables $\left(\frac{1}{n} \ell(Y_n) \right)_{n \geq 1}$ satisfies the weak large deviation principle with a convex rate function $I \colon \mathbb{R}_{\geq 0} \to [0, \infty]$.

(2) If μ has a finite exponential moment, then I is a proper function and the sequence $\left(\frac{1}{n} \ell(Y_n) \right)_{n \geq 1}$ satisfies the full large deviation principle with rate function I.

(3) If μ has finite moment-generating function, then I is the Fenchel-Legendre transform of the limiting logarithmic moment generating function of the sequence $\left(\frac{1}{n} \ell(Y_n) \right)_{n \geq 1}$.

A close inspection of the proof of Lemma 4.2 reveals that the whole argument leading to Theorem 1.4 can be readily adapted to establish Proposition 1.3. In particular, the last two assertions of Theorem 1.4 remain equally valid in the setting of Proposition 1.3.

For a precise definition of all the terms involved in the statement of Theorem 1.4, we refer the reader to Sections 2 and 3. Let us just recall here that a probability measure μ on G is said

---

[1]After the first version of this paper appeared, Boulanger, Mathieu, Sert and Sisto [6] proved existence of the large deviation principle for random walks on geodesic hyperbolic spaces, thus encompassing the case of walks on Gromov-hyperbolic groups. The underpinning strategy does not differ substantially from our approach, though it relies on deeper geometric considerations.

to have *a finite exponential moment* if $\int_G \exp\left(\tau\ell(g)\right) d\mu(g) < \infty$ for some $\tau > 0$, and that we say it has *finite moment-generating function* if $\int_G \exp\left(\tau\ell(g)\right) d\mu(g) < \infty$ for every $\tau > 0$.

By taking $G_i = \mathbb{Z}$ for all $i = 1, \ldots, r$, we settle in particular the question of existence of the large deviation principle for random walks on free groups; in turn, this yields the result for nearest-neighbour random walks on locally finite regular trees (a straightfoward adaptation of the proof of Theorem 1.4 allows to deal with regular trees of odd degree as well). For the sake of simplicity, we state the corollary in the case relevant for applications to (possibly lazy[2]) simple random walks on trees.

Corollary 1.5. *Let G be a free group on $r \geq 1$ generators, and let S be a free set of generators. Assume μ is a probability measure on G whose support is contained in $S \cup S^{-1} \cup \{e\}$, and let $(Y_n)_{n\geq 0}$ be a right random walk on G with increments distributed according to μ. The sequence of random variables $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$, where ℓ is the length function on G determined by S, satisfies the large deviation principle with a proper, convex rate function, coinciding with the Fenchel-Legendre transform of the limiting logarithmic moment generating function of $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$.*

Notice that the case $r = 1$ of Corollary 1.5 is not covered in principle by Theorem 1.4; on the other hand, this case is a well-known, elementary instance of Cramer's theorem (cf. [10, Thm. 2.2.3]) on deviations of the empirical mean of independent, identically distributed real random variables. Incidentally, our method would be readily applicable to this case as well, as we point out in section 6, thus yielding an indirect proof of Cramer's theorem for simple random walks on $\mathbb{Z}$ (and $\mathbb{Z}^d$).

Remark 1.6. Several remarks about Theorem 1.4 and Corollary 1.5 are in order.

(1) A version of Grushko's theorem ([24]) asserts that every finitely generated group can be decomposed in an essentially unique way as a free product of finitely many groups, which are not further decomposable as non-trivial free products. Notwithstanding this structural result, the class of examples Theorem 1.4 deals with is restricted, because of the limitations imposed on the generating set S, whose peculiar structure is crucial to our approach (cf. Section 1.1 below).

On the other hand, the pattern-avoiding assumption on the semigroup Γ is by no means necessary for the result to hold; it is only a convenient manner of identifying a large class of examples to which our method applies[3]. Therefore, it stands to reason to expect that a technical refinement of our method would allow to weaken the assumption on the support of the driving measure, and deal with the case in which no conjugate of the semigroup Γ lies in a single factor. In this respect, see the proof of Lemma 4.2. A similar result would yield, notably, that existence of the LDP for the length function is stable under taking free products.

(2) The result in Corollary 1.5 might also be derived, when $2r = p + 1$ for a positive prime p, from the large deviation principle for random walks on linear algebraic groups[4] (see [35, Thm. 3.3]), by choosing an appropriate representation of the free group in the projective special linear group $\mathrm{PSL}_2(\mathbb{Q}_p)$[5]. Our approach is different in that it

---

[2]A G-random walk $(Y_n)_n$ is customarily called *lazy* if $\mu(e) \geq 1/2$; here, for convenience, we employ the terminology in order to refer to the more general case $\mu(e) > 0$.

[3]It becomes clear from the proofs that the very same method takes care, in addition, of some cases such as $\mathrm{supp}\,\mu \subset \{(ab)^n : n \in \mathbb{Z}\}$ in $G = \langle a, b\rangle$ a free group on two generators, in which the semigroup generated by $\mathrm{supp}\,\mu$ is not pattern-avoiding. Ruling out such trivial examples, it doesn't seem unlikely that a failure of the pattern-avoidance condition actually forces a conjugate of Γ to lie in one of the factors.

[4]This has been pointed out to the author by C. Sert.

[5]The rank-one algebraic group $\mathrm{PSL}_2(\mathbb{Q}_p)$ acts by isometries on its Bruhat-Tits tree $\mathcal{T}$, which is regular of degree $p + 1$ (for the construction, we refer to Serre's book [32]). Hyperbolic elements of $\mathrm{PSL}_2(\mathbb{Q}_p)$ act on $\mathcal{T}$ as hyperbolic elements in the geometric sense (cf. [30, Sec. 6]). Choosing a base vertex $o \in \mathcal{T}$, the translation

resorts to the intrinsic geometric properties of the free group, rather than appealing to a representation.

(3) Sharp large deviations estimates for the word-length functional of finite-range random walks on free groups are already present in the work of Lalley[6] ([21, Thm. 7.2]). The techniques adopted there differ significantly from ours, hinging on an extension of the Perron-Frobenius theory of nonnegative matrices to certain inhomogeneous matrix products; they yield finer information on the rate function, notably strict convexity, but require the assumption of *aperiodicity* of the random walk (cf. [21]), which our method does not necessitate.

Remark 1.7. Our hypothesis on the support of μ is unrelated to the choice of the generating set S. This makes Theorem 1.4 applicable, for instance, to the following circumstance, in which the driving measure has *apriori* no connection with the generating set. Let G be a finitely generated group, $H < G$ a finite-index subgroup (hence H is finitely generated by Schreier's subgroup lemma), $S \subset H$ a finite generating set of H, $T \subset G$ a set of representatives of right cosets of H in G, $\tilde{S} = \{st : s \in S, t \in T\}$ the corresponding finite generating set of G. Suppose that $\tilde{\mu}$ is a probability measure on G whose support is contained $\tilde{S} \cup \tilde{S}^{-1} \cup \{e\}$, thus giving rise to a nearest-neighbour random walk $(Y_n)_{n \in \mathbb{N}}$ on the Cayley graph $\mathrm{Cay}(G, \tilde{S})$ of G with respect to $\tilde{S}$. Let $\tau_1 < \tau_2 < \cdots \tau_n < \cdots$ be the strictly increasing sequence of stopping times defined by the successive instants in which the random walk visits H; they are all finite $\mathbb{P}$-almost surely, since H has finite index in G. By an iterative application of the strong Markov property ([20, Chap. 17]) to the process $(Y_n)_{n \in \mathbb{N}}$, it follows that H-valued process $(Y_{\tau_n})_{n \in \mathbb{N}}$ (where we agree that $Y_{\tau_0} = e$) is a right random walk on H driven by a measure μ having finite moment-generating function with respect to the word length determined by S; if H is a non-trivial free product of finitely generated groups, all conclusions of Theorem 1.4 hold.

An example of interest is the arithmetic group $\mathrm{SL}_2(\mathbb{Z})$, which contains a multitude of finite-index free subgroups (cf. [17, Chap. II]).

1.1. **Outline of the strategy.** To illustrate the overarching strategy of our proof of Theorem 1.4, it is informative to recall the indirect approach to the proof of Cramer's theorem for i.i.d. real random variables, put forward by Lanford ([22]). If $(X_n)_{n \geq 1}$ is a sequence of i.i.d. $\mathbb{R}$-valued random variables and $S_n = \sum_{i=1}^{n} X_i$ denotes the sequence of partial sums, then, for every $x \in \mathbb{R}$ and $\varepsilon > 0$, the limit $\lim_n \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}S_n \in (x - \varepsilon, x + \varepsilon)\right)$ exists in $[-\infty, 0]$ by super-multiplicativity of the sequence $\mathbb{P}\left(\frac{1}{n}S_n \in (x - \varepsilon, x + \varepsilon)\right)$, which in turn is given by additivity of the the process $(S_n)_{n \geq 1}$. The weak LDP now follows from a standard result in large deviations' theory (see Proposition 3.4). Similarly, the weak LDP holds for any additive functional[7] ℓ' of a random walk $(Y_n)_{n \in \mathbb{N}}$ on a group G.

The major obstacle, when attempting to transport this argument to our context, lies in the defect of additivity of length functions on discrete groups; subadditivity only ensures super-multiplicativity of the sequence $\mathbb{P}(\frac{1}{n}\ell(Y_n) \in I)$ for intervals of the form $I = (-\infty, x)$. Still, if the random walk can be restricted to subsets in which the length function is almost additive (cf. Lemma 4.2 and the terminology introduced thereunder) without sizeable loss in the exponential decay rate of the corresponding probabilities, then Lanford's approach carries over almost unaffectedly. Specifically, the structure of the generating set S, obtained by concatenating generating sets of the various factors, enables us to quantify neatly the lack

---

distance from o corresponds, up to a multiplicative factor, to the operator norm on $\mathrm{PSL}_2(\mathbb{Q}_p)$ derived from a choice of a K-invariant ultrametric norm on the local field $\mathbb{Q}_p$, where $K < \mathrm{PSL}_2(\mathbb{Q}_p)$ is the compact stabilizer of o. Selecting hyperbolic elements which generate a Zariski-dense free subgroup of $\mathrm{PSL}_2(\mathbb{Q}_p)$ amounts to defining an isometric embedding of the corresponding free group in $\mathrm{PSL}_2(\mathbb{Q}_p)$.

[6]We thank S. Müller for drawing our attention to this reference.

[7]This means that $\ell'(Y_{n+m}) = \ell'(Y_n) + \ell'(Y_m)$ for every $n, m \geq 0$.

of additivity in terms of the reduced-word expansion of the elements involved; the pattern-avoiding assumption on the semigroup Γ can then be leveraged to confine the attention to subsets on which the length function is *weakly additive*, and which are attained by the random walk with sufficiently high probability on an exponential scale. This is detailed in Lemma 4.2. Once a uniform lower bound for the loss of additivity is achieved, it is possible to deduce that, if $\gamma = \limsup_n \frac{1}{n} \log \mathbb{P}\big(\frac{1}{n}\ell(Y_n) \in (x-a, x+a)\big)$ for given $x, a \in \mathbb{R}_{>0}$, then the bound $\mathbb{P}\big(\frac{1}{n_k}\ell(Y_{n_k}) \in (x-a, x+a)\big) \geq e^{n_k(\gamma - \eta)}$ (η being an arbitrarily small parameter) holds for a non-lacunary sequence of integers $(n_k)_k$. The arithmetic nature of such a sequence permits to deduce a lower bound $\liminf_n \frac{1}{n} \log \mathbb{P}\big(\frac{1}{n}\ell(Y_n) \in (x-b, x+b)\big) \geq \gamma - \eta$, at the minor cost of choosing b strictly larger than a; this is the purpose of Lemma 4.1, which in a sense plays the role of Fekete's lemma in Lanford's original argument. The actual proof of the weak LDP is then articulated in Proposition 4.3. Convexity of the rate function follows from entirely analogous arguments, as the proof of Proposition 5.1 highlights. Finally, the remaining properties of the rate function mentioned in the statement of Theorem 1.4 are inferred from well-known foundational results in the theory of large deviations (cf. Proposition 3.3, Theorem 3.5 and Sections 5.1, 5.2).

As a concluding comment, let us point out that the strategy outlined here parallels arguments employed in [35] to deal with large deviations of the Cartan projection of random matrix products; in this context, a weak form of additivity for the Cartan projection is satisfied on (r, ε)-Schottky semigroups, as shown by Benoist ([2]). The restriction of the random walk to such semigroups is then made possible by a result of Abels-Margulis-Soifer ([1]), establishing the ubiquity of (r, ε)-proximal elements in Zariski-dense semigroups.

1.2. **Outline of the article.** We begin with some preliminaries on random walks on finitely generated groups in Section 2, which mainly serve the purpose of fixing notation and elucidating the nature of the pattern-avoiding assumption we impose on the semigroup Γ. In Section 3 we recall some standard terminology from the theory of large deviations, together with a few general facts which are employed in the proof of Theorem 1.4. Sections 4 and 5 are devoted to the proof our main result 1.4; specifically, in Section 4 we establish existence of the large deviation principle, while in Section 5 we prove convexity of the rate function, which, together with properness, allows us to identify it as the convex conjugate of a logarithmic moment generating function. Finally, in Section 6 we assemble ideas on possible generalizations of Theorem 1.4, list some open questions and formulate related conjectures.

2. Random walks on groups

2.1. **Word length and metric on a finitely generated group.** Convenient sources for the material presented hereunder are [17, 25, 36].

Let G be a finitely generated group with identity element e, $S \subset G$ a finite generating set. Let $S^{-1} = \{s^{-1} : s \in S\}$ denote the set of inverses of the elements in S, so that

$$G = \{s_1 \cdots s_n : n \geq 1, s_i \in S \cup S^{-1} \text{ for all } 1 \leq i \leq n\}.$$

We define the *word length* ℓ detemined by the generating set S as the function $\ell \colon G \to \mathbb{N}$ given by

$$\ell(g) = \inf\{n \in \mathbb{N} : \text{ there exist } s_1, \ldots, s_n \in S \cup S^{-1} \text{ such that } g = s_1 \cdots s_n\}$$

for every $g \in G$, with the understanding that $\ell(e) = 0$. Then ℓ is a length function, meaning that it satisfies the following properties:

- $\ell(g) \geq 0$ for all $g \in G$ and $\ell(g) = 0$ if and only if $g = e$;
- $\ell(g^{-1}) = \ell(g)$ for all $g \in G$;
- $\ell(g_1 g_2) \leq \ell(g_1) + \ell(g_2)$ for all $g_1, g_2 \in G$.

The word length ℓ determines a distance function d on G, called the *word metric* associated to the generating set S, defined by $d(g_1, g_2) = \ell(g_1^{-1} g_2)$ for all $g_1, g_2 \in G$. The word metric d is invariant for the action of G on itself by left translation, namely $d(gg_1, gg_2) = d(g_1, g_2)$ for all $g, g_1, g_2 \in G$.

We denote by $\mathrm{Cay}(G, S) = (V, E)$ the Cayley graph of G with respect to S; we recall that this is the simple, undirected graph whose vertex set V is the group G, where two vertices $g_1, g_2 \in V$ are connected by an edge $e = \{g_1, g_2\} \in E$ if and only if $d(g_1, g_2) = 1$. In other words, there is an edge connecting g_1 to g_2 if and only if there is $s \in S \cup S^{-1} \setminus \{e\}$ such that $g_2 = g_1 s$. The graph $\mathrm{Cay}(G, S)$ is connected, transitive and locally finite of degree $|S \cup S^{-1} \setminus \{e\}|$. The word metric d on G corresponds, via this identification, to the path distance on the vertex set V (cf. [25, Chap. 3]).

Let $B^G(T) = \{g \in G : \ell(g) \leq T\}$ be the closed d-ball of radius T centered at the identity, for any $T \in \mathbb{R}_{\geq 0}$. As the sequence $\left(|B^G(n)|\right)_{n \geq 1}$ is submultiplicative, the limit $\gamma_S = \lim_n |B^G(n)|^{1/n}$ exists; we say that G has *subexponential growth* if $\gamma_S = 1$, a property which is actually independent of the generating set S. Recall that a broad class of finitely generated groups with subexponential (in fact, polynomial) growth consists of nilpotent groups ([37]).

If $G = G_1 * \cdots * G_r$ is the free product (cf. [7]) of finitely generated groups $G_1, \ldots, G_r$, we shall always restrict our considerations to the following kind of generating sets (and corresponding word lengths): we fix generating sets $S_i \subset G_i$ for each factor G_i of the free product, and take the union $S = \bigcup_{i=1}^{r} S$ as generating set for G.

2.2. Free products and pattern-avoiding subsets.

Let $r \geq 2$ be an integer, $G_1, \ldots, G_r$ non-trivial finitely generated groups, and let $G = G_1 * \cdots * G_r$ be the free product of the G_i's. We shall identify each G_i, $1 \leq i \leq r$, with its isomorphic copy embedded in G.

Lemma 2.1 ([17, Chap. II, Prop. 1]). *For any non-trivial element $g \in G$, there exist uniquely determined non-trivial elements $x_1 \in G_{i_1}, \ldots, x_m \in G_{i_m}$, with $i_j \neq i_{j+1}$ for all $1 \leq j \leq m - 1$, such that $g = x_1 x_2 \cdots x_m$.*

Any product $x_1 \cdots x_m$ as in Lemma 2.1 is referred to as a *reduced word* of *type size* m in the free product; correspondingly, we shall also say that $g = x_1 \cdots x_m$ is an element of type size m. For any $i \in \{1, \ldots, m\}$, we call the element x_i the i-th *letter* of the reduced word $x_1 \cdots x_m$.

Remark 2.2. Suppose that we fix a generating set $S_i \subset G_i$ for each factor of the free product, and let ℓ_i denote the associated word length on G_i. Then, if ℓ is the word length determined by the generating set $S = \bigcup_{i=1}^{r} S_i \subset G$ and if $g, x_1, \ldots, x_m$ are as in Lemma 2.1, it holds $\ell(g) = \ell(x_1) + \cdots + \ell(x_m)$. Observe in particular that, while the *word length* of an element $g \in G$ depends on the choice of the generating sets for the factors, the *type size* of g does not.

Let $\omega = y_1 \cdots y_d$ be a reduced word of type size d, $g \in G$ an element of type size at least 2, with reduced-word decomposition $g = x_1 \cdots x_m$. We shall say that g

- *starts with ω* if $x_1 \cdots x_{\inf\{d, \lfloor m/2 \rfloor\}} = y_1 \cdots y_{\inf\{d, \lfloor m/2 \rfloor\}}$, and
- *ends with ω* if $x_{m - \inf\{d, \lfloor m/2 \rfloor\} + 1} \cdots x_m = y_1 \cdots y_{\inf\{d, \lfloor m/2 \rfloor\}}$,

where $\lfloor a \rfloor$ indicates the integer part of a real number a. Notice that the definition is independent of any choice of generating sets for the factors $G_1, \ldots, G_r$ of the free product.

Example 2.3. If $G = \langle a, b \rangle$ is a free group on two generators a and b, then the element $abab$ starts with ab and ends with ab, while the element $abab^{-1}a^{-1}$ starts with ab and ends with $b^{-1}a^{-1}$. Also, according to our definition, the latter element starts with any word $ab\omega'$ obtained by juxtaposing a reduced word ω' to ab in such a way that $ab\omega'$ is again a reduced word.

A subset $\mathcal{T} \subset G$ is called *pattern-avoiding* if there exists a positive integer $D > 0$ such that, for any reduced word $\omega = y_1 \cdots y_D$ of type size D in the free product, there exists $g \in \mathcal{T}$ such that g does not start with ω and does not end with $\omega^{-1} = y_D^{-1} \cdots y_1^{-1}$ (in particular, g has type size at least 2). In case we need to keep track of the integer D, we shall say that $\mathcal{T}$ *avoids patterns of type size D*. The examples presented below clarify the notion.

Example 2.4.   (1) Let $G = \langle a, b, c \rangle$ be a free group on three generators a, b and c. The sets

$$\mathcal{T}_1 = \{ab, bc\}, \ \mathcal{T}_2 = \{acb, a^3bca^{-2}\}, \ \mathcal{T}_3 = \{aba^{-1}, bab^{-1}\}$$

are pattern-avoiding, while the set

$$\mathcal{T}_4 = \{ab, ac^2, ca^{-1}\}$$

is not pattern-avoiding, as all its elements either start with a or end with a^{-1}.
   (2) If $\mathcal{S} \subset gG_ig^{-1}$ for some $i \in \{1, \ldots, r\}$ and some $g = x_1 \cdots x_m \in G$, then the semigroup Γ generated by $\mathcal{S}$ is not pattern-avoiding: all its elements start with $x_1 \cdots x_m$ and end with $(x_1 \cdots x_m)^{-1}$.
   (3) Suppose that there are indices $i \neq j \in \{1, \ldots, r\}$ such that $\mathcal{S} \cap (G_i \setminus \{e\}) \neq \emptyset$ and $\mathcal{S} \cap (G_j \setminus \{e\}) \neq \emptyset$. Then the semigroup Λ generated by $\mathcal{S}$ is pattern-avoiding: if $x \in \mathcal{S} \cap (G_i \setminus \{e\})$ and $y \in \mathcal{S} \cap (G_j \setminus \{e\})$, then $\{xy, yx\}$ is pattern-avoiding and contained in Λ.
   (4) The semigroup generated by $\{aba, a^2ba^2\}$ in $G = \langle a, b \rangle$ avoids patterns of type size 1, but does not satisfy the hypotheses of Proposition 1.3: its elements start and end in the factor $\langle a \rangle$.

Obviously, if $\mathcal{T}' \subset \mathcal{T} \subset G$ and $\mathcal{T}'$ is pattern-avoiding, then so is $\mathcal{T}$. Conversely, the following elementary observation is essential for our line of reasoning in Section 4: if $\mathcal{T}$ is pattern-avoiding, then there exists a finite subset $\mathcal{T}' \subset \mathcal{T}$ which is also pattern-avoiding[8].

2.3. **Random walks on finitely generated groups.** Let μ be a probability measure on the group G; equivalently, μ is a function defined on G taking non-negative real values and satisfying $\sum_{g \in G} \mu(g) = 1$. Then μ defines a right random walk on G as follows: let $(X_n)_{n \geq 1}$ be a sequence of independent, identically distributed G-valued random variables with common law μ. Implicitly, we consider them to be defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which will be fixed hereinafter. We define a G-valued stochastic process $(Y_n)_{n \in \mathbb{N}}$ by setting $Y_0 = e$, $Y_n = X_1 \cdots X_n$ for every integer $n \geq 1$. The process $(Y_n)_{n \in \mathbb{N}}$ is called a *right random walk* on G, issued from the origin e with increments distributed according to μ. Equivalently, one may defined the process $(Y_n)_{n \in \mathbb{N}}$ as a Markov chain on G issued from e with transition matrix $Q = (q(x, y))_{x, y \in G}$ given by $q(x, y) = \mu(x^{-1}y)$ for all $x, y \in G$ (cf. [36, Sec. 1.1]).

Let $\operatorname{supp} \mu = \{g \in G : \mu(g) > 0\}$ be the support of the measure μ. If $\operatorname{supp} \mu \subset S \cup S^{-1}$, then the process $(Y_n)_{n \in \mathbb{N}}$ can also be interpreted as a nearest-neighbour random walk on the Cayley graph $\operatorname{Cay}(G, S)$, where the walker in position x moves to xs with probability $\mu(s)$, for all $s \in S \cup S^{-1}, x \in G$. Notice that we are not excluding the case $\mu(e) > 0$, so that the walker may have positive probability of remaining where it is.

---

[8]A simple enumeration of all possibilities shows that $\mathcal{T}'$ can be chosen with cardinality at most 3.

Let $\mathbb{E}[X]$ denote the expectation of a random variable $X \colon \Omega \to \mathbb{R}$ with respect to the probability measure $\mathbb{P}$. If μ has finite first moment, the sequence of renormalized averaged lengths

$$\frac{1}{n}\mathbb{E}[\ell(Y_n)], \ n \geq 1,$$

is a subadditive real sequence, and as such converges to a limit $\lambda \in \mathbb{R}_{\geq 0}$, called the escape rate or speed of the random walk $(Y_n)_{n \in \mathbb{N}}$. As mentioned in the introduction (Theorem 1.1), $\mathbb{P}$-almost every trajectory $(y_n)_{n \geq 0} \in G^{\mathbb{N}}$ of the random walk actually satisfies $\frac{1}{n}\ell(y_n) \overset{n \to \infty}{\longrightarrow} \lambda$.

Remark 2.5. (1) We could equally well consider random walks issued at any initial vertex $g_0 \in G$, by defining $Y_0' = g_0$, $Y_n' = g_0 X_1 \cdots X_n$ for any $n \geq 1$. It is then natural to consider the renormalized distance $\frac{1}{n}d(g_0, Y_n')$ which, by invariance of d under left translations, equals precisely $\frac{1}{n}d(e, X_1 \cdots X_n) = \frac{1}{n}\ell(Y_n)$. Hence, for the purpose of our considerations, there is no loss of generality in assuming that the random walk starts at the origin.

(2) Similarly, restricting to *right* random walks does not result in any loss of generality; if $Y_n' = X_n \cdots X_1$, $n \geq 1$, is a left random walk issued from the origin with driving measure μ, then $(Y_n^{-1})_{n \in \mathbb{N}}$ is a right random walk with driving measure $\iota_* \mu$, given by $\iota_* \mu(g) = \mu(g^{-1})$ for every $g \in G$, and $\ell(Y_n^{-1}) = \ell(Y_n)$ for every $n \in \mathbb{N}$.

3. LARGE DEVIATION PRINCIPLE

In this section, we briefly review some of the terminology that is usually employed in the theory of large deviations. For a comprehensive introduction to the subject, the reader is referred to [10].

Throughout this section, X denotes a Hausdorff regular topological space, endowed with the Borel σ-algebra $\mathcal{B}$. Let $(\mu_n)_{n \geq 1}$ be a sequence of Borel probability measures on X, $I \colon X \to [0, \infty]$ a lower semicontinuous function. The *effective domain* of I is the set $D_I = \{x \in X : I(x) < \infty\}$.

Definition 3.1. We say that the sequence $(\mu_n)_{n \geq 1}$ satisfies the large deviation principle (or, in abridged form, LDP) with rate function I if, for any Borel measurable set $\Lambda \subset X$,

$$-\inf_{x \in \Lambda^\circ} I(x) \leq \liminf_{n \to \infty} \frac{1}{n}\log \mu_n(\Lambda) \leq \limsup_{n \to \infty} \frac{1}{n}\log \mu_n(\Lambda) \leq -\inf_{x \in \overline{\Lambda}} I(x) \ ,$$

where Λ° and $\overline{\Lambda}$ denote the interior and the closure of Λ, respectively.

We observe in passing that, for a given sequence $(\mu_n)_{n \geq 1}$, there is at most one lower semicontinuous function I for which the LDP can hold ([10, Lem. 4.1.4]).

In Definition 3.1, it is obviously equivalent to require that

$$\liminf_{n \to \infty} \frac{1}{n}\log \mu_n(V) \geq -\inf_{x \in V} I(x) \text{ for every open set } V \subset X \tag{3.1}$$

and

$$\limsup_{n \to \infty} \frac{1}{n}\log \mu_n(F) \leq -\inf_{x \in F} I(x) \text{ for every closed set } F \subset X. \tag{3.2}$$

If the lower bound (3.1) holds for any open set $V \subset X$, while the upper bound (3.2) holds just for all compact sets $K \subset X$, then we say that the sequence $(\mu_n)_{n \geq 1}$ satisfies the weak large deviation principle (weak LDP) with rate function I.

If $(Z_n)_{n \geq 1}$ is a sequence of X-valued random variables, and μ_n denotes the law of Z_n for every $n \geq 1$, we shall say that $(Z_n)_{n \geq 1}$ satisfies the (weak) LDP if the sequence $(\mu_n)_{n \geq 1}$ satisfies the (weak) LDP.

Under certain conditions, we may retrieve the full LDP from the existence of the weak LDP. The most common of these conditions involves the notion of exponential tightness.

Definition 3.2. We say that a sequence $(\mu_n)_{n\geq 1}$ of Borel probability measures on X is exponentially tight if, for every $\alpha \in \mathbb{R}_{\geq 0}$, there exists a compact set $K \subset X$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(X \setminus K) < -\alpha .$$

In other words, the mass is concentrated on compact sets, on an exponential scale.

It is intuitively clear that exponential tightness enables to pass from a weak form of the LDP to a strong form, something which we clarify in the following proposition (cf. [10, Lem. 1.2.18]).

Proposition 3.3. *Let $(\mu_n)_{n\geq 1}$ be an exponentially tight sequence of Borel probability measures on X. Assume that $(\mu_n)_{n\geq 1}$ satisfies the weak LDP with rate function I. Then:*

(1) $(\mu_n)_{n\geq 1}$ satisfies the LDP with rate function I;

(2) I is a proper function.

The following statement establishes a criterion to determine whether the weak LDP holds, without knowing the rate function in advance. It will be the key tool to prove existence of the weak LDP in our context.

Proposition 3.4 ([10, Thm. 4.1.11]). *Let $(\mu_n)_{n\geq 1}$ be a sequence of Borel probability measures on X. Define the function $I \colon X \to [0, \infty]$ by*

$$I(x) = \sup_{x \in V \text{ open}} -\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(V) \text{ for all } x \in X. \tag{3.3}$$

Then I is lower semicontinuous. Moreover, if

$$I(x) = \sup_{x \in V \text{ open}} -\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(V) \text{ for all } x \in X, \tag{3.4}$$

then the sequence $(\mu_n)_{n\geq 1}$ satisfies the weak LDP with rate function I.

Let us observe that, both in (3.3) and in (3.4), we may clearly replace the whole collection of open sets containing the point $x \in X$ by any fundamental system of open neighborhoods of x.

Assume now that X is a locally convex, Hausdorff topological vector space over $\mathbb{R}$, and let X^* denote its topological dual. In case the sequence $(\mu_n)_{n\geq 1}$ satisfies the LDP on X with a proper, convex rate function I, it is possible to give an alternative expression for the rate function itself, provided that a certain logarithmic moment generating function exists. More precisely, define the *logarithmic moment generating function* of the measure μ_n, for each integer $n \geq 1$, as the function $\Lambda_n \colon X^* \to (-\infty, \infty]$ given by

$$\Lambda_n(\varphi) = \log \int_X e^{\langle \varphi, x \rangle} d\mu_n(x) \quad \text{for all } \varphi \in X^*,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard dual pairing between X^* and X. The *limiting logarithmic moment generating function* of the sequence $(\mu_n)_{n\geq 1}$ is then defined as

$$\Lambda(\varphi) = \limsup_{n\to\infty} \frac{1}{n} \Lambda_n(n\varphi) \in (-\infty, \infty] \quad \text{for all } \varphi \in X^*.$$

Given a function $f \colon X \to (-\infty, \infty]$, not identically infinite, we define its *Fenchel-Legendre transform* $f^* \colon X^* \to (-\infty, \infty]$ as

$$f^*(\varphi) = \sup_{x \in X} \{\langle \varphi, x \rangle - f(x)\} \quad \text{for all } \varphi \in X^*.$$

If $g \colon X^* \to (-\infty, \infty]$ is a function defined on the dual space, we shall view its Fenchel-Legendre transform g^* as a function defined just on X, rather than on the entire bidual X^{**}.

A remarkable consequence of Varadhan's integral lemma ([10, Thm. 4.3.1]), in conjunction with Fenchel-Moreau's duality theorem ([8, Thm. 1.11]), is the following characterization of the rate function (cf. [10, Thm. 4.5.10]).

Theorem 3.5. *Let $(\mu_n)_{n\geq 1}$ be a sequence of Borel probability measures on a locally convex, Hausdorff topological vector space X. Assume the following:*

(1) the limiting logarithmic moment generating function $\Lambda\colon X^* \to (-\infty, \infty]$ of the sequence $(\mu_n)_{n\geq 1}$ is finite for every $\varphi \in X^*$;

(2) the sequence $(\mu_n)_{n\geq 1}$ satisfies the LDP with a proper, convex rate function I.

Then the rate function I is the Fenchel-Legendre transform of Λ, namely

$$I(x) = \sup_{\varphi \in X^*} \{\langle \varphi, x \rangle - \Lambda(\varphi)\} \text{ for every } x \in X.$$

Theorem 3.5 reveals the importance of knowing *a priori* the existence of the LDP with a proper, convex rate function.

4. EXISTENCE OF LDP

We now set out to prove our main Theorem 1.4. Specifically, the objective of the present section is twofold: in Proposition 4.3, we address existence of the weak LDP, with a certain rate function, under the pattern-avoiding assumption for the semigroup generated by the support of the driving measure, while in Proposition 4.4 the result is upgraded to the full LDP, under the additional requirement of finiteness of some exponential moment. Convexity of the rate function, and the ensuing identification of it as a Fenchel-Legendre transform, are dealt with in Section 5.

For a start, we briefly recall the setup. Let $G_1, \ldots, G_r$ be a finite collection of non-trivial finitely generated groups of subexponential growth, $G = G_1 * \cdots * G_r$ their free product. For any $i \in \{1, \ldots, r\}$, $S_i \subset G_i$ is a finite set of generators of G_i, so that $S = \bigcup_{i=1}^r S_i$ is a finite generating set for G, with associated word length $\ell\colon G \to \mathbb{N}$. Let μ be a probability measure on G, $(Y_n)_{n\geq 0}$ a right random walk on G issued from the identity with steps distributed according to μ. For every integer $n \geq 1$, let μ_n be the law of the random variable $\frac{1}{n}\ell(Y_n)$.

Henceforth, we shall denote by $B(y, \varepsilon)$ the open interval $(y - \varepsilon, y + \varepsilon) \subset \mathbb{R}$, for any $y \in \mathbb{R}$ and any $\varepsilon > 0$. Furthermore, for any positive integer k, we let

$$kB(y, \varepsilon) = \{kz : z \in B(y, \varepsilon)\}.$$

We precede the statement of Proposition 4.3 by two technical lemmas which, taken together, essentially allow to reduce the problem of establishing LDP in this context to a setup akin to the standard case of i.i.d. real random variables, in which (almost-)additivity of the process can be put to good use.

The first of the two lemmas allows to deduce a lower bound for the asymptotic exponential decay rate of the probabilities $\mu_n(B(x, b))$ from a uniform lower bound on a non-lacunary sequence of times.

Lemma 4.1. *Suppose that there exist $a > 0, \gamma \in \mathbb{R}$, a strictly increasing sequence $(n_k)_{k\geq 1}$ of positive integers with $\lim_{k\to\infty} n_{k+1}/n_k = 1$, such that*

$$\mu_{n_k}(B(x, a)) \geq e^{n_k \gamma} \text{ for all } k \geq 1. \tag{4.1}$$

Then, for all $b > a$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B(x, b)) \geq \gamma.$$

Proof. Choose a finite set $\mathcal{F} \subset G$ such that $\sum_{g\in\mathcal{F}} \mu(g) > 1/2$. For any $k \geq 1$, set

$$M_k = \sup\{\ell(x_1 \cdots x_{n_{k+1}-n_k}) : x_i \in \mathcal{F} \cup \{e\} \text{ for all } 1 \leq i \leq n_{k+1} - n_k\},$$

and notice that the upper bound $M_k \leq (n_{k+1} - n_k)M_1$ holds by subadditivity of ℓ.

Now let $N \geq n_1$ be arbitrary; there exists a unique $k = k(N) \geq 1$ such that $n_k \leq N < n_{k+1}$. As $b - a > 0$, the assumption $n_{k+1}/n_k \to 1$ implies that there exists $k_0 \in \mathbb{N}$ such that

$$\{\ell(Y_{n_k}) \in n_k B(x,a)\} \cap \{X_{n_k+1} \in \mathcal{F}, \dots, X_N \in \mathcal{F}\} \subset \{\ell(Y_N) \in NB(x,b)\} \text{ for all } k \geq k_0;$$

this follows from the double inequality $|\ell(g) - \ell(h)| \leq \ell(gh) \leq \ell(g) + \ell(h)$, holding for every $g, h \in G$. Now, if $k \geq k_0$ and $N \in \{n_k, \dots, n_{k+1} - 1\}$, we may estimate

$$\mu_N(B(x,b)) = \mathbb{P}(\ell(Y_N) \in nB(x,b)) \geq \mathbb{P}(\ell(Y_{n_k}) \in n_k B(x,a), X_{n_k+1} \in \mathcal{F}, \dots, X_N \in \mathcal{F})$$
$$= \mu_{n_k}(B(x,a)) \mu(\mathcal{F})^{N - n_k} \geq e^{n_k \gamma} 2^{-(n_{k+1} - n_k)},$$

the last two inequalities being given, respectively, by independence and stationarity of the process $(X_n)_{n \geq 1}$, and by the assumption of the lemma. Taking the logarithm and dividing by N, we obtain

$$\frac{1}{N} \log \mu_N(B(x,b)) \geq \frac{n_k}{N} \gamma - \frac{n_{k+1} - n_k}{N} \log 2.$$

Taking the inferior limit as $N \to \infty$ on both sides, and observing that the assumption on $(n_k)_k$ implies $\lim_{N \to \infty} n_{k(N)}/N = 1$, we achieve the proof. □

The next lemma expresses the possibility of restricting the random walk to subsets on which the length function ℓ is almost additive, without losing consistently on the exponential decay rate of the probabilities involved.

For every $T \in \mathbb{R}_{\geq 0}$, set $\theta_T = \sup\{|B^{G_i}(T)| : i = 1, \dots, r\}$.

Lemma 4.2. *Let ν be a probability measure on G, $\mathcal{T} \subset G$ a finite subset avoiding patterns of type size D for a certain integer $D > 0$. Set $L := \sup\{\ell(g) : g \in \mathcal{T}\}$. Then, for any $T \in \mathbb{R}_{\geq 0}$ and any set $F \subset B^G(T) \setminus \{e\}$, there exist a subset $A \subset F$ with $\nu(A) \geq (r\theta_T)^{-2D} \nu(F)$ and an element $g \in \mathcal{T}$ such that at least one of the following holds:*

(1) for any integer $k \geq 1$ and any collection $g_1, \dots, g_k \in A$

$$\ell(g_1 \cdots g_k) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2LD);$$

(2) for any integer $k \geq 1$ and any collection $g_1 g, \dots, g_k g \in A \cdot g$

$$\ell(g_1 g \cdots g_k g) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2LD).$$

Observe that $T/\log \theta_T \overset{T \to \infty}{\longrightarrow} \infty$ due to the subexponential growth of $G_1, \dots, G_r$; as a consequence, the factor $(r\theta_T)^{-2D}$, quantifying the maximal loss in probability, is negligible on an exponential scale (cf. the proof of Proposition 4.3).

Proof. The proof consists of a repeated application of the union bound for ν, in order to extract various subsets of F with predetermined letters in their reduced-word expression.

To begin with, there exist $(i_1, j_1) \in \{1, \dots, r\}^2$ and $F_1 \subset F$ such that $\nu(F_1) \geq r^{-2} \nu(F)$ and, for any $g \in F_1$, the first letter of g is in G_{i_1} and the last one is in G_{j_1}. If $i_1 \neq j_1$, then $\ell(g_1 \cdots g_k) = \ell(g_1) + \cdots \ell(g_k)$ for any $g_1, \dots, g_k \in F_1$, so that $A = F_1$ fulfils the statement. If $i_1 = j_1$, we might choose a subset $E_1 \subset F_1$ and elements $y_1, z_1 \in G_{i_1}$ such that $\nu(E_1) \geq \theta_T^{-2} \nu(F_1)$ and, for each $g \in E_1$, the first letter of g is y_1 and the last one is z_1. We distinguish three cases.

– Suppose $\ell(y_1) > L, \ell(z_1) > L$; if g is chosen in $\mathcal{T} \setminus G_{i_1}$, it is easy to check that $\ell(g_1 g \cdots g_k g) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2L)$ for any $g_1 g, \dots, g_k g \in E_1 \cdot g$, so that we may set $A = E_1$ and conclude.

– If just one between y_1 and z_1 has length exceeding L, or alternatively if $\ell(y_1) \leq L, \ell(z_1) \leq L$ and $z_1 \neq y_1^{-1}$, then $\ell(g_1 \cdots g_k) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2L)$ for any $g_1, \dots, g_k \in E_1$; again, setting $A = E_1$ allows to conclude.

The only remaining case is $\ell(y_1) \leq L, z_1 = y_1^{-1}$. We then carry out the same procedure, selecting $F_2 \subset E_1$, $(i_2, j_2) \in \{1, \ldots, r\}^2$, with $\nu(F_2) \geq r^{-2}\mu(E_1)$ and so that, for each $g \in F_2$, the second letter of g is in G_{i_2} and the second-to-last one is in G_{j_2}. If $i_2 \neq j_2$, then $\ell(g_1 \cdots g_k) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2L)$ for any $g_1, \ldots, g_k \in F_2$. If instead $i_2 = j_2$, then choose $E_2 \subset F_2$ and elements $y_2, z_2 \in G_{i_2}$ so that $\nu(E_2) \geq \theta_T^{-2}\nu(F_2)$ and, for each $g \in E_2$, the second letter of g is y_2 and the second-to-last one is z_2. Notice that, by assumption, $\mathcal{T}$ is not contained in any conjugate of any factor G_i by any word ω of type size not exceeding D. Therefore, unless $\ell(y_2) \leq L$ and $z_2 = y_2^{-1}$, we can set $A = E_2$ and conclude as before.

Proceeding in this way, we select, if needed at each successive step, nested subsets $E_2 \supset F_3 \supset E_3 \supset \cdots \supset E_D$. The set E_D has the property that $\nu(E_D) \geq (\theta_T)^{-2}\nu(F_D) \geq (r\theta_T)^{-2D}\nu(F)$; furthermore, there are letters $y_3, \ldots, y_D, z_D$ such that, for any $g \in E_D$, the reduced-word expression of g is $y_1 \cdots y_D \cdots z_D y_{D-1}^{-1} \cdots y_1^{-1}$. It remains to deal with three possibilities, as above.

– Suppose $\ell(y_D) > L, \ell(z_D) > L$, and set $\omega = y_1 \cdots y_{D-1}$. If g is chosen in $\mathcal{T} \setminus \omega G_{i_D}\omega^{-1}$, where G_{i_D} is the factor to which both y_D and z_D belong[9], then $\ell(g_1 g \cdots g_k g) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2DL)$ for any $g_1 g, \ldots, g_k g \in E_D \cdot g$, so that we may set $A = E_D$ and conclude.
– If just one between y_D and z_D has length exceeding L, or alternatively if $\ell(y_D) \leq L, \ell(z_D) \leq L$ and $z_D \neq y_D^{-1}$, then $\ell(g_1 \cdots g_k) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2DL)$ for any $g_1, \ldots, g_k \in E_1$; conclude by setting $A = E_1$.
– Finally, assume $\ell(y_1) \leq L, z_D = y_D^{-1}$, and choose $g \in \mathcal{T}$ not starting with $y_1 \cdots y_D$ nor ending with $(y_1 \cdots y_D)^{-1}$. Then again $\ell(g_1 g \cdots g_k g) \geq \ell(g_1) + \cdots + \ell(g_k) - k(2DL)$ for any $g_1 g, \ldots, g_k g \in E_D \cdot g$. The set $A = E_D$ satisfies the conclusion.

The argument is finalized. $\qquad\square$

If a set A (resp. $A \cdot g$) satisfies the conclusion of Lemma 4.2, then we say that A (resp. $A \cdot g$) has the *weak length additivity property* of order LD.

We are now in a position to prove existence of the weak LDP.

Proposition 4.3. *Let G, S, ℓ, μ be as above, $(Y_n)_{n \geq 0}$ a right random walk on G issued from the identity with increments distributed according to μ. Suppose that the support of μ generates a pattern-avoiding semigroup $\Gamma \subset G$. Then the sequence of $\mathbb{R}$-valued random variables $\left(\frac{1}{n}\ell(Y_n)\right)_{n \geq 1}$ satisfies the weak LDP with a rate function $I: \mathbb{R}_{\geq 0} \to [0, \infty]$.*

Proof. We rely on the criterion phrased in Proposition 3.4, checking that the condition expressed therein is satisfied. Arguing by contradiction, suppose that there exists $x \in \mathbb{R}_{\geq 0}$ such that

$$I(x) \neq \sup_{x \in V \text{ open}} - \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(V). \tag{4.2}$$

As the left-hand side of (4.2) always dominates the right-hand side by definition, this yields

$$I(x) > \sup_{x \in V \text{ open}} - \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(V). \tag{4.3}$$

Notice first that, necessarily, x is strictly positive; indeed, for $x = 0$ the criterion in Proposition 3.4 is trivially satisfied, as $\lim_n \frac{1}{n} \log \mu_n(B(0, \varepsilon))$ exists in $[-\infty, 0]$ for every $\varepsilon > 0$, by subadditivity of ℓ.

As a consequence of (4.3), there exist $\delta, \eta > 0$ such that

$$- \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(B(x, \delta)) > \left(\sup_{\rho > 0} - \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(B(x, \rho)) \right) + \eta. \tag{4.4}$$

---

[9]To select an element g of this sort, concatenate any letter y_D' with ω, in such a way that $\omega y_D'$ is a reduced word; using that $\mathcal{T}$ avoids patterns of type size D, pick $g \in \mathcal{T}$ not starting with $\omega y_D'$ nor ending with $(\omega y_D')^{-1}$.

Fix a positive real number ρ such that $\rho < \inf\{x, \delta\}$; then, by (4.4), there are infinitely many positive integers $n_j, j \geq 1$, for which

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(B(x, \delta)) < \frac{1}{n_j} \log \mu_{n_j}((B(x, \rho))) \; - \eta \; . \tag{4.5}$$

For notational simplicity, denote by

$$\alpha = \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(B(x, \delta)), \; \beta_j = \frac{1}{n_j} \log \mu_{n_j}((B(x, \rho))) \text{ for every } j \geq 1. \tag{4.6}$$

We claim that, if j is taken to be sufficiently large, the inequality $\alpha \geq \beta_j - \eta$ holds, which is opposite to what is given by (4.5), giving the desired contradiction.

The hypothesis on the semigroup Γ ensures the existence of a finite subset $\mathcal{T} \subset \Gamma \setminus \{e\}$ with the following property: there exists an integer $D > 0$ such that, for any reduced word ω of type size D in G, we can find $g \in \mathcal{T}$ not starting in ω and not ending in ω^{-1} (cf. Section 2.2). For any $g \in \mathcal{T}$, choose $t(g) \in \mathbb{N}_{\geq 1}$ and $p(g) \in \mathbb{R}_{>0}$ such that the random walk attains g in $t(g)$ steps with probability $p(g)$, that is $\mathbb{P}(Y_{t(g)} = g) = p(g)$. Define $L = \sup\{\ell(g) : g \in \mathcal{T}\}$, $p = \inf\{p(g) : g \in \mathcal{T}\}$, $t = \sup\{t(g) : g \in \mathcal{T}\}$. Keeping with our earlier notation, let

$$\theta_T = \sup\{|B^{G_i}(T)| : i = 1, \ldots, r\} \quad \text{for any } T \in \mathbb{R}_{\geq 0}.$$

Now choose an integer $j_0 \geq 1$ so that

$$n_{j_0} \geq \sup\big\{(2LD + tx)(\delta - \rho)^{-1}, \eta^{-1}\big(2D(\log r + \log \theta_{n_{j_0}(x+\rho)}) - \log p\big)\big\} \; ;$$

this exists since $T/\log \theta_T \overset{T \to \infty}{\longrightarrow} \infty$ by the subexponential-growth assumption on the factors $G_1, \ldots, G_r$. Define $F = \{g \in G : \ell(g) \in n_{j_0} B(x, \rho)\}$, so that $e^{\beta_{j_0} n_{j_0}} = \mathbb{P}(Y_{n_{j_0}} \in F)$ by (4.6). Notice also that F does not contain the identity as $n_{j_0}(x - \rho) > 0$. Applying Lemma 4.2, with ν being the law of the random variable $Y_{n_{j_0}}$, we can manufacture a set $A \subset F$ and an element $g \in \mathcal{T}$ such that

– $\mathbb{P}(Y_{n_{j_0}} \in A) \geq (r\theta_{n_{j_0}(x+\rho)})^{-2D} e^{\beta_{j_0} n_{j_0}}$ and
– either A or $A \cdot g$ has the weak length additivity property of order LD.

We distinguish two cases.

- First case: A has the weak length additivity property of order LD.

    Define the sequence $n_k = kn_{j_0}, k \geq 1$. Since $n_{j_0} \geq 2LD(\delta - \rho)^{-1}$, there exists $\rho' < \delta$ such that $\rho' - \rho \geq 2n_{j_0}^{-1}LD$. For such a choice of ρ', we have that $\ell(g_1 \cdots g_k) \in n_k B(x, \rho')$ whenever $g_1, \ldots, g_k$ are chosen from A. Therefore, we may estimate, for each $k \geq 1$,

$$\mu_{n_k}(B(x, \rho')) = \mathbb{P}(\ell(Y_{n_k}) \in n_k B(x, \rho')) \geq \mathbb{P}(X_1 \cdots X_{n_{j_0}} \in A, \ldots, X_{n_{k-1}+1} \cdots X_{n_k} \in A)$$

$$\geq \mathbb{P}(Y_{n_{j_0}} \in A)^k \geq \big((r\theta_{n_{j_0}(x+\rho)})^{-2D} e^{\beta_{j_0} n_{j_0}}\big)^k \geq e^{n_k(\beta_{j_0} - \eta)} \; ,$$

    where the middle inequality is given by independence and stationarity of the process $(X_n)_{n \geq 1}$, while the last one comes from our choice $n_1 = n_{j_0} \geq 2D\eta^{-1}(\log r + \log \theta_{n_{j_0}(x+\rho)})$. Lemma 4.1 gives

$$\alpha = \liminf_{n \to \infty} \frac{1}{n} \log \mu_n B(x, \delta) \geq \beta_{j_0} - \eta \; ,$$

    as desired.

- Second case: $A \cdot g$ has the weak length additivity property of order LD.

    Define the sequence $n_k = k(n_{j_0} + t(g)), k \geq 1$. Since $n_{j_0} \geq (\delta - \rho)^{-1}(2LD + tx)$, it is possible to select $\rho' < \delta$ so that $\rho' - \rho \geq n_{j_0}^{-1}(2LD + tx)$. It is straightforward to verify that this choice of ρ' ensures $\ell(g_1 g \cdots g_k g) \in n_k B(x, \rho')$ for every $g_1, \ldots, g_k \in A$.

As before, we may thus estimate

$$\mu_{n_k}(B(x, \rho')) = \mathbb{P}(\ell(Y_{n_k}) \in n_k B(x, \rho')) \geq \mathbb{P}(X_1 \cdots X_{n_{j_0}} \in A, X_{n_{j_0}+1} \cdots X_{n_{j_0}+t(g)} = g)^k$$

$$\geq \mathbb{P}(Y_{n_{j_0}} \in A)^k p(g)^k \geq \left((r\theta_{n_{j_0}(x+\rho)})^{-2D} e^{\beta_{j_0} n_{j_0}}\right)^k p^k \geq e^{n_k(\beta_{j_0} - \eta)}$$

for each $k \geq 1$. This time, the last inequality stems from our choice $n_{j_0} \geq \eta^{-1}\big(2D(\log r + \log \theta_{n_{j_0}(x+\rho)}) - \log p\big)$. Applying Lemma 4.1 once more, we deduce again that $\alpha \geq \beta_{j_0} - \eta$.

The proof is concluded. $\qquad\square$

Proposition 4.4. *In the setting of Proposition 4.3, assume further that μ has a finite exponential moment. Then the rate function I governing the weak LDP for the sequence $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$ is proper, and the sequence $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$ satisfies the full LDP with rate function I.*

Proof. As before, we let μ_n be the law of the random variable $\frac{1}{n}\ell(Y_n)$, for every $n \geq 1$. In light of Proposition 3.3, it suffices to show that the sequence $(\mu_n)_{n\geq 1}$ is exponentially tight. By the assumption, there exists a real number $\tau > 0$ such that $C := \int_G \exp(\tau\ell(g))d\mu(g) < \infty$.

Fix $M > 0$. Then

$$\mu_n([0, M]^c) = \mathbb{P}(\ell(Y_n) > nM) = \mathbb{P}(\exp\tau\ell(Y_n) > \exp\tau nM) \leq \frac{\mathbb{E}[\exp\tau\ell(Y_n)]}{\exp\tau nM} ,$$

the last upper bound being given by Markov's inequality. Subadditivity of the length function ℓ, together with independence and stationarity of the process $(X_n)_{n\geq 1}$, gives

$$\mathbb{E}[\exp(\tau\ell(Y_n))] \leq \mathbb{E}\left[\exp\tau\left(\sum_{i=1}^n \ell(X_n)\right)\right] = \mathbb{E}\left[\prod_{i=1}^n \exp\tau\ell(X_i)\right] = \prod_{i=1}^n \mathbb{E}[\exp\tau\ell(X_i)]$$

$$= (\mathbb{E}[\exp\tau\ell(X_1)])^n = \left(\int_G e^{\tau\ell(g)}d\mu(g)\right)^n = C^n \text{ for every } n \geq 1.$$

Combining the previous two estimates, taking the logarithm and dividing by n, we obtain $\frac{1}{n}\log\mu_n([0, M]^c) \leq \log C - \tau M$. As a result,

$$\limsup_{n\to\infty} \frac{1}{n}\log\mu_n([0, M]^c) \xrightarrow{M\to\infty} -\infty ,$$

which establishes exponential tightness of the sequence $(\mu_n)_{n\geq 1}$. $\qquad\square$

5. Convexity of the rate function

The chief aim of this section is the proof of convexity of the rate function associated to the LDP for the sequence $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$. In the last part, we gather some further properties of the rate function, and deduce its characterization expressed in the last sentence of Theorem 1.4. As in the foregoing section, we let μ_n denote the law of the random variable $\frac{1}{n}\ell(Y_n)$, for $n \geq 1$.

Recall that, if X is a real vector space, a function $f: X \to (-\infty, +\infty]$ is convex if, for any $x_1, x_2 \in X$ and any $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) ; \tag{5.1}$$

the function f is mid-point convex if the previous inequality holds for $\lambda = 1/2$, that is if

$$f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) \leq \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2)$$

for all $x_1, x_2 \in X$.

Suppose now X is a topological (real) vector space. By iteration, a mid-point convex function f satisfies the inequality (5.1) for any $\lambda \in \{k/2^n : n \in \mathbb{N}, k \in \{0, \ldots, 2^n\}\}$. The latter set being dense in $[0, 1]$, (5.1) can be extended to all $\lambda \in [0, 1]$ by a standard approximation argument,

provided that we know that f is lower semicontinuous. To wrap up, a lower semicontinuous, mid-point convex function $f\colon X \to (-\infty. +\infty]$ is convex.

Proposition 5.1. *Let $G, S, \ell, \mu, (Y_n)_{n\geq 0}$ be as in Proposition 4.3. Then the rate function I, governing the LDP for the sequence of $\mathbb{R}$-valued random variables $\left(\frac{1}{n}\ell(Y_n)\right)_{n\geq 1}$, is convex.*

The proof bears a lot of resemblance with the proof of Proposition 4.3; for the sake of conciseness, we shall omit a few details.

Proof. As observed in the previous paragraph, it suffices to show that I is mid-point convex, since we already know I that is lower semicontinuous. Again, we argue by contradiction: assume there exist $x_1 < x_2 \in \mathbb{R}$ such that

$$I\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) > \frac{1}{2}I(x_1) + \frac{1}{2}I(x_2) . \tag{5.2}$$

Recall that we have

$$I(x) = \sup_{x\in V \text{open}} -\liminf_{n\to\infty} \frac{1}{n}\log\mu_n(V) = \sup_{x\in V \text{open}} -\limsup_{n\to\infty} \frac{1}{n}\log\mu_n(V) \quad \text{for all } x \in X;$$

therefore, (5.2) implies that there exist $\delta, \eta > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\log\mu_n\left(B\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \delta\right)\right) <$$
$$< \frac{1}{2}\left(\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(B(x_1,\rho_1)) + \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(B(x_2,\rho_2))\right) - \eta \tag{5.3}$$

for any $\rho_1, \rho_2 > 0$. Notice that this forces in particular $x_1, x_2 \in \mathbb{R}_{\geq 0}$. Choose $\rho := \rho_1 = \rho_2 < \delta$. For a sufficiently large n_0 and every $n \geq n_0$, we claim that there exists $\phi(n) \in \{2n, \ldots, 2n+t\}$ such that

$$\frac{1}{\phi(n)}\log\mu_{\phi(n)}\left(B\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \delta\right)\right) \geq \frac{1}{2}\left(\frac{1}{n}\log\mu_n(B(x_1,\rho)) + \frac{1}{n}\log\mu_n(B(x_2,\rho))\right) - \eta . \tag{5.4}$$

Letting n vary over an arithmetic progression for which the corresponding sequence of $\phi(n)$ is strictly increasing, it is clear that we obtain a contradiction to (5.3).

It remains to prove the claim just stated. Let $\mathcal{T} \subset \Gamma \setminus \{e\}$ be a finite set avoiding patterns of type size D, and fix $n \geq n_0$; let $F_i = \{g \in G : \ell(g) \in nB(x_i, \rho)\}$, $i = 1, 2$. Adapting the proof of Lemma 4.2 appropriately[10], we deduce that there is an element $g \in \mathcal{T}$ and subsets $A_i \subset F_i$ such that $\mathbb{P}(Y_n \in A_i) \geq (r\theta_{n(x+\rho_i)})^{-D}\mathbb{P}(Y_n \in F_i)$ and

  – either for any $g_1 \in A_1, g_2 \in A_2$ it holds $\ell(g_1 g_2) \geq \ell(g_1) + \ell(g_2) - 2LD$,
  – or for any $g_1 \in A_1, g_2 \in A_2$, $\ell(g_1 g g_2) \geq \ell(g_1) + \ell(g_2) - 2LD$.

In the first case, we get the inequality (5.4) for $\phi(n) = 2n$, by observing that $g_1 \in A_1, g_2 \in A_2$ imply $\ell(g_1 g_2) \in 2nB((x_1 + x_2)/2, \delta)$; in the second case, we get it for $\phi(n) = 2n + t(g)$. We refer to the proof of Proposition 4.3 for the missing details. $\qquad\square$

5.1. **Further properties of the rate function.** We list below some additional properties of the rate function, emphasizing connections with other relevant quantities associated to the random walk, such as the rate of escape and the spectral radius.

  (1) Since $\frac{1}{n}\ell(Y_n)$ converges to the escape rate λ almost surely, I has a zero at $x = \lambda$.

---

[10]There is a minor nuisance here if $x_1 = 0$, as F_1 contains the identity; replacing F_1 with $F_1 \setminus \{e\}$ results in harmless modifications of the probabilities involved.

(2) Convexity of the rate function I gives, as an immediate corollary, that its effective domain D_I is a convex subset of $\mathbb{R}_{\geq 0}$, hence a (possibly degenerate[11]) sub-interval of the positive half-line. Standard properties of convex functions defined on sub-intervals of the real line imply that, on the open interval D_I°, the rate function I is continuous, admits left and right derivatives at every point, and it is differentiable outside a countable set of points. In particular, continuity on D_I° gives that

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_n(V) = -\inf_{x \in V} I(x) \quad \text{for every open set } V \subset D_I^\circ \; ;$$

in other words, the exponential decay rate of the sequence $(\mu_n(V))_{n \geq 1}$ is well-defined whenever V is an open subset of D_I°.

(3) Define the *spectral radius* of the random walk as

$$\rho = \limsup_{n \to \infty} \mathbb{P}(Y_n = e)^{\frac{1}{n}} \in [0, 1] \; .$$

If the measure μ is symmetric, that is $\mu(g) = \mu(g^{-1})$ for every $g \in G$, this quantity coincides with the spectral radius of the *Markov operator* associated with the random walk (cf. [25, Chap. 6]). For every $\delta > 0$, we have

$$\mu_n(B(0, \delta)) = \mu_n([0, \delta)) \geq \mu_n(0) = \mathbb{P}(\ell(Y_n) = 0) = \mathbb{P}(Y_n = e),$$

which implies

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(B(0, \delta)) \geq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Y_n = e) = \log \rho \; ,$$

with the understanding that $\log \rho = -\infty$ if $\rho = 0$. The previous inequality holding for every $\delta > 0$, we infer that

$$I(0) = \sup_{\delta > 0} - \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(B(0, \delta)) \leq -\log \rho \; . \tag{5.5}$$

As a consequence, we deduce that $0 \in D_I$ provided that the spectral radius is strictly positive. This occurs, for instance, whenever the semigroup Γ generated by $\operatorname{supp} \mu$ contains e: if $n_0 \in \mathbb{N}$ is any integer for which $\mathbb{P}(Y_{n_0} = e) > 0$, then

$$\rho \geq \limsup_{k \to \infty} \mathbb{P}(Y_{kn_0} = e)^{\frac{1}{kn_0}} \geq \limsup_{k \to \infty} \left(\mathbb{P}(Y_{n_0} = e)^k \right)^{\frac{1}{kn_0}} > 0.$$

It is worth mentioning that equality $I(0) = -\log \rho$ actually holds[12], whenever the LDP for the word length functional is verified and the measure μ driving the random walk satisfies $\inf\{\mu(g) : g \in \operatorname{supp} \mu\} > 0$ (see [27, Lem. 2.8]).

(4) As far as the least upper bound of D_I is concerned, assume that the support of μ is bounded, and let $L = \sup\{\ell(g) : g \in \operatorname{supp} \mu\} < \infty$. Then $I \equiv \infty$ on the open half-line (L, ∞), as subadditivity of ℓ implies $\ell(Y_n) \leq nL$ $\mathbb{P}$-almost surely for any $n \geq 1$. Therefore, in this case, D_I is contained in $[0, L]$.

If no restriction is placed on the size of $\operatorname{supp} \mu$, then $\sup D_I$ may be infinite[13].

---

[11]In general, the rate function I can be as degenerate as possible: for instance, if $G = \langle a, b \rangle$ is a free group on two generators, and $\mu(a) = p = 1 - \mu(b)$ for some $p \in [0, 1]$, then $I(1) = 0$ and $I(x) = \infty$ for any $x \in \mathbb{R}_{\geq 0} \setminus \{1\}$, as $\ell(Y_n) = n$ $\mathbb{P}$-almost surely for every n.

[12]We thank S. Müller for communicating this fact.

[13]Consider, once again, $G = \langle a, b \rangle$ a free group on two generators, and choose a measure μ with $\operatorname{supp} \mu = \langle a \rangle$. Then $\mathbb{P}(\ell(Y_n) = nk) \geq (\mu(a^k))^n$ for all integers $n, k \geq 1$, so that $I(k) < \infty$ for any $k \geq 1$. In this example, we have thus $D_I = \mathbb{R}_{\geq 0}$.

5.2. **The rate function as a Fenchel-Legendre transform.** It remains to prove the final statement of Theorem 1.4, under the assumption that μ has finite moment-generating function. By virtue of Theorem 3.5, it suffices to prove that the limiting logarithmic moment generating function of the sequence $(\mu_n)_{n \geq 1}$, given by

$$\Lambda(z) = \limsup_{n \to \infty} \frac{1}{n} \log \int_{\mathbb{R}} e^{nz \cdot x} d\mu_n(x) = \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{z \cdot \ell(Y_n)}] \ , \ z \in \mathbb{R},$$

is finite everywhere, where we have canonically identified $\mathbb{R}$ with its dual space, and the dual pairing with the standard product of real numbers.

Fix $z \in \mathbb{R}_{\geq 0}$; then $\mathbb{E}[e^{z \cdot \ell(Y_1)}] = \int_G \exp(z\ell(g)) \, d\mu(g) < \infty$, since all exponential moments of μ are finite. Moreover, for any $n, m \geq 1$, we have

$$\mathbb{E}[e^{z \cdot \ell(Y_{n+m})}] \leq \mathbb{E}[e^{z \cdot \ell(X_1 \cdots X_n)} e^{z \cdot \ell(X_{n+1} \cdots X_{n+m})}] = \mathbb{E}[e^{z \cdot \ell(Y_n)}]\mathbb{E}[e^{z \cdot \ell(Y_m)}] \ ;$$

the first inequality comes from subadditivity of the length function ℓ, whereas the second follows from independence and stationarity of the process $(X_n)_{n \geq 1}$. Therefore, the sequence

$$a_n = \log \mathbb{E}[e^{z \cdot \ell(Y_n)}] \ , \ n \geq 1, \tag{5.6}$$

is subadditive, that is $a_{n+m} \leq a_n + a_m$ for every $n, m \geq 1$; Fekete's lemma ([25, Ex. 3.9]) gives

$$\Lambda(z) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{z \cdot \ell(Y_n)\rangle}] = \inf_{n \geq 1} \frac{1}{n} \log \mathbb{E}[e^{z \cdot \ell(Y_n)\rangle}] \leq \mathbb{E}[e^{z \cdot \ell(Y_1)}] < \infty \ .$$

If $z \in \mathbb{R}_{<0}$, a similar argument shows that the sequence (5.6) is superadditive, and $\Lambda(z) < \infty$ follows all the same.

6. Concluding remarks and open questions

6.1. **Groups with strongly connected finite-state automata.** We mention another class of examples to which our method would apply: finitely generated groups whose cone type automaton with respect to a given generating set is finite and strongly connected.

Let G be a finitely generated group, S a finite set of generators, ℓ the word length defined by S on G. For every element $g \in G$, we define the *cone type* of g as the set

$$C(g) = \{h \in G : \ell(gh) = \ell(g) + \ell(h)\}.$$

Notice that the usual definition of cone type which appears in the literature ([9, 12, 28]) involves geodesic words in the alphabet S, rather that actual group elements of G; our definition is more convenient for the purposes of this discussion.

The cone type of an element selects those geodesic segments that can be attached (in algebraic terms, multiplied) to it on the right so that the concatenation is again a geodesic segment. Observe that it is precisely this notion that, implicitly, comes into play both in the proof of existence of LDP and in the proof of convexity of the rate function.

Cone types offer an algorithmic way to label geodesics in the group G, in other words to identify those strings $(s_1, \ldots, s_n)$ of letters in the alphabet S such that $\ell(s_1 \cdots s_n) = n$. This is achieved through the construction of a finite state automaton (cf. [12]), called the *cone type automaton* of G with respect to the language given by S. Assume there are only finitely many cone types $C_0 = C(e), C_1, \ldots, C_s$, which we view as vertices of a directed graph Δ whose edges are labelled by elements of S; more precisely, we connect the cone type $C(g)$ of an element g to the cone type of $C(gs)$, via a directed edge labelled by $s \in S$, if and only if $s \in C(g)$. It is immediate that the definition doesn't depend on the choice of g but only on its cone type. If $e \notin S$, there is a one-to-one correspondence between edge-paths in the directed graph Δ starting at C_0 and finite sequences $(s_1, \ldots, s_n) \in S^n$ such that $\ell(s_1 \cdots s_n) = n$, that is geodesic words in the alphabet S.

Now, the conditions we need to impose in order for the arguments of Sections 4 and 5 to carry over unaffectedly are:

(1) the finite directed graph Δ is strongly connected, meaning that there is a directed path joining any two of its vertices;

(2) every element of G belongs to the cone type of some non-trivial element; otherwise stated, for any geodesic word $\omega = (s_1, \ldots, s_n)$ in the alphabet S, there is a cone type $C \neq C_0$ from which we can follow a directed path in the graph Δ according to the labelling given by ω.

Example 6.1 (Simple random walks on integer lattices)**.** Consider $G = \mathbb{Z}^d$ with its standard symmetric set of generators $S = \{\pm e_i : 1 \leq i \leq d\}$. Any probability distribution μ with $\operatorname{supp} \mu \subset S$ gives rise to a simple random walk $(Y_n)_{n \in \mathbb{N}}$ on $\mathbb{Z}^d$. It is clear that there are exactly $2^d + 2d + 1$ different cone types (the 2^d quadrants, the $2d$ half-spaces delimited by the d coordinate planes, and the whole $\mathbb{Z}^d$). It takes a moment to realize that both conditions stated above are met. We thus recover, by elementary means, existence of the LDP with convex rate function for the process $\frac{1}{n} \|Y_n\|_1$ (where $\|(x_1, \ldots, x_d)\|_1 = |x_1| + \cdots |x_d|$ for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$), which is usually seen as a straightforward consequence of Cramer's theorem for the empirical mean of i.i.d. random vectors (see [10, Thm. 2.2.30]).

Finiteness of the number of cone types appears to be an intrinsic requirement when attempting to establish the LDP via the strategy presented here, while the two additional conditions on the cone type automaton mentioned above can be presumably lifted through a refinement of the method.

A large class of finitely generated groups having only finitely many cone types, with respect to any finite generating set, is given by Gromov-hyperbolic groups; indeed, in such groups the cone type of an element only depends on its k-*tail*, for a fixed positive integer k depending only on the group (see [9]). Our considerations thus provide substance to the claim that Theorem 1.4 holds for any Gromov-hyperbolic group[14].

6.2. **Some open problems.** Computing the exact expression of the rate function, in the cases treated by Theorem 1.4, is mostly out of reach; however, it is worth carrying through the computation in the easiest case of symmetric simple random walks on free groups, to get a flavour of what should happen in more general circumstances. This has already been performed in [33]: let G be a free group on $r \geq 1$ generators, $S = \{a_1, \ldots, a_r\}$ a free generating set, μ the uniform probability measure on $S \cup S^{-1}$, i.e. $\mu(a_i) = \mu(a_i^{-1}) = (2r)^{-1}$ for any $i \in \{1, \ldots, r\}$. The rate function governing the LDP for the sequence $\left(\frac{1}{n}\ell(Y_n)\right)_{n \geq 1}$ is given by the following expression:

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

where we agree that $0 \log 0 = 0$. The function I is analytic in $(0, 1)$ and strictly convex in its effective domain $[0, 1]$, and hence admits a unique zero at $\lambda = 1 - 1/r$, corresponding to the escape rate of the random walk; as a consequence thereof, the probability $\mathbb{P}\left(|\frac{1}{n}\ell(Y_n) - \lambda| \geq \varepsilon\right)$ that the renormalized length deviates largely from the escape rate decays exponentially fast

---

[14](Added in revision) Gouëzel has shown ([14, Lem. 2.4]) that a non-elementary hyperbolic group G equipped with a word length ℓ satisfies the following geometric property: there exist constants $c, C > 0$ such that, for any $x, y \in G$, there is an element $a \in G$ of length at most C such that $\ell(xay) \geq \ell(x) + \ell(y) - c$. The result has been subsequently extended in [11, Lem. 5.3] to relatively hyperbolic groups. It can be used as a replacement of almost length additivity throughout the proof of Theorem 1.4, thereby proving its validity for irreducible random walks on any relatively hyperbolic group, with respect to any word length. The resulting argument simplifies the proof of [6, Thm. 1.2], which however addresses more general spaces and walks, and yields a finer result on the rate function.

with n for any $\varepsilon > 0$. Furthermore, the value of I at 0 is equal (in absolute value) to the logarithm of the spectral radius, as expected. Lastly, we notice that the right derivative $I'(0)$ at 0 is finite, while the left derivative $I'(1)$ at 1 is infinite.

This motivates the following questions:

(1) Is the rate function I in Theorem 1.4 always strictly convex? In particular, does it always have a unique zero at $x = \lambda$?
(2) What are the finer regularity properties of the rate function? What is the behaviour of the (one-sided) derivatives of I at the extreme points of its effective domain?

Assuming the validity of Theorem 1.4 for Gromov-hyperbolic groups, the same questions can obviously be phrased in this broader context as well.

References

[1] H. Abels, G. Margulis and A. Soifer, *Semigroups containing proximal linear maps*, Israel J. Math. **91** (1995), 1-30.
[2] Y. Benoist, *Propriétés asymptotiques des groupes linéaires*, Geom. Func. Anal. **7** (1997), 1-47.
[3] Y. Benoist, J.F. Quint, *Central limit theorem for linear groups*, Ann. Probab. **44** (2016), 1308-1340.
[4] Y. Benoist, J.F. Quint, *Central limit theorem on hyperbolic groups*, Izv. Akad. Nauk Ser. Mat. **80** (2016), 3-23.
[5] M. Bjorklund, *Central limit theorem for Gromov hyperbolic groups*, J. Theoret. Probab. **23** (2010), 871-887.
[6] A. Boulanger, P. Mathieu, C. Sert and A. Sisto, *Large deviations for random walks on hyperbolic spaces*, arXiv:2008.02709v1 (2020).
[7] N. Bourbaki. *Éléments de mathématique, Algèbre*.
[8] H. Brezis. *Functional analysis, Sobolev spaces and Partial Differential Equations*. Springer, New York, 2011.
[9] M.R. Bridson, A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 1999.
[10] A. Dembo, O. Zeitouni. *Large deviations Techniques and Applications*. Stochastic Modelling and Applied Probability, Springer, Berlin, 2010.
[11] M. Dussaule, *Local limit theorems in relatively hyperbolic groups I: rough estimates*, Ergodic Theory Dynam. Systems (2021), 1-41, doi.10.1017/etds.2021.7.
[12] D.B.A Epstein et al. *Word Processing in Groups*. Johns and Bartlett Publishers, Boston, MA, 1992.
[13] E. Ghys, P. de la Harpe. *Sur les groupes hyperboliques d'après Mikhael Gromov*. Progress in Mathematics, Birkhaüser, Boston, MA, 1990.
[14] S. Gouëzel, *Local limit theorem for symmetric random walks in Gromov-hyperbolic groups*, J. Amer. Math. Soc. **27** (2014), 893-928.
[15] M. Gromov, *Hyperbolic groups*, MSRI Publ. **8** (1987), 75-263.
[16] Y. Guivarc'h, *Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire*, Astérisque **74** (1980), 47-98.
[17] P. de la Harpe. *Topics in Geometric Group Theory*. Chicago Lectures in Mathematics, The University of Chicago Press, Chicago, 2000.
[18] H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336-354.
[19] J.F.C. Kingman, *The ergodic theory of subadditive processes*, J. Roy. Statist. Soc. Ser. B **30** (1968), 499-510.
[20] A. Klenke. *Probability theory. A Comprehensive Course*. Second Edition. Universitext, Springer-Verlag, London, 2014.
[21] S.P. Lalley, *Finite range random walks on free groups and homogeneous trees*, Ann. Probab. **21** (1993), 2087-2130.
[22] O.E. Lanford, *Entropy and equilibrium states in classical statistical mechanincs*, Statistical Mechanics and Mathematics Problems, Lecture Notes in Physics, Springer-Verlag, Berlin, 1973.
[23] F. Ledrappier, *Some asymptotic properties of random walks on free groups*, CRM Proceedings and Lecture Notes **28** (2001), 117-152.
[24] R.C. Lyndon, *Grushko's theorem*, Proc. Amer. Math. Soc. **16** (1965), 822-826.
[25] R. Lyons, Y. Peres. *Probability on Trees and Networks*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York, 2016.
[26] P. Mathieu, A. Sisto, *Deviation inequalities for random walks*, Duke Math. J. **169** (2020), 961-1036.
[27] S. Müller, *Recurrence for branching Markov chains*, Electron. Commun. Probab. **13** (2008), 576-605.

[28] K. Ohshika. *Discrete groups.* Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 2002.

[29] G. Polya, *Über eine Aufgabe der Wahrscheinlichkeitstheorie betreffend die Irrfahrt im Straßennetz*, Math. Ann. **84** (1921), 149-160.

[30] J.F. Quint, *Cônes limites des sous-groupes discrets des groupes réductifs sur un corps local*, Transformation groups **7** (2002), 247-266.

[31] S. Sawyer, T. Steger, *The rate of excape for anisotropic random walks in a tree*, Probab. Theory Related Fields **76** (1987), 207-230.

[32] J.P. Serre. *Arbres, Amalgames,* SL$_2$. Astérisque, Soc. Math. France, 1977.

[33] C. Sert, *Joint Spectrum and Large Deviation Principles for Random Matrix Products*, PhD Thesis, Université Paris-Sud, 2016.

[34] C. Sert, *Joint spectrum and large deviation principle for random matrix products*, C. R. Acad. Sci. Paris, Ser. I **355** (2017), 718-722.

[35] C. Sert, *Large deviation principle for random matrix products*, Ann. Probab. **47** (2019), 1335-1377.

[36] W. Woess. *Random Walks on Infinite Graphs and Groups.* Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2000.

[37] J. Wolf, *Growth of finitely generated solvable groups and curvature of Riemannian manifolds*, J. Diff. Geom. **2** (1968), 421-446.

(E. C.) ETH Zürich, Rämistrasse 101 CH-8092 Zürich Switzerland

Email address: emilio.corso@math.ethz.ch