Persistent Hall rays for Lagrange spectra at cusps of Riemann surfaces

PERSISTENT HALL RAYS FOR LAGRANGE SPECTRA AT CUSPS OF RIEMANN SURFACES

MAURO ARTIGIANI, LUCA MARCHESE, AND CORINNA ULCIGRAI

Abstract. We study Lagrange spectra at cusps of finite area Riemann surfaces. These spectra are penetration spectra that describe the asymptotic depths of penetration of geodesics in the cusps. Their study is in particular motivated by Diophantine approximation on Fuchsian groups. In the classical case of the modular surface and classical Diophantine approximation, Hall proved in 1947 that the classical Lagrange spectrum contains a half-line, known as a Hall ray. We generalize this result to the context of Riemann surfaces with cusps and Diophantine approximation on Fuchsian groups. One can measure excursion into a cusp both with respect to a natural height function or, more generally, with respect to any proper function. We prove the existence of a Hall ray for the Lagrange spectrum of any non co-compact, finite covolume Fuchsian group with respect to any given cusp, both when the penetration is measured by a height function induced by the imaginary part as well as by any proper function close to it with respect to the Lipschitz norm. This shows that Hall rays are stable under (Lipschitz) perturbations. As a main tool, we use the boundary expansion developed by Bowen and Series to code geodesics and produce a geometric continued fraction-like expansion and some of the the ideas in Hall's original argument. A key element in the proof of the results for proper functions is a generalization of Hall's theorem on the sum of Cantor sets, where we consider functions which are small perturbations in the Lipschitz norm of the sum.

1. Introduction

The classical Lagrange spectrum is a well studied subset of the real line, which can be described either in terms of Diophantine approximation or dynamics, as penetration spectrum of geodesics on the modular surface (see Section 1.1). Hall proved in 1947 that the classical Lagrange spectrum contains a semi-infinite interval, known as a Hall ray. We generalize this result to the context of Riemann surfaces with cusps and Diophantine approximation on Fuchsian groups, answering a question which was left open despite many results in the literature for various geometric generalizations of these spectra (see Section 1.2). Definitions of the Lagrange spectra we study and main results are presented in Section 1.3 and Section 1.5.

1.1. The classical Lagrange spectrum.

Classical Diophantine approximation is the study of how well one can approximate a real number α by rational ones. The well-known results of Dirichlet and Hurwitz imply that for all irrational real numbers α there are infinitely many p/q, $p \in \mathbb{Z}, q \in \mathbb{N}$, such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{\sqrt{5}q^2}$$

and that this is the best possible result for *every* real number α, as one can see by considering the golden mean $\alpha = \frac{1+\sqrt{5}}{2}$. A natural question is hence if *fixing* α one can improve the constant appearing in the denominator. This leads to the introduction of the (classical) *Lagrange spectrum* $\mathcal{L} \subset \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ as follows. For a given $\alpha \in \mathbb{R}$, let $L(\alpha) \in \overline{\mathbb{R}}$ be such that

$$L(\alpha) := \sup\{k : |\alpha - p/q| < 1/kq^2 \text{ for infinitely many } q \in \mathbb{N},\ p \in \mathbb{Z}\}.$$

Then $\mathcal{L}$ is the collection of values $\{L(\alpha), \alpha \in \mathbb{R}\}$. Equivalently, one can also write (see for example [26])

$$\mathcal{L} = \left\{ L(\alpha) = \limsup_{q,p\to+\infty} \frac{1}{q|q\alpha - p|}, \quad \alpha \in \mathbb{R} \right\} \subset \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}. \tag{1.1}$$

One can see that for almost every α one has $L(\alpha) = \infty$, but $L(\alpha) < \infty$ for a set of full Hausdorff dimension, which consists exactly of so called *badly approximable* (or bounded type) numbers.

Date: May 8, 2023.

2010 *Mathematics Subject Classification.* Primary 11J06, 37D40; Secondary 30F35.

Key words and phrases. Lagrange spectrum, Hall ray, Diophantine Approximation, Fuchsian groups, Boundary Expansions, Penetration spectra, Cantor sets.

A close relative of the Lagrange spectrum is the *Markoff spectrum* $\mathcal{M}$ obtained by replacing the lim sup in (1.1) with a sup. Both Markoff and Lagrange spectra have been intensively studied by many authors, and much of their beautiful and rich structures is known. Both $\mathcal{L}$ and $\mathcal{M}$ are closed subsets of $\overline{\mathbb{R}}$, with the strict inclusion $\mathcal{L} \subset \mathcal{M}$. It is for instance known that:

- The minimum of $\mathcal{L}$ and $\mathcal{M}$ is $\sqrt{5}$, which is known as *Hurwitz constant* [21, 25];
- $\mathcal{L} \cap (0,3) = \mathcal{M} \cap (0,3)$ is an explicit discrete set that accumulates to 3 (see for example [8]);
- $\mathcal{L}$ contains a semi-infinite interval $[R, \infty)$ [16]. This part of the Lagrange spectrum is called the *Hall ray*. The exact value of R where the Hall ray begins (i.e. the smallest r such that $[r, \infty) \subset \mathcal{L}$) is known after the work of Freiman [13] and hence known as *Freiman constant*.

The (classical) Lagrange spectrum admits also a geometro-dynamical interpretation as the spectrum of asymptotic depths of penetration for the geodesic flow into the cusp of the modular surface $X = \mathrm{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$. More precisely, the Lagrange spectrum is the set of values $L \in \overline{\mathbb{R}}$ which can be realized as

$$L = \limsup_{t\to+\infty} \mathrm{height}(\gamma(t)), \tag{1.2}$$

for a parametrized geodesic $\gamma(t)$ on $X = \mathrm{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$, where height$(x)$ denotes the hyperbolic *height* in the cusp, given by the imaginary part of the unique lift of x from the modular surface to its classical fundamental domain $\mathcal{F} = \{z \in \mathbb{H} : |z| > 1, |\operatorname{Re}(z)| < \frac{1}{2}\}$.

Remark 1.1. If one replaces the lim sup for $t \to +\infty$ in (1.2) with a lim sup for $|t| \to +\infty$, the value on a given γ might change (since it then also depends on the backward endpoint of γ), but one can show, choosing symmetric geodesics that the set of such values taken on all geodesics still produces exactly $\mathcal{L}$ (see the Appendix of [8] for a proof).

Over the course of time, both the Markoff and the Lagrange spectrum have been generalized to many different contexts, either from the Diophantine approximation or from the geometric point of view, exploiting their dynamical definition as penetration spectra.

1.2. **A brief history of generalizations of Markoff and Lagrange spectra.** We do not attempt here to summarize all the developments in this area, which started more than a century ago and has seen a surge of recent developments, but we will only briefly survey some of the results, in particular those which are closer to the main topic of this paper, namely the presence of Hall rays. The interested reader can find further information in the monograph [8] by Cusick and Flahive, the introduction of [20] and the recent survey by Matheus [26], and refer to the references therein.

The first natural generalization of the classical Lagrange and Markoff spectra is obtained by replacing the modular group $\mathrm{PSL}(2,\mathbb{Z})$ with a more general Fuchsian group. Both the dynamical and Diophantine approximation definition extend naturally to this context (see Section 1.3). In particular, important classes of examples are given by the cases of Hecke groups and more generally triangle groups. The minimum value in these spectra, which is also called, as in the classical case, the *Hurwitz constant*, is computed for Hecke and triangle groups respectively by Haas and Series in [15] and Vulakh in [42]. Markoff spectra of Fuchsian groups were studied in detail by Vulakh in [43, 44]; in particular, in [43] the author gives the complete description of the discrete part of the Markoff spectrum (and hence of the Lagrange spectrum) of any Hecke group.

Another natural generalization leads to study Markoff and Lagrange spectra for quotient of higher dimensional hyperbolic spaces by discrete subgroups [40]. In particular, the case of Bianchi groups has connections with the approximation of a complex number with numbers from a given imaginary quadratic number field, see [41, 27].

Penetration spectra and more general objects, such as spiraling spectra, can be studied more generally in the context of (variable) negative curvature, see in particular the works by Paulin in collaboration with Hersonsky [18] and Parkkonen [32]. In [32] it is shown that both the Lagrange and Markoff spectrum of a finite volume Riemannian manifold with sectional curvature less than -1 and *dimension at least* 3 contain a Hall ray. Remarkably, Parkkonen and Paulin also managed to obtain a *universal* estimate on the beginning of both spectra.

In the case of surfaces, Schmidt and Sheingorn proved in [35] that the Markoff spectrum of a hyperbolic surface of constant negative curvature -1 contains a Hall ray. Recently Moreira and Romaña proved that, for generic small perturbations of dynamically defined analogues of the Lagrange and Markoff spectra on negatively curved surfaces, these spectra contain intervals arbitrarily close to infinity. We stress that neither of these two results does however imply the existence of Hall rays for Lagrange spectra. In a similar spirit, in [28] continuity of the Hausdorff dimension of the Spectra, when intersected with the open interval $(-\infty, t)$ for $t \in \mathbb{R}$, is proved. Very recently, in [7], the same result is proved for generic perturbations of dynamically defined spectra on negatively curved surfaces. An introduction of Moreira's work as well as the classical theory of these spectra can be found in [26].

Another generalization is introduced by Hubert and two of the authors in [20], where Lagrange spectra are defined in the context of translation surfaces and interval exchange transformations. Also in this case one has a geometric definition as penetration spectra for the Teichmüller geodesic flow, as well as an interpretation motivated by Diophantine approximation for interval exchange maps, see [20]. A version of the latter already appears in the work by Boshernitzan, see [4]; different types of Lagrange spectra for interval exchange transformations, in particular in the case of 3 interval exchanges, are also studied by Ferenczi in [10]. For Lagrange spectra of *strata* of translation surfaces, the existence of Hall rays was in [20] and the Hurwitz constant was recently found by Boshernitzan and Delecroix [5]. The authors proved in [1] that also for the particularly symmetric class of translation surfaces made of Veech surfaces, the Lagrange spectrum contained a Hall ray and the first values of the Lagrange spectrum for a particular example of a square-tiled Veech surface are studied in detail in [19].

The generalizations of Lagrange spectra we study in this paper are in the context of Diophantine approximation on Fuchsian groups (see Section 1.3) and penetration spectra for Riemann surfaces with cusps with respect to proper functions (see Section 1.5).

1.3. **Lagrange spectra and Diophantine approximation in Fuchsian groups.** The definition of Lagrange spectrum for a Fuchsian group G in terms of Diophantine approximation on G is due to Lehner [23], inspired by Ford's geometric proof of Hurwitz theorem [11] and was studied among others by Haas, Series, Vulakh [14, 15, 42, 43, 44]. In analogy with the classical Lagrange spectrum, we now define these Lagrange spectra first from the Diophantine approximation point of view, then interpret them geometrically in terms of essential heights of geodesics and, finally, in a more dynamical way.

We denote with $\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$ the upper half-plane with the hyperbolic metric. The group of isometries of $\mathbb{H}$ can be identified with $\mathrm{PSL}(2, \mathbb{R})$ (see the beginning of Section 2). Discrete subgroups of $\mathrm{PSL}(2, \mathbb{R})$ are called *Fuchsian groups*. Since Fuchsian groups act by isometries on $\mathbb{H}$ the quotient of the hyperbolic plane by any such group inherits a natural metric from the hyperbolic metric on $\mathbb{H}$. Thus, $X := G\backslash\mathbb{H}$ is a hyperbolic surface (possibly with orbifold singularities coming from fixed points of elliptic elements in G). A Fuchsian group G is a *lattice* if the quotient $X = G\backslash\mathbb{H}$ has finite volume, with respect to the natural volume form induced by the metric. We consider only Fuchsian groups that are so-called *non uniform* lattices, meaning that the quotient has finite volume but is not compact.

The action of $\mathrm{PSL}(2, \mathbb{R})$ extends by continuity to an action on the boundary $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ of $\mathbb{H}$. We recall that an element of $\mathrm{PSL}(2, \mathbb{R})$ is parabolic if it has trace equal to 2. The set of *cusps* of G is the set of points of $\overline{\mathbb{R}}$ fixed by a non trivial parabolic element of G. If $X = G\backslash\mathbb{H}$ has finite volume, then the set of cusps of X coincides with the set of *ends* of the surface itself. We are going to assume for now that ∞ is a parabolic fixed point for G (we later remove this assumption, see Remark 1.5 and Corollary 1.6). We call an element of $\overline{\mathbb{R}}$ *G-rational* if it is the fixed point under some non trivial parabolic transformation in G. The complement of G-rational numbers in $\overline{\mathbb{R}}$ is the set of *G-irrational numbers*.

Diophantine approximation on a Fuchsian group G consists in approximating G-irrational numbers by G-rational ones, or G-rational ones in the G-orbit of a fixed cusp. The definition of Lagrange spectrum $\mathcal{L}(G, \infty)$ in terms of Diophantine approximation on G introduced by Lehner [23] is the following. Given $g \in G$, we denote by $a(g)$ and by $c(g)$ the first entry on the first and second row of g respectively. For $\alpha \in \mathbb{R}$ define $L_G(\alpha)$ to be:

$$L_G(\alpha) := \sup\left\{k : |\alpha - g \cdot \infty| = \left|\alpha - \frac{a(g)}{c(g)}\right| < \frac{1}{kc(g)^2} \text{ for infinitely many } g \in G \right.$$
$$\left. \text{s. t. } g \cdot \infty \text{ are all distinct}\right\}.$$

Then, we define $\mathcal{L}(G, \infty) := \{L_G(\alpha), \alpha \in \mathbb{R}\}$. We remark that if we take $G = \mathrm{PSL}(2, \mathbb{Z})$ in the previous definitions, $L_G(\alpha)$ coincides with the one given in (1.1) for $L(\alpha)$, so that $\mathcal{L}(G, \infty)$ is indeed a generalization of the classical Lagrange spectrum $\mathcal{L}$.

We can interpret this definition geometrically as follows. We recall that a geodesic in the hyperbolic plane is uniquely determined by its two extremal points in $\overline{\mathbb{R}}$. Given two points x and y in $\overline{\mathbb{R}}$, throughout the paper we denote by $\gamma(x,y)$ the hyperbolic geodesic connecting x to y which has x (resp. y) as a backward (resp. forward) end point, i.e. if $\gamma(t)$ is the geodesic parametrization of γ,

$$\gamma(-\infty) := \lim_{t\to-\infty} \gamma(t) = x, \qquad \gamma(+\infty) := \lim_{t\to+\infty} \gamma(t) = y.$$

We define the *naive height* of the geodesic $\gamma = \gamma(x,y)$ by

$$\text{ht}(\gamma) = \begin{cases} \frac{1}{2}|x-y|, & \text{if } x,y \in \mathbb{R}, \\ \infty, & \text{otherwise.} \end{cases} \tag{1.3}$$

Thus, the naive height $\text{ht}(\gamma)$ is the Euclidean radius of the semi-circle which represents the geodesic γ in the upper half plane $\mathbb{H}$.

We say that two elements g and h in G are *equivalent modulo infinity* if there is an element $k \in G$ that fixes infinity and such that $g = kh$. We remark that if g and h are equivalent modulo infinity they differ by a horizontal translation and hence, for every geodesic γ in $\mathbb{H}$, $\text{ht}(g(\gamma)) = \text{ht}(h(\gamma))$. Choose a set G_∞ of representatives of the equivalence classes of G modulo infinity. The *essential height* of a geodesic γ on X is defined by

$$\text{ht}_G(\gamma) = \sup\{k : \text{ht}(g(\tilde{\gamma})) \geq k \quad \text{for infinitely many } g \in G_\infty\}, \tag{1.4}$$

where $\tilde{\gamma}$ is any lift of the geodesic γ from X to the universal cover $\mathbb{H}$.

The following Lemma provides a geometric interpretation of the constant $L_G(\alpha)$ in terms of essential height. We include below also its short proof, which can be found e.g. in [15], since it provides an educational example, for the non familiar reader, of the interplay between Diophantine approximation and penetration in the cusps.

Lemma 1.2 ([15]). *Let G be a non uniform lattice in* $\mathrm{PSL}(2,\mathbb{R})$*. For every real number α we have*

$$L_G(\alpha) = 2\,\text{ht}_G(\gamma(\infty,\alpha)),$$

where $\gamma(\infty,\alpha)$ is the vertical geodesic from ∞ to α.

Proof. Assume that $k > 0$ is such that there exists a sequence g_i of infinitely many elements of G such that

$$|\alpha - g_i \cdot \infty| < \frac{1}{kc(g_i)^2},$$

and the points $g_i \cdot \infty$ are all distinct. The vertical hyperbolic geodesic $\gamma(\infty,\alpha)$ intersects each of the Euclidean disks D_i of radius $1/kc(g_i)^2$ tangent to $\mathbb{R}$ at the points $g_i \cdot \infty$. Equivalently $g_i^{-1}(\gamma(\infty,\alpha)) \cap g_i^{-1}(D_i) \neq \varnothing$. Since the $g_i \cdot \infty$ are all distinct, we have that the elements g_i^{-1} are not equivalent modulo infinity. A simple computation shows that $g_i^{-1}(D_i) = \{z \in \mathbb{H} : \operatorname{Im} z \geq k/2\}$. Thus, the assumption that $\gamma(\infty,\alpha)$ crosses D_i implies that k is such that $\text{ht}(\gamma) \geq k/2$ for infinitely many elements of G_∞. Thus, k belongs to the set of which $L_G(\alpha)$ is supremum if and only if $k/2$ belongs to the set of which $\text{ht}_{G_\infty}(\gamma)$ is the supremum. This gives the desired equality. □

One can define the Lagrange spectrum $\mathcal{L}(X,\infty)$ of the hyperbolic surface $X = G\backslash\mathbb{H}$ with respect to the cusp at ∞ to be

$$\mathcal{L}(X,\infty) := \{2\,\text{ht}_G(\gamma), \gamma \text{ geodesic on } X = G\backslash\mathbb{H}\}.$$

The reason for the constant 2 appearing in the definition is apparent from Lemma 1.2, since one can use it to show that these definitions coincide if G is the uniformizing Fuchsian group of X.

Corollary 1.3. *If $X = G\backslash\mathbb{H}$ then we have that $\mathcal{L}(X,\infty) = \mathcal{L}(G,\infty)$.*

Proof. Lemma 1.2 directly gives the inclusion $\mathcal{L}(G,\infty) \subset \mathcal{L}(X,\infty)$. Conversely, if $\gamma = \gamma(\alpha^-,\alpha^+)$, consider the two vertical geodesics $\gamma^- := \gamma(\infty,\alpha^-)$ and $\gamma^+ := \gamma(\infty,\alpha^+)$. Suppose, with loss of generality, that $L_G(\alpha^+) > L_G(\alpha^-)$. Let us show that this implies that $\text{ht}_G(\gamma) = \text{ht}_G(\gamma^+)$. Let $\text{ht}_G(\gamma) = h$. This implies, by (1.4), that, for every $\varepsilon > 0$, there exists a sequence g_i of infinitely many elements of G such that $g_i(\gamma) \cap \mathcal{U}_{h-\varepsilon} \neq \varnothing$, where $\mathcal{U}_l = \{z \in \mathbb{H} : z = x+iy, y > r\}$. Equivalently, $\gamma \cap g_i^{-1}(\mathcal{U}_{h-\varepsilon}) \neq \varnothing$. It is enough now to observe that this can only happen to a portion of the geodesic γ bounded away from the past endpoint α^- for, otherwise, we would have that

$$|\alpha^- - g_i^{-1} \cdot \infty| < \frac{1}{2(h-\varepsilon)c(g_i^{-1})^2},$$

that is $L_G(\alpha^-) \geq 2h = L_G(\alpha^+)$, which is a contradiction. □

Finally, one can also interpret $\mathcal{L}(G,\infty)$ in a more *dynamical* way, analogously to what happens in the classical case (see (1.2) and Remark 1.1). Given any parametrization $t \mapsto \gamma(t)$ of γ (in particular the one given by the geodesic flow) one immediately sees that $\mathrm{ht}(\gamma) = \sup_{t\in\mathbb{R}} \mathrm{Im}\,\gamma(t)$. If $\mathrm{ht}_G(\gamma)$ is sufficiently large (greater than the starting point of the maximal Margulis neighborhood, see (1.7)), one can see that one equivalently has

$$\mathrm{ht}_G(\gamma) = \limsup_{|t|\to\infty} \mathrm{height}(\gamma(t)), \tag{1.5}$$

where, as before, $\mathrm{height}(x)$ is the imaginary part of the unique lift of a point from x to a chosen fundamental domain of X which has two vertical lines. This equivalence can be seen as a byproduct of the proof of Perron's formula for the essential height, see Lemma 3.4 for details.

Thus, the Lagrange spectrum $\mathcal{L}(X,\infty)$ describes asymptotic depths of penetration of the geodesics of X into the cusp $e=\infty$. This is the point of view that we will generalize in Section 1.5, where we consider more general ways of measuring the penetration into a cusp.

1.4. **Hall rays for Diophantine approximation in Fuchsian groups.** The first result we prove in this paper is the following generalization to Fuchsian groups of Hall's theorem on the existence of a Hall ray for the classical Lagrange spectrum $\mathcal{L}$, proved 1947 for the classical spectrum.

Theorem 1.4 (Hall ray for Fuchsian groups)**.** *Let $G \subset \mathrm{PSL}(2,\mathbb{R})$ be a non uniform lattice. Assume that ∞ is a cusp of G. The Lagrange spectrum $\mathcal{L}(G,\infty)$ of G with respect to ∞ contains a* Hall ray*, i.e. there exists an $L_0 = L_0(G,\infty) \in \mathbb{R}$ such that*

$$[L_0, +\infty] \subset \mathcal{L}(G,\infty).$$

The result extends also to other cusps of G as follows. Let us first remark that the presence of Hall rays does not depend on the choice of normalizations for the width of the cusp at ∞.

Remark 1.5. If $G' = \bar{g}\,G\,\bar{g}^{-1}$ is obtained by conjugating G by an element of $\bar{g} \in \mathrm{PSL}(2,\mathbb{R})$ which fixes infinity, $\mathcal{L}(G,\infty)$ and $\mathcal{L}(G',\infty)$ are obtained by each other by a smooth change of coordinates. In particular, $\mathcal{L}(G,\infty)$ contains a Hall ray if and only if $\mathcal{L}(G',\infty)$ does. More precisely, if $\bar{g} = \left(\begin{smallmatrix} \lambda & \nu \\ 0 & 1/\lambda \end{smallmatrix}\right)$, then one has that the entry $c(\bar{g}g\bar{g}^{-1}) = c(g)/\lambda^2$ for every $g \in G$. Thus, using the explicit form of $\bar{g}$, we have

$$\begin{aligned} L_{G'}(\bar{g}\cdot\alpha) &= \sup\left\{k : \left|\bar{g}\cdot\alpha - \bar{g}g\bar{g}^{-1}\cdot\infty\right| < \frac{1}{kc(\bar{g}g\bar{g}^{-1})^2} \text{ for infinitely many } g\in G\right\} \\ &= \sup\left\{k : \lambda^2\left|\alpha - g\cdot\infty\right| < \frac{\lambda^4}{kc(g)^2} \text{ for infinitely many } g\in G\right\} \\ &= \frac{1}{\lambda^2}L_G(\alpha). \end{aligned}$$

The Lagrange spectrum $\mathcal{L}(G,e)$ with respect to a different cusp e can be obtained by conjugating by an appropriate element of $\mathrm{PSL}(2,\mathbb{R})$ sending e to ∞, once a normalization has been chosen (see for example Section 1.5 or Section 3.1 for a natural one). By Remark 1.5 the presence of a Hall ray does not depend on the actual choice of the normalization. We have the following immediate corollary of Theorem 1.4.

Corollary 1.6. *Let $G \subset \mathrm{PSL}(2,\mathbb{R})$ be a non uniform lattice. For any e be a cusp of G, the Lagrange spectrum $\mathcal{L}(G,e)$ of G with respect to e contains a* Hall ray*.*

This results should be compared with the existence of Hall rays proved by Schmidt and Sheingorn in [35] for the Markoff spectrum in an analogous setup. In general, it is easier to construct values in the Markoff spectrum than in the Lagrange spectrum, essentially because of the presence of a supremum instead than a lim sup in Equation (1.1). In order to show that a certain value is in the Markoff spectrum is achieved, Schmidt and Sheingorn construct a geodesic which starts achieving a (sufficiently high) desired value of the height function. Then, to guarantee that this value is indeed the supremum, they use a symbolic coding (which essentially counts winding numbers in the cusp at ∞) and *slide* the endpoints of the geodesic to guarantee that further excursions in the cusp are of lower height. On the other hand, for the Lagrange spectrum, one needs to construct a *sequence* of increasing times for which the height tends to the desired value. To achieve this much more delicate form of control of the geodesic behavior, we also use symbolic coding (in the form of the boundary expansions first described by Bowen and Series) but then need to adapt to the Fuchsian setting Hall's original ideas in particular by reducing the result to the study of a sums of Cantor sets on the boundary. See Section 1.7 for more details on the strategy of proof.

1.5. **Hall rays for dynamical Lagrange spectra of Riemann surfaces.** The Lagrange spectra $\mathcal{L}(X,\infty)$, defined in terms of Diophantine approximation in Fuchsian groups, can be interpreted, as we saw in Section 1.3, as penetration spectra for geodesics at the cusp at ∞ with respect to the height function. From this point of view, it is natural to consider simultaneous penetration in other cusps and, more generally, different notions of *penetration*. Simultaneous penetration in the cusps can be defined with respect to any *proper* function from the surface to $\mathbb{R}^+$ (see below). The main results stated in this section (Theorem 1.7 and Theorem 1.8) concern these more general Lagrange penetration spectra and shows that Hall rays defined with respect to height functions are *stable*, i.e. persistent under (Lipschitz) perturbation, in a sense which is made precise in Section 1.6.

We consider in this section any Riemann surface X with genus g and n punctures, such that $\chi(X) := 2-2g-n < 0$. We adopt in this paper the convention (used for example by Beardon [3]) to call Riemann surfaces also two dimensional hyperbolic orbifolds (the modular *surface* is such an example since it has two orbifold singularities). These, also called *marked* or singular Riemann surfaces, are all finite quotients of smooth Riemann surfaces. The Uniformization theorem gives that $X = G\backslash\mathbb{H}$ is the quotient of $\mathbb{H}$ under some Fuchsian group G acting by the left action given by Möbius transformations. Orbifold singularities of X correspond to elliptic elements if G, so X is a smooth Riemann surface iff G contains no elliptic elements.

Let $h\colon \mathbb{H} \to \mathbb{R}_+$ be a G-invariant continuous function. Equivalently, h induces a function on the quotient $X = G\backslash\mathbb{H}$, which we will still denote by h. We assume that h is *proper*, meaning a function such that the preimage of a compact set is a compact sets. In particular, h diverges in the cusps of X.

One can define a generalization of the Lagrange spectrum by measuring the asymptotic excursion into the cusps with respect to the function $h\colon X = G\backslash\mathbb{H} \to \mathbb{R}_+$ as follows. Let $t \mapsto \gamma(t)$ be a geodesic on X and let

$$L_G(h,\gamma) = L(h,\gamma) := \limsup_{t\to+\infty} h(\gamma(t)). \tag{1.6}$$

We will often drop the explicit dependence on G since it is implicit in the symmetries of the function h. Then we call $\mathcal{L}(X,h)$ the corresponding Lagrange spectrum, given by the set of values $L(h,\gamma)$ for γ geodesics on X. These type of Lagrange spectra are also called *dynamical Lagrange spectra* in the literature. Dynamical spectra were in particular studied in the seminal works by [27, 32, 17] and have seen a recent surge of interest, see for example [1, 5, 7, 10, 19, 20, 26].

If the surface X has only one cusp at infinity, height$(\cdot)$ is an example of a *proper* function on X. Proper functions when there are more cusps can be build for example by measuring penetration in each cusp with respect to a height function in that cusp and either adding them up or taking the maximum of these functions.

A natural example of proper function to consider is given by Paulin and Parkkonen in [31]. For each cusp e, let $\beta_{X,e}$ be the Busemann function on X for the end e, normalized to converge to $+\infty$ towards e and to vanish on the boundary of the maximal open Margulis neighborhood (definitions can be found in [31]). We remark that $\beta_{X,\infty} = 2\log \mathrm{ht}_G$ (see [17]), so in particular these two penetrations have the same Hall rays. The *Busemann height function* β_X is defined to be the maximum of the functions $\beta_{X,e}$ over the cusps e of X.

The Lagrange spectrum $\mathcal{L}(X)$ of the Riemann surface X is then defined by Paulin and Parkkonen to be the dynamical spectrum $\mathcal{L}(X,\beta_X)$ with respect to the Busemann height function β_X. Let us first highlight, for its intrinsic interest, a result which will follow as a special case of the more general Theorem 1.8 that we will state in Section 1.6.

Theorem 1.7 (Hall ray for Riemann surfaces). *For any X non compact, finite volume Riemann surface with $\chi(X) < 0$, the Lagrange spectrum $\mathcal{L}(X) := \mathcal{L}(X,\beta_X)$ contains a* Hall ray.

As evidenced by the brief history in the previous section, the existence of a Hall ray was known for dynamical Markoff and Lagrange spectra dimension greater than 3 [32] and for Markoff spectra of Riemann surfaces [35], as well as for Lagrange spectra in other dynamical contexts, such as [20, 1]. Thus, our work deals with the only case that was surprisingly still open in the constant curvature case, namely Lagrange spectra in dimension 2. It is worth to remark that the existence of Hall rays is actually not expected to hold in general in dimension 2 with variable negative curvature, see [32].

1.6. **Persistence of Hall rays for Lagrange spectra of Riemann surfaces.** We state now the most general result we prove in this paper (of which Theorem 1.7 is a corollary), which shows in particular that the Hall ray for $\mathcal{L}(X,\beta_X)$ is *stable* under Lipschitz perturbations. We consider proper functions h which behave in at least one cusp as a Lipschitz perturbation of the essential height function in that cusp, in the following sense.

Recall that a horodisk at infinity is a set of the form $\mathcal{U}_l = \{z \in \mathbb{H} : z = x+iy, y > l\}$ for some $l > 0$, such that its image on the surface $X = G\backslash\mathbb{H}$ is a *Margulis neighborhood*, i.e. is homeomorphic to a punctured disk. The *fundamental horodisk* at infinity is

$$\mathcal{U}_m, \quad \text{where } m \text{ is the } \textit{minimal } l > 0 \text{ s.t. } \mathcal{U}_l \text{ is a Margulis neighbourhood.} \tag{1.7}$$

The projection of $\mathcal{U}_m$ on X is called the *maximal Margulis neighborhood* of the cusp at ∞. We will call m the *height* of the maximal Margulis neighborhood.

Let U be an open subset in $\mathbb{H}$ and let $g\colon U \to \mathbb{R}$ be a continuous function bounded on U. Recall that the uniform norm of g is

$$\|g\|_\infty := \sup_{z\in U} g(z).$$

Recall also that when g is a *Lipschitz function*, the *Lipschitz constant* of g is

$$\mathrm{Lip}(g) := \sup_{z,w\in U} \frac{|g(z)-g(w)|}{|z-w|}. \tag{1.8}$$

Finally, if $g\colon U \to \mathbb{R}$ is a bounded Lipschitz function we define its *Lipschitz norm* as

$$\|g\|_{\mathrm{Lip}} := \|g\|_\infty + \mathrm{Lip}(g).$$

The following result shows that the presence of a Hall ray is open under perturbations in the Lipschitz norm.

Theorem 1.8 (Hall ray for perturbations). *Let $G \subset \mathrm{PSL}(2,\mathbb{R})$ be a non uniform lattice. Assume that ∞ is a cusp of G. Let $h\colon \mathbb{H} \to \mathbb{R}_+$ be a G-invariant continuous function such that the induced function on $X = G\backslash\mathbb{H}$ is proper.*

There exists a constant $\delta_G > 0$ such that, if there exists an $l_0 > 0$ so that

$$\left\|\big(h - \mathrm{Im}(\cdot)\big)|_{\mathcal{U}_{l_0}}\right\|_{Lip} < \delta_G, \tag{1.9}$$

then the Lagrange spectrum $\mathcal{L}(X,h)$ contains a Hall ray.

Remark 1.9. More precisely, we show that if G is normalized so that the height of the maximal Margulis neighborhood is equal to 1, then $\mathcal{L}(X,h)$ contains a *Hall ray* as long as, for some $l_0 > 0$, we have

$$\left\|\big(h - \mathrm{Im}(\cdot)\big)|_{\mathcal{U}_{l_0}}\right\|_{\mathrm{Lip}} < \frac{1}{4\sqrt{2}}. \tag{1.10}$$

Remark 1.10. As in the case of Theorem 1.4 (see Corollary 1.6)), the assumption that ∞ is a cusp is not a real restriction. Given a Riemann surface X and a proper function $\overline{h}\colon X \to \mathbb{R}_+$, the spectrum $\mathcal{L}(X,\overline{h})$ contains a Hall ray as as long as there exists a cusp e of X and a uniformization $X = G\backslash\mathbb{H}$ such that e lifts to ∞ and $m = 1$, and a lift $h\colon \mathbb{H} \to \mathbb{R}^+$ of $\overline{h}$ for which (1.10) is satisfied.

1.7. **Some ideas in the proofs.** Let us now give some details on the way in which we prove the main results that we stated in the two previous sections. Let us first recall the classical strategy introduce by Hall in [16]. The starting point of Hall's approach is a classical formula, due to Perron [33], that allows to compute the Lagrange value of a real number α given its continued fraction expansion. If $\alpha = [a_0; a_1, a_2 \dots]$, then Perron's formula for its Lagrange value is the following expression:

$$L(\alpha) = \limsup_{n\to\infty} [0; a_{n-1}, a_{n-2}, \dots, a_0] + a_n + [0; a_{n+1}, a_{n+2}, \dots]. \tag{1.11}$$

The expression inside the lim sup consist of *central digit*, a_n, and two *tails* given by continued fraction expansions. It is well known that the set $\mathbb{K}_N$ of numbers in $[0,1]$ whose continued fractions expansion digits are all bounded by an integer N form a Cantor set. At the heart of Hall's work, there is a statement about these Cantor sets: he proves that if $N \geq 4$ the Cantor set $\mathbb{K}_N$ is sufficiently *thick* so that the sum set $\mathbb{K}_N + \mathbb{K}_N$ contains an interval. Hall's idea is then to construct real numbers α that realize any sufficiently large value L in the Lagrange spectrum by their continued fraction expansion, using larger and larger blocks formed by a large central digit a_n set to be the integer part of L, and two carefully selected finite tails (with bounded digits), which converge to two elements in the Cantor sets which add up to the fractional part of L. Evaluating Perron's formula on this special sequence then yields the desired Lagrange value.

The starting point for our work is that, since the naive height of a geodesic is (half) the difference of its end points on the real line, one can carry a strategy similar to Hall's one, with the endpoints playing the role of the *tails*. Of course there are several difficulties to overcome. The first tool we introduce to carry our this strategy is a nice symbolic coding, which we use to code geodesics and to provide a geometric substitute for classical continued fractions expansions. We use Bowen-Series boundary expansions with respect to a finite index subgroup $\Gamma < G$ without elliptic points. The idea of passing to the finite index subgroup Γ is a key, since it allows to use the simplest and neater form of the coding (which in general is otherwise defined only under some assumptions on the fundamental domain of the Fuchsian group): even though this reduction possibly looses information about the *bottom* part of the spectrum, we show that it still allows to study the *top* part of the spectrum (and in particular the existence of Hall's rays).

To study Lagrange spectra with respect to a proper function in presence of several cusps, it is key that through this coding one can *see* (large) excursions into *each* cusp. To control these excursions, we use *decomposition into cuspidal words*, a notion which was introduced in our previous work [1]. A delicate point we show is that by (locally) bounding the lengths of cuspidal words one is able to estimate the penetration into *all* cusps (see Lemma 5.1), in order to then be able to achieve Lagrange values by prescribing larger excursions in a given chosen cusp.

For the general setting considered in Theorem 1.8, the final key tool is a generalization of Hall classical result on sums of Cantor sets, which gives a sufficient condition for such a sum to contain an interval. We prove what we call a *stable* version of Hall's result, where stability is meant here under (bounded size) perturbations, with respect to the Lipschitz norm, of the sum function. We give more details and a precise formulation of this result in the following Section 1.8.

Finally, let us point out that Theorem 1.4 is *morally* a special case of Theorem 1.8. We say morally since one cannot formally deduce Theorem 1.4 from Theorem 1.8. In fact the function height$(\cdot)$ is not proper if X has more than one cusp and, moreover, $L_G(\cdot)$ can be expressed in terms of the essential height of a geodesic as a $\limsup$ as $|t| \to \infty$ (see (1.5)) and not as $t \to \infty$. However both these points can be easily taken into account via simple technical tricks (i.e. artificially creating a proper function with the same spectrum and using symmetric geodesics, as in Equation (1.5) for the modular surface). We chose to present in Section 3 an independent proof of Theorem 1.4 for two reasons: first of all, the arguments required to prove Theorem 1.4 are much more direct and essentially exploit coding and geometric arguments, combined with a generalization to the Fuchsian context of Hall's original strategy. Secondly, we believe that the proof of Theorem 1.4 might serve as a gentle guide for the reader to the ideas exploited in the rather more technical Theorem 1.8. Indeed, while the main strategy is the same, additional layers of technical difficulty in the proof of Theorem 1.8 come from the need to simultaneously control excursions in all cusps and to generalize Hall's result on the sum of Cantor sets in order to be able to deal with Lipschitz perturbations of the height function. We also remark that the full strength of the symbolic coding we use, in particular of the decomposition into cuspidal words (see Section 2.3), is only used for Theorem 1.8. For Theorem 1.4 it would in principle be enough to control, through the coding, only the excursions into the cusp at ∞. We instead use the same Cantor set ($\mathbb{B}_N \subset \partial\mathbb{D}$ and its image $\mathbb{K}_N \subset \mathbb{R} = \partial\mathbb{H}$, see Section 4) for both the proof of Theorem 1.4 and the one of Theorem 1.8, in order to prove only once the distortion and gaps estimates needed to apply results on the sum or the perturbation of the sum of Cantor sets.

1.8. **A stable version of Hall's theorem on the sum of Cantor sets.** We conclude this section by formulating the generalization of Hall's theorem on the sum of Cantor sets on which our result on perturbed Lagrange spectra, namely Theorem 1.8, is based. This statement is an abstract result on Cantor sets, which might be of independent interest, given the large literature on these type of questions (for example [30, 2, 29]). In order to formulate the result, we need first to give a series of definitions on the way a Cantor set is constructed and the properties of its *holes*.

Let $\mathbb{K}$ be any Cantor set in $\mathbb{R}$. One can present $\mathbb{K}$ as intersection $\bigcap_{n\in\mathbb{N}} \mathbb{K}(n)$ of unions $\mathbb{K}(n)$ of closed disjoint intervals. We now define the notion of *slow subdivision*: intuitively, the reader should keep in mind that this definition convey that the Cantor set is build step by step by removing exactly one *hole* at each stage. We adopt the following notation. For any K be compact interval and any B open interval with $B \subset K$ (where the inclusion is obviously strict), we denote by K^L and K^R the two closed subintervals of

K such that

$$K = K^L \sqcup B \sqcup K^R.$$

As suggested by the notation, we assume that K^L is on the left side of B and K^R is on the right side of the *hole* B.

A *slow subdivision* of $\mathbb{K}$ is a family of closed sets $\big(\mathbb{K}(n)\big)_{n\in\mathbb{N}}$ with $\mathbb{K}(n+1) \subset \mathbb{K}(n)$ for any $n \in \mathbb{N}$ which satisfies the following properties.

(1) Any set $\mathbb{K}(n)$ is the union of $n+1$ disjoint closed intervals, where in particular we have

$$\mathbb{K}(0) = [\min \mathbb{K}, \max \mathbb{K}].$$

(2) For any n there is exactly one compact interval K in $\mathbb{K}(n)$ and a non-empty open subinterval B_K of K such that

$$K \cap \mathbb{K}(n+1) = K \setminus B_K = K^L \sqcup K^R,$$

where K^L and K^R are two disjoint, non-empty, closed subintervals in $\mathbb{K}(n+1)$.

(3) We have

$$\bigcap_{n\in\mathbb{N}} \mathbb{K}(n) = \mathbb{K}.$$

The *holes* of a Cantor set $\mathbb{K}$ are the connected components of its complement which are contained in the interval $[\min \mathbb{K}, \max \mathbb{K}]$. We remark that holes are maximal open intervals in the complement.

Remark 1.11. In a slow subdivision the holes are naturally ordered: for any $n \in \mathbb{N}$, the n^{th} *hole* is the unique $B_K \subset K$, given by condition (2), which is removed from the connected component K of $\mathbb{K}(n)$ at stage $n+1$, i.e. such that $K \cap \mathbb{K}(n+1) = K \setminus B_K$.

We say that a slow subdivision $\big(\mathbb{K}(n)\big)_{n\in\mathbb{N}}$ of the Cantor set $\mathbb{K}$ satisfies the ε-*stable gap condition* for some $\varepsilon > 0$ if for any n, the n^{th} hole B_K (according to the terminology introduced in Remark 1.11) and its right and left closed intervals K^L and K^R satisfy

$$\frac{|B_K|}{|K^L|} < 1-\varepsilon \quad \text{and} \quad \frac{|B_K|}{|K^R|} < 1-\varepsilon. \tag{1.12}$$

We say that the Cantor set $\mathbb{K}$ satisfies the ε-*stable gap condition* if it admits a slow subdivision which satisfies the ε-stable gap condition.

Given two Cantor sets $\mathbb{K}$ and $\mathbb{F}$, we say that the pair of Cantor sets $(\mathbb{K}, \mathbb{F})$ satisfies the ε-*size condition* if we have

$$|B| \leq (1-\varepsilon)|\mathbb{K}| \quad \text{and} \quad |C| \leq (1-\varepsilon)|\mathbb{F}|, \tag{1.13}$$

where $|\mathbb{K}| := \max \mathbb{K} - \min \mathbb{K}$ is the length of $\mathbb{K}$, $|\mathbb{F}| := \max \mathbb{F} - \min \mathbb{F}$ is the length of $\mathbb{F}$, and B and C are any pair of holes in $\mathbb{K}$ and in $\mathbb{F}$ respectively.

We can now state our stable version of Hall's Theorem. Consider the *sum* function $S_0 \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $S_0(x_1, x_2) := x_1 + x_2$. Fix an open subset $U \subset \mathbb{R}^2$; abusing the notation we still denote S_0 the restriction of S_0 to U.

Theorem 1.12 (Stable Hall Theorem)**.** *Fix $\varepsilon > 0$ and let $\mathbb{K}$ and $\mathbb{F}$ be two Cantor sets in $\mathbb{R}$, each one satisfying the ε-stable gap condition and such that $\mathbb{K} \times \mathbb{F} \subset U$. Assume that the pair $(\mathbb{K}, \mathbb{F})$ satisfies the ε-size condition. Then for any function $S \colon U \to \mathbb{R}$ such that*

$$\frac{1 - \mathrm{Lip}(S - S_0)}{1 + \mathrm{Lip}(S - S_0)} > 1 - \varepsilon \tag{1.14}$$

we have

$$S(\mathbb{K} \times \mathbb{F}) = S\left([\min \mathbb{K}, \max \mathbb{K}] \times [\min \mathbb{F}, \max \mathbb{F}]\right).$$

Remark 1.13. Let us show that this is indeed a generalization of the classical Hall's theorem. Observe that S as in the statement is automatically continuous, indeed it is the sum of S_0 and a Lipschitz function. Thus, if K and F are closed intervals, then $S(K \times F)$ is a closed interval too, by continuity of S. It hence follows that, in the special case when $S = S_0$ (and $U = \mathbb{R}^2$) is the sum function, the theorem shows that $\mathbb{K} + \mathbb{F}$ contains an interval, which is the conclusion of Hall's theorem. The assumptions of Hall, on the other hand, correspond to the ε-stable and ε-size condition for the limit case $\varepsilon = 0$.

Structure of the paper. The rest of the paper is organized as follows. In Section 2 we describe the symbolic coding of geodesics that we use. We first recall the simplest case of Bowen-Series coding and the notion of boundary expansions (see Section 2). We also introduce the notions of *cuspidal words* and decomposition into cuspidal words. In Section 3 we present the proof of Theorem 1.4, in particular introducing Hall's arguments. The starting point is a generalization of Perron's formula for the classical spectrum through symbolic coding, see Lemma 3.4 in Section 3.3. The only result needed whose proof is given later is Proposition 3.2 on the difference of Cantor sets.

In Section 4 we describe the Cantor sets $\mathbb{B}_N \subset \mathbb{D}$ consisting of endpoints of geodesics such that the lengths of cuspidal words is bounded by N. We then prove some distortions estimates on Möbius transformations (see Section 4.2) which are then applied to show that the image $\mathbb{K}_N \subset \mathbb{R}$ of the Cantor sets $\mathbb{B}_N$ satisfy (and their rigid images) satisfy the assumptions of the Stable Hall theorem (see Lemma 4.4). In Section 4.4 we can then prove Proposition 3.2 thus completing the proof of Theorem 1.4.

In the following Section 5 we show that by locally bounding the lengths of cuspidal words one can control the distance from a compact core of X (see in particular Lemma 5.1). Theorem 1.8 is then proved using these preliminary results and the Stable Hall Theorem 1.12 in Section 6. In particular, the key step to implement Hall's strategy is this context is provided by Proposition 6.7, proved in Section 6.3.

In Section 7 we then give the proof of the Stable Hall Theorem 1.12. Finally, two Appendices contain respectively the proofs of some Lemmas on parabolic words (Appendix A) and some estimates which relate the Lipschitz norm of h to the Lipschitz norm of the auxiliary function H introduced in Section 6.1, proved in Appendix B.

2. Symbolic coding

Let $\mathbb{N} = \{0, 1, \dots\}$ denote the natural numbers. We will use $\mathbb{H}$ and the unit disk $\mathbb{D} = \{z \in \mathbb{C}, |z| < 1\}$ interchangeably, by using the identification $\mathscr{C} \colon \mathbb{H} \to \mathbb{D}$ (here $\mathscr{C}$ stays for *Cayley* map) given by

$$\mathscr{C}(z) = \frac{z-i}{z+i}, \qquad z \in \mathbb{H}. \tag{2.1}$$

Let $\mathrm{SL}(2,\mathbb{R})$ be the set of 2×2 matrices with real entries and determinant one, and similarly for $\mathrm{SL}(2,\mathbb{C})$. The group $\mathrm{SL}(2,\mathbb{R})$ acts on $\mathbb{H}$ by *Möbius transformations* (or *homographies*). Given $g \in \mathrm{SL}(2,\mathbb{R})$ we will denote by $g \cdot z$ the action of g on $z \in \mathbb{H}$ given by

$$z \mapsto g \cdot z = \frac{az+b}{cz+d}, \qquad \text{if} \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$ induces an action on the unit tangent bundle $T^1\mathbb{H}$, by mapping a unit tangent vector at z to its image under the derivative of g in z, which is a unit tangent vector at $g \cdot z$. This action is transitive but not faithful and its kernel is exactly $\{\pm \mathrm{Id}\}$, where Id is the identity matrix. Thus, it induces an isomorphism between $T^1\mathbb{H}$ and $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm \mathrm{Id}\}$. Throughout the paper, we will often write $A \in \mathrm{PSL}(2,\mathbb{R})$ and denote by $A \in \mathrm{SL}(2,\mathbb{R})$ the equivalence class of the matrix A in $\mathrm{PSL}(2,\mathbb{R})$. Equality between matrices in $\mathrm{PSL}(2,\mathbb{R})$ must be intended as equality as equivalence classes. The group $\mathrm{SL}(2,\mathbb{C})$ also acts by Möbius transformations on the Riemann sphere $\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ and we will denote this action with $g \cdot z$ too.

2.1. Cutting sequences.

For a special class of Fuchsian groups, Bowen and Series developed in [6] a geometric method of symbolic coding of points on $\partial\mathbb{D}$, known as *boundary expansion*, that allows to represent the action of a set of suitably chosen generators of the group as a subshift of finite type. Boundary expansions can be thought of as a geometric generalization of the continued fraction expansion, which is related to the boundary expansion of the geodesic flow on the modular surface (see [36] for this connection). We will now recall two equivalent definitions of the simplest case of boundary expansions, either as cutting sequences of geodesics on $X = \Gamma\backslash\mathbb{H}$ or as itineraries of expanding maps on $\partial\mathbb{D}$. For more details and a more general treatment we refer to the expository introduction to boundary expansions given by Series in [37].

Let $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ be a Fuchsian group and assume in this section that Γ be a co-finite, non cocompact and does not contain elliptic elements[1]. One can see that Γ admits a fundamental domain which is an *ideal polygon* $\mathcal{F}$ in $\mathbb{D}$, that is a hyperbolic polygon having finitely many vertices ξ all lying on $\partial\mathbb{D}$ (see for

[1]The choice of a different name, Γ, for the Fuchsian group in this section is deliberate. To study the Lagrange spectrum of of general co-finite, non cocompact Fuchsian group G, that can a priori contain elliptic elements, we will exploit a subgroup $\Gamma < G$ without elliptic elements, as in this section.

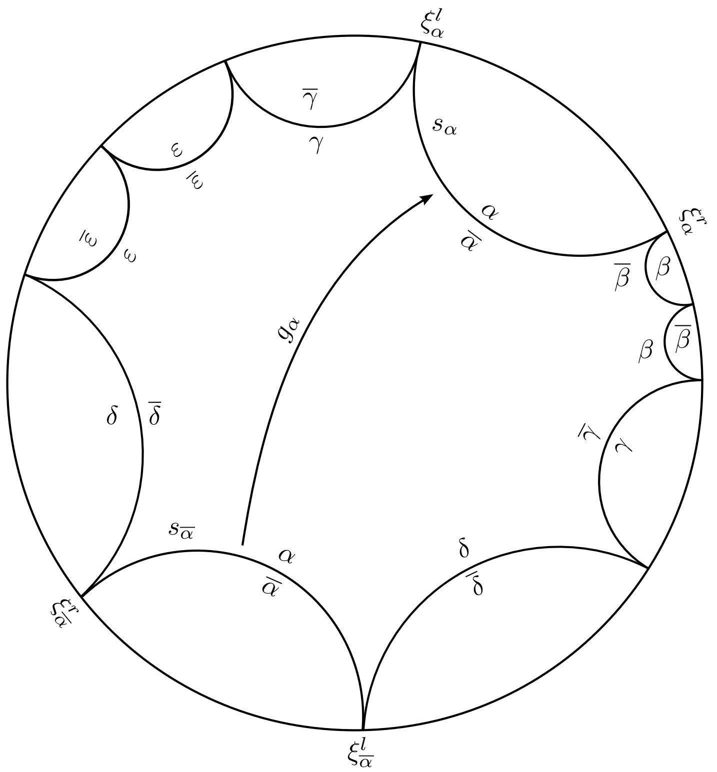

FIGURE 1. A hyperbolic fundamental domain, with sides labeling and the action of the generator g_α.

example Tukia [39]). We will denote by s the sides of $\mathcal{F}$, which are geodesic arcs with endpoints in $\partial\mathbb{D}$. Geodesic sides appear in pairs, i.e. for each s there exists a side $\overline{s}$ and an element g of Γ such that the image $g(s)$ of s by g is $\overline{s}$. Let $2d$ ($d \geq 2$) be the number of sides of of $\mathcal{F}$. Let $\mathscr{A}_0$ be a finite alphabet of cardinality d and label the $2d$-sides ($d \geq 2$) of $\mathcal{F}$ by letters in

$$\mathscr{A} = \mathscr{A}_0 \cup \overline{\mathscr{A}_0} = \{\alpha \in \mathscr{A}_0\} \cup \{\overline{\alpha}, \alpha \in \mathscr{A}_0\}$$

in the following way. Assign to a side s an internal label α and an external one $\overline{\alpha}$. The side $\overline{s}$ paired with s has $\overline{\alpha}$ as internal label and α as the external one. We then see that the pairing given by $g(s) = \overline{s}$ transports coherently the couple of labels of the side s onto the couple of labels of the side $\overline{s}$. Let us denote by s_α the side of $\mathcal{F}$ whose *external* label is α. A convenient set of generators for Γ is given by the family of isometries $g_\alpha \in \mathrm{PSL}(2,\mathbb{R})$ for $\alpha \in \mathscr{A}_0$, where g_α is the isometry which sends the side $s_{\overline{\alpha}}$ onto the side s_α, and their inverses $g_{\overline{\alpha}} := g_\alpha^{-1}$ for $\alpha \in \mathscr{A}_0$, such that $g_\alpha^{-1}(s_\alpha) = s_{\overline{\alpha}}$, see Theorem 3.5.4 in [22]. Thus, $\mathscr{A}$ can be thought as the set of labels of generators, see Figure 1. It is convenient to define an involution on $\mathscr{A}$ which maps $\alpha \mapsto \overline{\alpha}$ and $\overline{\alpha} \mapsto \overline{\overline{\alpha}} = \alpha$.

Since $\mathcal{F}$ is an ideal polygon, Γ is a free group. Hence every element of Γ as a unique representation as a *reduced word* in the generators, i.e. a word in which an element is never followed by its inverse. We transport the internal and external labeling of the sides of $\mathcal{F}$ to all its copies in the tessellation by ideal polygons given by all the images $g(\mathcal{F})$ of $\mathcal{F}$ under $g \in \Gamma$. We label a side s of a copy $g(\mathcal{F})$ of $\mathcal{F}$ with the labels of the side $g^{-1}(s) \in \partial\mathcal{F}$. We remark that this is well defined since we have assigned an internal and an external label to each side of $\mathcal{F}$, and this takes into account the fact that every side of a copy $g(\mathcal{F})$ belongs also to another adjacent copy $g'(\mathcal{F})$.

Let γ be a hyperbolic geodesic ray, starting from the center 0 of the disk and ending at a point $\xi \in \partial\mathbb{D}$. The *cutting sequence* of γ is the infinite reduced word obtained by concatenating the *exterior* labels of the sides of the tessellation crossed by γ, in the order in which they are crossed. In particular, if the cutting sequence of γ is $a_0, a_1, \dots,$ the i^{th} crossing along γ is from the region $g_{a_0} \dots g_{a_{i-1}}(\mathcal{F})$ to $g_{a_0} \dots g_{a_i}(\mathcal{F})$ and the sequence of sides crossed is

$$g_{a_0} g_{a_1} \cdots g_{a_{n-1}}(s_{a_n}), \qquad n \in \mathbb{N}. \tag{2.2}$$

We remark that, since two distinct hyperbolic geodesics meet at most in one point, a word arising from a cutting sequence is reduced. In other words, hyperbolic geodesics *do not backtrack.*

We complete this section explaining how to code *complete* geodesics passing through $\mathcal{F}$ at time zero. If γ is a complete geodesic, parametrized in such a way that $\gamma(0) \in \mathcal{F}$, let $\gamma_\pm(t) \colon \mathbb{R}_+ \to \mathbb{D}$ defined by $\gamma_\pm \colon t \mapsto \gamma(\pm t)$. In other words, γ_+ is the ray obtained moving along γ forward in time and γ_- is the one obtained moving along γ backwards in time. Code the first one by $(b_n)_{n\in\mathbb{N}}$ and the second one by $(c_n)_{n\in\mathbb{N}}$. Then the cutting sequence of γ is the infinite word $(a_n)_{n\in\mathbb{Z}}$ defined by

$$a_n = \begin{cases} b_n, & \text{if } n \geq 0, \\ \overline{c_{n-1}}, & \text{if } n < 0. \end{cases}.$$

The bar for negative n's is due to the fact that we were moving along γ in the reverse orientation when defining the sequence $(c_n)_{n\in\mathbb{N}}$.

2.2. **Boundary expansions.** Let us now explain how to recover cutting sequences of geodesic rays by itineraries of an expanding map on $\partial\mathbb{D}$. The action of each $g \in \Gamma$ extends by continuity to an action on $\partial\mathbb{D}$ which will be denoted by $\xi \mapsto g(\xi)$. Let $\mathcal{A}[\alpha]$ be the closed arc on $\partial\mathbb{D}$ such that $\mathcal{A}[\alpha] \cup s_\alpha$ is the boundary of the connected component which is disjoint from the interior of $\mathcal{F}$. Then it is easy to see from the geometry that the action $g_\alpha \colon \partial\mathbb{D} \to \partial\mathbb{D}$ associated to the generator g_α of Γ sends the complement of $\mathcal{A}[\overline{\alpha}]$ to $\mathcal{A}[\alpha]$. Moreover, if for each $\alpha \in \mathscr{A}$ we denote by ξ^l_α and ξ^r_α the endpoints of the side s_α, with the convention that the right follows the left moving in clockwise sense on $\partial\mathbb{D}$, we have

$$g_\alpha(\xi^r_{\overline{\alpha}}) = \xi^l_\alpha \qquad \text{and} \qquad g_\alpha(\xi^l_{\overline{\alpha}}) = \xi^r_\alpha. \tag{2.3}$$

Some times it will be useful to write $\xi^l_\alpha = \inf \mathcal{A}[\alpha]$ and $\xi^r_\alpha = \sup \mathcal{A}[\alpha]$. Let $\mathcal{A} = \bigcup_\alpha \mathring{\mathcal{A}[\alpha]} \subseteq \partial\mathbb{D}$, where $\mathring{\mathcal{A}[\alpha]}$ denotes the arc $\mathcal{A}[\alpha]$ without endpoints. Define $F \colon \mathcal{A} \to \partial\mathbb{D}$ by

$$F(\xi) = g_\alpha^{-1}(\xi), \qquad \text{if } \xi \in \mathring{\mathcal{A}[\alpha]} .$$

Let us call a point $\xi \in \partial\mathbb{D}$ *cuspidal* if it is a vertex of the ideal tessellation with fundamental domain $\mathcal{F}$ and *non-cuspidal* otherwise. One can see that ξ is non-cuspidal point if and only if $F^n(\xi)$ is defined for any $n \in \mathbb{N}$. One can *code* a trajectory $\{F^n(\xi), n \in \mathbb{N}\}$ of a non-cuspidal point $\xi \in \partial\mathbb{D}$ with its *itinerary* with respect to the partition into arcs $\{\mathcal{A}[\alpha], \alpha \in \mathscr{A}\}$, that is by the sequence $(a_n)_{n\in\mathbb{N}}$, where $a_n \in \mathscr{A}$ are such that $F^n(\xi) \in \mathcal{A}[a_n]$ for any $n \in \mathbb{N}$. We will call such sequence the *boundary expansion* of ξ.

Moreover, in analogy with the continued fraction notation, we will write

$$\xi = [a_0, a_1, \dots]_{\partial\mathbb{D}}.$$

When we write the above equality or say that ξ has boundary expansion $(a_n)_{n\in\mathbb{N}}$ we implicitly assume that ξ is non-cuspidal.

One can show that the only restrictions on letters which can appear in a boundary expansion $(a_n)_{n\in\mathbb{N}}$ is that α cannot be followed by $\overline{\alpha}$, that is

$$a_{n+1} \neq \overline{a_n} \qquad \text{for any } n \in \mathbb{N}. \tag{2.4}$$

We will call this property the *no-backtracking condition.* Boundary expansions can be defined also for cuspidal points (see Remark 4.3 in [1]) but are unique exactly for non-cuspidal points. Every sequence in $\mathscr{A}^\mathbb{N}$ which satisfies the no-backtracking condition can be realized as a boundary expansion (of a cuspidal or non-cuspidal point).

We will adopt the following notation. Given a sequence of letters $a_0, a_1, \dots, a_n$, let us denote by

$$\mathcal{A}[a_0, a_1, \dots, a_n] = \overline{\mathcal{A}[a_0] \cap F^{-1}(\mathcal{A}[a_1]) \cap \dots \cap F^{-n}(\mathcal{A}[a_n])}$$

the closure of set of points on $\partial\mathbb{D}$ whose boundary expansion starts with $a_0, a_1, \dots, a_n$. One can see that $\mathcal{A}[a_0, \dots, a_n]$ is a connected arc on $\partial\mathbb{D}$ which is non-empty exactly when the sequence satisfies the no-backtracking condition (2.4). From the definition of F, one can work out that

$$\mathcal{A}[a_0, a_1, \dots, a_n] = g_{a_0} \cdots g_{a_{n-1}} \mathcal{A}[a_n]. \tag{2.5}$$

Thus two such arcs are *nested* if one word contains the other as a beginning. For any fixed $n \in \mathbb{N}$, the arcs of the form $\mathcal{A}[a_0, a_1, \dots, a_n]$, where $a_0, a_1, \dots, a_n$ vary over all possible sequences of n letters in $\mathscr{A}$ which satisfy the no-backtracking condition, will be called an *arc of level n.* To produce the arcs of level $n+1$, each arc of level n of the form $\mathcal{A}[a_0, a_1, \dots, a_n]$ is partitioned into $2d-1$ arcs, each of which has

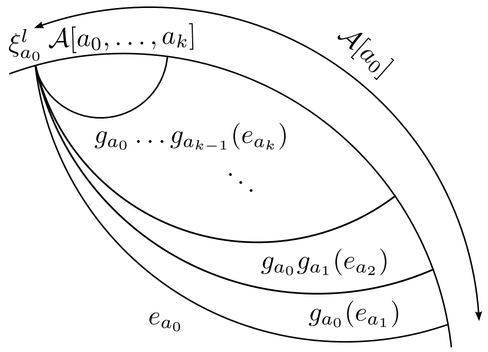

FIGURE 2. A left cuspidal word $a_0 \dots a_k$.

the form $\mathcal{A}[a_0, a_1, \dots, a_{n+1}]$ for $a_{n+1} \in \mathscr{A} \setminus \{\overline{a_n}\}$. Each one of these arcs corresponds to one of the arcs cut out by the sides of the ideal polygon $a_0 a_1 \dots a_n \mathcal{F}$ and contained in the previous arc $\mathcal{A}[a_0, a_1, \dots, a_n]$.

We summarize the previous discussion in the next result.

Proposition 2.1 (Bowen-Series). *If $(a_n)_{n\in\mathbb{N}}$ is the boundary expansion of $\xi \in \partial\mathbb{D}$ we have*

$$\xi = \bigcap_{n\in\mathbb{N}} g_{a_0} \cdots g_{a_n} \mathcal{A}[a_{n+1}].$$

Moreover, if $t \mapsto \gamma(t)$ is a hyperbolic geodesic ray with $\gamma(0) \in \mathcal{F}$ and ending at $\xi = \gamma(+\infty) \in \partial\mathbb{D}$, then the cutting sequence $(a_n)_{n\in\mathbb{N}}$ of γ coincides with the boundary expansion of ξ.

The Bowen-Series map $F\colon \partial\mathbb{D} \to \partial\mathbb{D}$ acts as the right shift on the space $\Sigma \subset \mathscr{A}^{\mathbb{N}}$ of those infinite words $(a_n)_{n\in\mathbb{N}}$ satisfying the no-backtracking condition (2.4). In other words we have

$$F\big([a_0, a_1, a_2, \dots]_{\partial\mathbb{D}}\big) = [a_1, a_2, \dots]_{\partial\mathbb{D}}.$$

Notice that the combinatorial no-backtracking condition (2.4) corresponds to the no-backtracking geometric phenomenon between hyperbolic geodesics we mentioned earlier.

2.3. **Cuspidal words and cuspidal sequences.** We now define an acceleration of the boundary expansion. The acceleration is obtained by grouping together all steps which correspond to excursions in the same cusp, in a similar way to how the Gauss map is obtained from the Farey map in the theory of classical continued fractions expansions.

Definition 2.2. A *left cuspidal word* (respectively a *right cuspidal word*) is a word $a_0 \dots a_k$ in the alphabet $\mathscr{A}$ which satisfies the no-backtracking condition (2.4) and such that the $k+1$ arcs

$$\mathcal{A}[a_0], \quad \mathcal{A}[a_0, a_1], \quad \dots \quad \mathcal{A}[a_0, \dots, a_{k-1}], \quad \mathcal{A}[a_0, \dots, a_k]$$

all share as a common left endpoint the left endpoint $\xi^l_{a_0}$ of $\mathcal{A}[a_0]$ (respectively as right endpoint the right endpoint $\xi^r_{a_0}$ of $\mathcal{A}[a_0]$), see Figure 2. We simply write that $a_0 \dots a_k$ is a *cuspidal word* when left or right is not specified. We say that a sequence $(a_n)_{n\in\mathbb{N}}$ is a *cuspidal sequence* if any word of the form $a_0 \dots a_n$ for $n \in \mathbb{N}$ is a cuspidal word and that it is *eventually cuspidal* if there exists $k \in \mathbb{N}$ such that $(a_{n+k})_{n\in\mathbb{N}}$ is a cuspidal sequence.

Equivalently, $a_0 \dots a_k$ is a left (respectively right) cuspidal word exactly when the arc $\mathcal{A}[a_0, \dots, a_k] \subset \partial\mathbb{D}$ has a vertex of $\mathcal{F}$ as its left (respectively right) endpoint. We remark that given an ideal vertex ξ, there is a *unique* left (right) cuspidal word of length $k+1$ such that the arc $\mathcal{A}[a_0, \dots, a_k]$ has ξ as left (right) endpoint. Indeed, such word can be obtained as follows. Let a_0 be such that $\mathcal{A}[a_0]$ has ξ as its left (right) endpoint. For any $0 \le i < k$, the arc $\mathcal{A}[a_0, \dots, a_i]$ of level i is subdivided at level $i+1$ into $2d-1$ arcs of level $i+1$ and $\mathcal{A}[a_0, \dots, a_{i+1}]$ is the unique one which contains the left (respectively right) endpoint of $\mathcal{A}[a_0, \dots, \alpha_i]$.

We will use cuspidal words to decompose an infinite word into blocks. Let us begin with a geometric description of this process. Let $\gamma\colon [0, +\infty) \to \mathbb{D}$ be a geodesic such that $\gamma(0) \in \mathcal{F}$ and that does not converge to a cuspidal point. Call $(a_n)_{n\in\mathbb{N}}$ its cutting sequence. Let $(t_n)_{n\in\mathbb{N}}$ be the sequence of times t_n when γ crosses a side of the tessellation of $\mathbb{D}$ given by $\mathcal{F}$. More precisely let t_0 such that $\gamma(t_0) \in s_{a_0}$ and

$$\gamma(t_n) \in g_{a_0} \circ \cdots \circ g_{a_{n-1}}(s_{a_n}), \tag{2.6}$$

for any $n \geq 1$. Define $n(0) = 0$ and, inductively for $r \in \mathbb{N}$, define $n(r+1)$ such that the segment $\gamma[t_{n(r)}, t_{n(r+1)})$ only intersects copies $g(s)$ of sides s of $\mathcal{F}$ all sharing one common endpoint.

Having this picture in mind, we define the cuspidal decomposition of an infinite word as follows. Consider an infinite word $(a_n)_{n\in\mathbb{N}}$ satisfying the no-backtracking condition (2.4) and that is not eventually cuspidal. Let $n(0) = 0$ and define $n(1) \geq 1$ to be the minimum time such that the arcs

$$\mathcal{A}[a_{n(0)}] \qquad \text{and} \qquad \mathcal{A}[a_{n(0)}, \dots, a_{n(1)}]$$

do not have a common endpoint. In other words, $n(1)$ is the minimum time such that the word $a_{n(0)} \dots a_{n(1)}$ is not cuspidal. Then set $C_0 = a_0 \dots a_{n(1)-1}$ to be the first maximal cuspidal word in the infinite word $(a_n)_{n\in\mathbb{N}}$. Similarly, for $r \geq 0$ define

$$n(r+1) = \min\{n > n(r) : a_{n(r)} \dots a_{n(r+1)} \text{ is not a cuspidal word}\},$$

and $C_r = a_{n(r)} \dots a_{n(r+1)-1}$ to be the r-th cuspidal maximal word in $(a_n)_{n\in\mathbb{N}}$. It is clear then that concatenating the cuspidal words $(C_r)_{r\in\mathbb{N}}$ we get the same infinite word as $(a_n)_{n\in\mathbb{N}}$ that is $a_0 a_1 \dots a_n \dots = C_0 C_1 \dots C_r \dots$.

In the sequel, we will also decompose bi-infinite words into cuspidal subwords. This is done as before, the only difference is that C_0 is the maximal cuspidal word containing a_0, and hence $n(0) \leq 0$.

2.4. **Parabolic words.** We end this section with two Lemmas that give a combinatorial description of cuspidal words, by showing that cuspidal words are obtained by repeating *parabolic words* (defined below), which are in one to one correspondence with cusps (see Corollary 2.5). The Lemmas were essentially proved in [1] (see Lemmas 4.8 and 4.9 in [1]). For completeness, we include their easy proofs in Appendix A.

Lemma 2.3. *Consider a word $a_0 \dots a_n$ in the alphabet $\mathscr{A}$ which satisfies the no-backtracking condition (2.4). Then*

(1) *The word $a_0 \dots a_n$ is left cuspidal if and only if*
$$g_{a_k}(\xi^l_{a_{k+1}}) = \xi^l_{a_k} \qquad \textit{for any} \qquad k = 0, \dots, n-1.$$
(2) *The word $a_0 \dots a_n$ is right cuspidal if and only if*
$$g_{a_k}(\xi^r_{a_{k+1}}) = \xi^r_{a_k} \qquad \textit{for any} \qquad k = 0, \dots, n-1.$$
(3) *The word $a_0 \dots a_n$ is left (resp. right) cuspidal if and only if $\overline{a_n} \dots \overline{a_0}$ is right (resp. left) cuspidal.*

The next Lemma connects cuspidal words and parabolic elements in Γ.

Lemma 2.4. *Let $a_0 \dots a_n$ be a left cuspidal word such that $a_0 \dots a_n a_0$ is a left cuspidal word too. Then $g = g_{a_0} \circ \dots \circ g_{a_n}$ is a parabolic element of Γ whose unique fixed point is*

$$\xi^l_{a_0} = \xi^r_{\overline{a_n}} \in \partial\mathbb{D}.$$

A word $a_0 \dots a_n$ as in the Lemma before is called a *left parabolic word* if it has minimal length. In the same way one defines a *right parabolic word*. Let us remark that $a_0 \dots a_n$ is left parabolic if and only if its *inverse word* $\overline{a_n} \dots \overline{a_0}$ is right parabolic, and the corresponding fixed point is $\xi^l_{a_0} = \xi^r_{\overline{a_n}}$. We write simply parabolic word when left or right is not specified.

From the Lemma, the following combinatorial description of cuspidal sequences follows (see point (2) of Lemma 4.9 in [1]).

Corollary 2.5. *For any right (left) cuspidal sequence $(a_n)_{n\in\mathbb{N}}$ there exists an integer $k \geq 1$ and a right (left) parabolic word $a_0 a_1 \cdots a_{k-1}$ such that $(a_n)_{n\in\mathbb{N}}$ is obtained repeating the word periodically, i.e. $a_n = a_{n \mod k}$ for every $n \in \mathbb{N}$.*

Remark 2.6. If $g_\alpha = g^{-1}_{\overline{\alpha}}$ is parabolic generator of Γ, the two sides s_α, $s_{\overline{\alpha}}$ share a common vertex ξ and $\xi = g_\alpha(\xi) = g^{-1}_{\overline{\alpha}}(\xi)$ is a cusp described by the (length one) parabolic words α and $\overline{\alpha}$. More in general, one can see that cusps of $\Gamma\backslash\mathbb{D}$ are in bijection with parabolic words, modulo inversion operation and cyclical permutation of the entries.

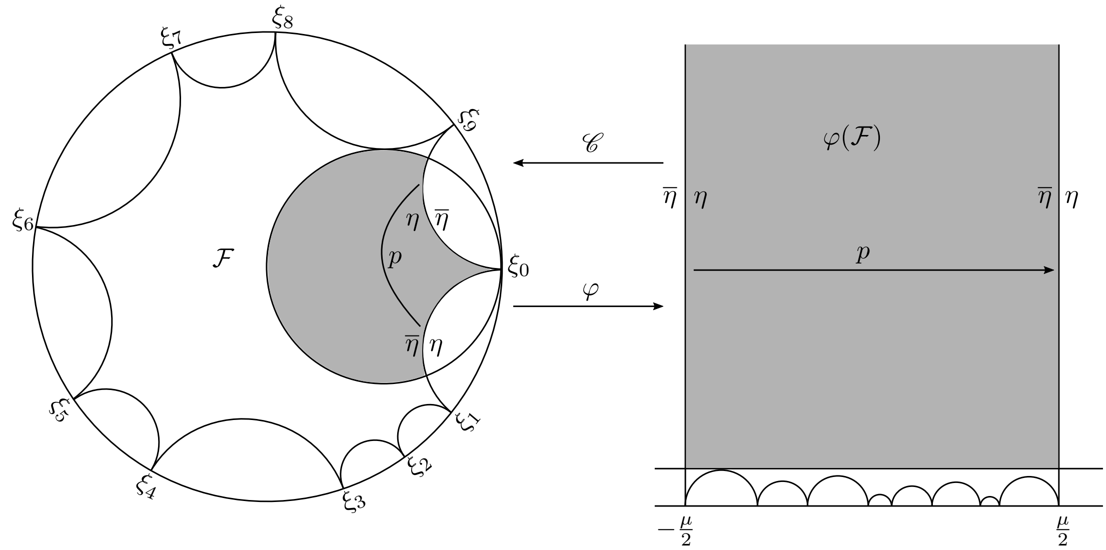

FIGURE 3. The fundamental domain $\mathcal{F}$ described in Lemma 3.1, with the horodisk $\mathcal{U}_1$ in grey in both figures.

3. Hall ray for the hyperbolic height

In this section we prove Theorem 1.4, namely the existence of Hall rays for Diophantine approximation on Fuchsian groups. This section is also meant as a guiding line for the following ones (and in particular Section 6), where the more difficult and technical proof of Theorem 1.8 will be presented. In particular, Section 3.4 introduces (the adaptation of) Hall's original argument using boundary expansions as a replacement of continued fraction which is used in both proofs (and referred to in Section 1.7). In Section 3.1 we choose a convenient fundamental domain for G and define the tessellation with respect to which to code geodesics. We then describe the Cantor set on the boundary $\partial\mathbb{D}$ of the disk corresponding to endpoints of geodesics which have bounded penetration in the cusps (see Section 3.2). This Cantor set will be proven to satisfy the assumptions of the Stable Hall Theorem later on, in Section 4.4. We will prove a Perron-like formula for sufficiently high geodesics in Section 3.3 and then prove Theorem 1.4.

3.1. **Preliminaries to the proofs of the main results.** In this section we are going to prepare the ground for the proofs of our two main results, Theorem 1.4 and Theorem 1.8. Let G be a fixed Fuchsian group. We assume that G is a non-uniform lattice and denote by $X = G\backslash\mathbb{H}$ the corresponding finite volume, not compact (orbifold) surface. We also assume that it is zonal, namely that ∞ is fixed by a parabolic element of G and hence projects to a cusp of X.

Conjugating G with an appropriate element of $\mathrm{PSL}(2,\mathbb{R})$ which fixes ∞, we normalize G so that $m = 1$, where m is the height of the fundamental horodisk at infinity (see (1.7)). We remark that, for Theorem 1.4, we are not losing any generality, since, by Remark 1.5, the presence of Hall rays in $\mathcal{L}(G,\infty)$ is preserved by this conjugation. For Theorem 1.8, we will first treat the case when G has $m = 1$, then we will show how to deduce a result for $m \neq 1$ from the result for $m = 1$ (see the proof of Theorem 1.8).

Let $\mu > 0$ the *width* of this cusp after this normalization, meaning that the matrix $p = \left(\begin{smallmatrix} 1 & \mu \\ 0 & 1 \end{smallmatrix}\right)$ is in G, and that p is not the power of another element in G.

Since G has finite covolume, it is finitely generated. Any finitely generated Fuchsian group contains a finite index normal subgroup Γ not containing any elliptic elements (see [12, 9]). It is well known that such Γ admits a fundamental domain for the action on $\mathbb{D}$ which is an *ideal* polygon, that is a hyperbolic polygon having finitely many vertices all lying on $\partial\mathbb{D}$ (see [39]). For technical reasons, we require some additional properties (in particular Condition (5) in the Lemma below) for the fundamental domain. We will hence construct a suitable fundamental domain $\mathcal{F}$ for the action of Γ on $\mathbb{D}$. The following Lemma summarizes the choice of $\mathcal{F}$. The reader can refer to Figure 3 for an example of a fundamental domain satisfying the requirements of the Lemma.

Here, and in the rest of the paper, we denote by $|\mathcal{A}[\alpha]|_{\partial\mathbb{D}}$ the length of the arc $\mathcal{A}[\alpha]$. Recall that $\varphi\colon \mathbb{D} \to \mathbb{H}$ is the inverse of the Cayley map defined in Equation (2.1).

Lemma 3.1. *There exists a fundamental domain $\mathcal{F} \subset \mathbb{D}$ for Γ with the following properties:*

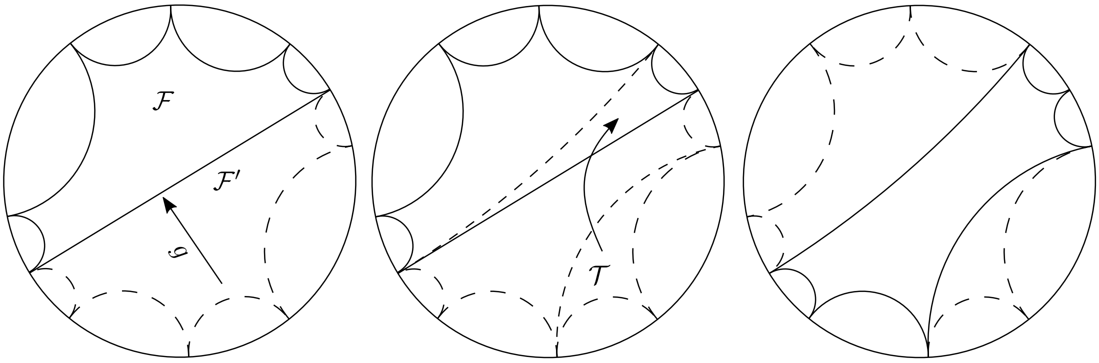

FIGURE 4. The surgery of the domain described in Section 3.1.

(1) *$\mathcal{F}$ is an ideal polygon with vertices $\xi_0, \dots, \xi_{2d-1} \in \partial\mathbb{D}$, where $\xi_0 := 1$ and hence $\varphi(\xi_0) = \infty$;*
(2) *the parabolic element p is a generator and identifies the side of $\mathcal{F}$ which share ξ_0 as endpoint; more precisely $p = g_\eta$, where $\eta \in \mathscr{A}$ is such that $\xi^l_\eta = \xi_0 = \xi^r_{\overline{\eta}}$;*
(3) *The endpoints of s_η and $s_{\overline{\eta}}$ different than ξ_0, that correspond to ξ_1 and ξ_{2d-1}, are such that*
$$\varphi(\xi^r_\eta) = \varphi(\xi_1) = \frac{\mu}{2}, \qquad \varphi(\xi^l_{\overline{\eta}}) = \varphi(\xi_{2d-1}) = -\frac{\mu}{2};$$
(4) *the origin of the disk $\mathbb{D}$ belongs to $\mathcal{F}$;*
(5) *for every arc $\mathcal{A}[\alpha]$ underlying a side s_α of $\mathcal{F}$, we have*
$$|\mathcal{A}[\alpha]|_{\partial\mathbb{D}} < \pi, \qquad \forall \alpha \in \mathscr{A}. \tag{3.1}$$

As we said above, Condition (5) is needed for technical reasons (more precisely for the distortion estimates in Section 4.2).

Proof. We will first construct a fundamental domain in $\mathbb{H}$ so that it verifies Condition (1), (2) and (3), then lift it to $\mathbb{D}$ and modify the choice so that also the other Conditions are verified.

Let H be the subgroup of Γ generated by p. Since p acts on the hyperbolic plane $\mathbb{H}$ by $z \mapsto z + \mu$, a fundamental domain for the action of H on the hyperbolic plane $\mathbb{H}$ is given by any vertical strip of width μ with the two vertical geodesics are identified by p. We choose the one centered on the vertical axis $\{z : -\frac{\mu}{2} < \mathrm{Re}(z) < \frac{\mu}{2}\}$. Recall that, given a matrix $g \in \mathrm{PSL}(2,\mathbb{R})$, not fixing ∞, its *isometric circle* I_g is the Euclidean semi-circle centered at $g^{-1} \cdot \infty = -d/c$ with radius $r_g = 1/|c|$. Then a fundamental domain for Γ is given by the intersection of a fundamental domain for H with the points that lie outside *every* isometric circle I_g given by the elements $g \in \Gamma \setminus H$. For more details we refer the interested reader to page 57 of [24]. The transformations that identify a pair of boundary sides of this fundamental domain are a set of generators for Γ.

Let us remark that the fundamental domain such constructed cannot have vertices inside $\mathbb{H}$, since each such point is necessarily fixed by some elliptic transformation. We notice also that the construction implies that p is one of the side pairings, and hence a generator for the group.

We call $\mathcal{F}$ the fundamental domain in $\mathbb{D}$ obtained transporting the fundamental domain we just built from $\mathbb{H}$ to $\mathbb{D}$ via φ (the inverse of the Caley map). By construction, $\mathcal{F}$ is an ideal polygons and has $1 = \varphi(\infty)$ as a vertex, which we will denote ξ_0. The images of the vertical lines with real part $\pm\mu/2$ are the to sides which share ξ_0 as a vertex. As in Figure 3, we will label by η (resp. $\overline{\eta}$) the side such that $\xi^l_\eta = \xi_0$ (resp. $\xi^r_{\overline{\eta}} = \xi_0$), so that $\varphi(\mu/2) = \xi^l_\eta$ (resp. $\varphi(-\mu/2) = \xi^r_{\overline{\eta}}$). This shows that Conditions (1)-(3) hold.

Moreover, since we are assuming that $m = 1$, we have that i belongs to the closure of the maximal Margulis neighborhood, which belongs by construction to the fundamental domain. Hence, the origin 0 of the disk $\mathbb{D}$ belongs to the closure of $\mathcal{F}$. In particular, this means that $|\mathcal{A}[\alpha]|_{\partial\mathbb{D}} \leq \pi$ for every $\alpha \in \mathcal{A}$. Thus, to ensure simultaneously Conditions (4) and (5), we just need to ensure that 0 does not lie on the boundary of the fundamental domain. This means that s_α is not a straight line in $\mathbb{D}$ for all α or equivalently that all the inequalities $|\mathcal{A}[\alpha]|_{\partial\mathbb{D}} \leq \pi$ inequalities are all strict.

We can always assume that this is the case up to performing the following surgery of the fundamental domain. If 0 lies on the boundary of the fundamental domain, it belongs to a side formed by a diameter, shared by two copies of the fundamental domain that we will call $\mathcal{F}$ and $\mathcal{F}'$ (see the left part of Figure 4).

Consider an ideal triangle $\mathcal{T}$, contained in $\mathcal{F}'$, bounded by the side s of $\mathcal{F}'$ that is paired with the diameter and an adjacent side of $\mathcal{F}'$, as shown in the middle picture in Figure 4. If g is the element of G that pairs s and the diameter, we choose as new fundamental domain $\big(\mathcal{F}' \setminus \mathcal{T}\big) \cup g(\mathcal{T})$, as in the right of Figure 4. By construction, the origin is an internal point of the new fundamental domain, which implies that (3.1) is satisfied.

Let us remark that, after the surgery we just explained, g still identifies two sides of the new fundamental domain, namely the two sides coming from the internal side of $\mathcal{T}$ and its image, that are dashed in the middle of Figure 4 and become solid in the right of the same picture. However, the generator that was matching the third side of $\mathcal{T}$ to some other side of $\mathcal{F}$ is changed. We need to take care that this side is *not* s_η or $s_{\overline{\eta}}$. This is always possible unless $\mathcal{F}$ has only 4 sides, necessarily identified in pairs by parabolic transformations. In this case, X must be the thrice punctured sphere (see p. 275 of [3]), which is unique in its isometry class (see, e.g. Theorem 9.8.8 of [34]). A fundamental domain in $\mathbb{D}$ for the thrice punctured sphere satisfying all the assumptions of the Lemma is given by the ideal quadrilateral with vertices $\{\pm 1, \pm i\}$, and the two sides that share the point 1 (resp. -1) identified. This completes the proof. $\square$

From now on, $\mathcal{F}$ will be a fundamental domain for the subgroup $\Gamma < G$ given by the Lemma. Let us remark that the fundamental domain $\mathcal{F}$ is a finite cover of a fundamental domain for G (obtained by *unfolding* the elliptic points) and hence the induced tessellation of the hyperbolic disk by $\mathcal{F}$ has tiles which are finite union of copies of a fundamental domain for G.

We will use the tessellation on $\mathbb{D}$ induced by the ideal polygon $\mathcal{F}$ to code geodesics using Bowen-Series coding explained in the previous Section. Let us stress that we do *not* pass to a *finite cover* of the surface X, which is fixed, but only code geodesics in $\mathbb{D}$ according to a super-tessellation, which is better suited to our purposes than the one corresponding to G. This is similar to what happens in the continued fractions case, where instead of coding the geodesics with respect to the tessellation given by the classical fundamental domain for $\mathrm{PSL}(2,\mathbb{Z})$ one uses the Farey tessellation, made by ideal triangles.

3.2. **Cantor sets and their sums.** Since the vertex ξ_0 of the fundamental domain $\mathcal{F}$ chosen in the previous section is such that $\varphi(\xi_0) = \infty$ the partition of $\partial\mathbb{D}$ given by the arcs $\mathcal{A}[\alpha]$ for $\alpha \in \mathscr{A}$ induces a partition of $\mathbb{R}$, and not only one of $\overline{\mathbb{R}} = \partial\mathbb{H}$. Given an infinite word $(a_n)_{n\in\mathbb{N}}$ that satisfies the no-backtracking condition (2.4), it will be useful to write

$$[a_0, \ldots, a_n, \ldots]_{\partial\mathbb{H}} := \varphi\big([a_0, \ldots, a_n, \ldots]_{\partial\mathbb{D}}\big).$$

Now, fix a positive integer $N \geq 2$ and let $\mathbb{B}_N = \mathbb{B}_N^\eta \subset \partial\mathbb{D}$ be the set of points ξ whose boundary expansion $(a_k)_{k\in\mathbb{N}}$ does not contain any cuspidal word of length $N+1$ and whose first letter is different from η and $\overline{\eta}$. One can show that the set $\mathbb{B}_N$ is a Cantor set: we are going to briefly describe its structure and its gaps in Section 4.2. Denote with $\mathbb{K}_N = \varphi(\mathbb{B}_N)$ its image in $\partial\mathbb{H}$. We remark that this is a compact set as $\mathbb{B}_N$ does not contain ξ_η^r nor a neighborhood around it and $\varphi(\xi_\eta^r) = \infty$.

Let $m_N := \min \mathbb{K}_N$, $M_N := \max \mathbb{K}_N$ and for $s \in \mathbb{N}$ let $\mathbb{K}_N^s = \mathbb{K}_N + s\mu$ denote the translates by $z \mapsto z + \mu$ of the Cantor set $\mathbb{K}_N$, so

$$\mathbb{K}_N^s := [m_N + s\mu, M_N + s\mu]\,.$$

The next Proposition is the analogue in our set up of Hall's theorem on the sum (difference) of Cantor sets given in terms of continued fractions.

Proposition 3.2. *There exists a natural number N_0 such that if $N \geq N_0$, for every integer $s \geq 0$, both Cantor sets $\mathbb{K}_N^s \pm \mathbb{K}_N$ contain an interval of size at least μ. More precisely, we have*

$$\begin{aligned} &\mathbb{K}_N^s + \mathbb{K}_N = [2m_N + s\mu, 2M_N + s\mu], && |\mathbb{K}_N^s + \mathbb{K}_N| = 2(M_N - m_N) > \mu, \\ &\mathbb{K}_N^s - \mathbb{K}_N = [-(M_N - m_N) + s\mu, M_N - m_N + s\mu], && |\mathbb{K}_N^s - \mathbb{K}_N| = 2(M_N - m_N) > \mu. \end{aligned}$$

We will prove the Proposition in Section 4.4. Let us remark that it can be proved as an application of the classical result by Hall on the sum of Cantor sets (the proof is very similar to the one given in [1] for similar Cantor sets). Since we need in any case to verify that the Cantor sets $\mathbb{K}_N$ and $\mathbb{K}_N^s$ satisfy the assumptions of the Stable Hall Theorem 1.12 for the proof of Theorem 1.8, we prove Proposition 3.2 in Section 4.4 as a special case of the Stable Hall Theorem.

As a Corollary, we have the following result, which is the starting point to build values in the Hall ray.

Corollary 3.3. *For any $L \geq \mu/2$, there exist two real numbers x_1, x_2 and an integer $s \geq 1$ such that $x_1, x_2 \in \mathbb{K}_N$ and*

$$L = s\mu + x_2 - x_1.$$

Proof. Remark that $\mathbb{K}_N^{s+1} - \mathbb{K}_N = (\mathbb{K}_N^s - \mathbb{K}_N) + \mu$. Thus, since by Proposition 3.2, the length of each $\mathbb{K}_N^s - \mathbb{K}_N$ is greater than μ, the intervals $\mathbb{K}_N^s - \mathbb{K}_N$, $s \in \mathbb{N}$, overlap and hence

$$\bigcup_{s\geq 1}(\mathbb{K}_N^s - \mathbb{K}_N) = [-(M_N - m_N) + \mu, +\infty) \supset \left[\frac{\mu}{2}, +\infty\right),$$

where the last inclusion follows since $M_N - m_N \geq \mu/2$ (also by Proposition 3.2). In particular for any $L \geq \mu/2$ there exists an integer $s \geq 1$ such that $L \in \mathbb{K}_N^s - \mathbb{K}_N$. Since $\mathbb{K}_N^s = \mathbb{K}_N + s\mu$, this means that there exist $x_1, x_2 \in \mathbb{K}_N$ such that $L = (x_2 + s\mu) - x_1$ as desired. □

3.3. **A generalized Perron formula via boundary expansions.** The starting point for the proof of existence of a Hall ray is a generalization of Perron's formula (1.11) for values of the Lagrange spectrum, in which classical continued fractions are replaced by the Bowen-Series boundary expansions with respect to the finite index subgroup $\Gamma < G$ defined in Section 3.1.

Let us remark that given a geodesic $\gamma = \gamma(x, y)$ whose cutting sequence with respect to the tessellation defined in Section 3.1 is $(a_n)_{n\in\mathbb{Z}}$, the two endpoints x and y of γ are given by

$$y = [a_0, \dots, a_n, \dots]_{\partial\mathbb{H}} \qquad \text{and} \qquad x = [\overline{a_{-1}}, \dots, \overline{a_{-n}}, \dots]_{\partial\mathbb{H}}.$$

The bars in the expression for x are due to the fact that moving from $\mathcal{F}$ to $\mathbb{R}$ towards x we are traveling backwards along γ, as explained at the very end of Section 2.1.

Hence, introduce the following notation:

$$[a_0, \dots, a_n, \dots]_{\partial\mathbb{H}}^{-} = [\overline{a_0}, \dots, \overline{a_n}, \dots]_{\partial\mathbb{H}}. \tag{3.2}$$

Let $\mathrm{ht}_G(\gamma)$ denote the essential height of the geodesic γ, see Section 1.3.

Lemma 3.4 (Perron's formula for the essential height). *Let γ be a complete geodesic with $\gamma(0) \in \mathcal{F}$ and cutting sequence $(a_n)_{n\in\mathbb{Z}}$. Suppose that $\mathrm{ht}_G(\gamma) > 1$. Then*

$$\mathrm{ht}_G(\gamma) = \frac{1}{2}\limsup_{n\in\mathbb{Z}} |[a_n, a_{n+1}, \dots]_{\partial\mathbb{H}} - [a_{n-1}, a_{n-2}, \dots]_{\partial\mathbb{H}}^{-}|. \tag{3.3}$$

We recall that are here assuming that $m = 1$, where m is the height of the maximal Margulis neighborhood. More in general, the formula Equation (3.3) holds for $\mathrm{ht}_G(\gamma) > m$. In the proof of Lemma 3.4, given below (and in the rest of the paper) we will use the following observation, which follows from the definition of the Bowen-Series coding and boundary expansions.

Lemma 3.5. *For any non-zero integer j, let γ_j be the geodesic defined by*

$$\gamma_j(t) := \begin{cases} g_{a_{j-1}}^{-1} g_{a_{j-2}}^{-1} \cdots g_{a_0}^{-1} \cdot \gamma(t), & \text{if } j \geq 1; \\ g_{a_j}^{-1} g_{a_{j-1}}^{-1} \cdots g_{a_{-1}}^{-1} \cdot \gamma(t), & \text{if } j < 0. \end{cases}$$

The cutting sequence of γ_j is $(a_{n+j})_{n\in\mathbb{Z}}$ and the endpoints of γ_j are

$$y_j := [a_j, a_{j+1}, \dots]_{\partial\mathbb{H}}, \qquad x_j := [a_{j-1}, a_{j-2}, \dots]_{\partial\mathbb{H}}^{-}.$$

The geodesic γ_j has the property that $\gamma_j(t) \in \mathcal{F}$ for $t_j < t < t_{j+1}$ where t_j is the j^{th}-crossing with the coding tessellation (compare with Equation (2.2)). We will call it the j^{th} normalized geodesic.

Proof. For $j \geq 0$, the statements follow from the definitions of coding and boundary expansion, see Proposition 2.1 and also Equation (2.2). When $j < 0$, consider the geodesic $\gamma'(t) := \gamma(-t)$, whose cutting sequence $(a'_n)_{n\in\mathbb{N}}$ is hence given by $a'_n = \overline{a_{n-1}}$, and apply the previous case. □

Proof of Lemma 3.4. We first claim that, if the essential height (1.4) of a geodesic $\gamma\colon \mathbb{R} \to X$ is larger than 1, it is sufficient to consider the elements $g \in G$ such that $g \cdot \gamma \cap \varphi(\mathcal{F}) \neq \varnothing$. In fact, consider an arbitrary element $g \in G$ and let $\mathcal{U}_1$ be the fundamental horodisk defined in (1.7). If $g \cdot \gamma \cap \mathcal{U}_1 = \varnothing$, then we have that $\mathrm{ht}(g \cdot \gamma) \leq 1$. Otherwise, since the fundamental domain $\varphi(\mathcal{F}) \subset \mathbb{H}$, by construction (see Lemma 3.1 and recall that $m = 1$), contains a Euclidean rectangle delimited by $\operatorname{Im} z = 1$ and two vertical lines at $-\frac{\mu}{2}$ and $\frac{\mu}{2}$, there exists an integer k such that $(p^k \cdot g \cdot \gamma) \cap \varphi(\mathcal{F}) \neq \varnothing$. Clearly $p^k \cdot g$ and g are equivalent modulo infinity. Moreover we have $\mathrm{ht}(p^k g \cdot \gamma) = \mathrm{ht}(g \cdot \gamma) \geq 1$. This proves the claim.

Let us now remark that the elements $g \in G$ that satisfy $g \cdot \gamma \cap \varphi(\mathcal{F}) \neq \varnothing$, i.e., bring back a piece of the geodesic γ to the fundamental domain, can be exactly obtained using the cutting sequence of γ, i.e. are exactly the elements of the form $g_{a_k}^{-1} \cdots g_{a_0}^{-1}$ for $k \geq 0$ and $g_{\overline{a_k}}^{-1} \cdots g_{\overline{a_{-1}}}^{-1}$ for $k < 0$. Thus, by Lemma 3.5 and the definition of hyperbolic naive height (1.3), we get (3.3). □

Finally, let us record as a Lemma some simple observations, which follows from the choice and geometry of the fundamental domain (we refer the reader to Figures 3 and 5).

Lemma 3.6. *Let $\gamma(x, y)$ be a geodesic with initial endpoint x and final endpoint y, whose cutting sequence, assuming that $\gamma(0) \in \mathcal{F}$, is $(a_n)_{n\in\mathbb{Z}}$. Then:*

1. *if $y > x$, then if $a_0 \neq \eta$ we have $y < \mu/2$ and if $a_{-1} \neq \eta$ then $x > -\mu/2$;*
2. *if $x > y$, then if $a_0 \neq \overline{\eta}$ we have $y > -\mu/2$ and if $a_{-1} \neq \overline{\eta}$ then $x < \mu/2$;*
3. *combining (1) and (2), if both a_0 and a_{-1} do not belong to $\{\eta, \overline{\eta}\}$, we have that*

$$\mathrm{ht}(\gamma) = \frac{|x - y|}{2} \leq \frac{\mu}{2}.$$

Proof. For Part (1) (resp. Part (2)), simply recall that the fundamental domain $\varphi(\mathcal{F}) \subset \mathbb{H}$ is bounded by the two vertical lines $\{\mathrm{Re}\, z = -\mu/2\}$ and $\{\mathrm{Re}\, z = \mu/2\}$, whose *external* labels are $\overline{\eta}$ and η respectively (see Figure 3). Thus, for a geodesic with $\gamma(0) \in \varphi(\mathcal{F})$ to cross the side labeled by η (resp. $\overline{\eta}$), so that $a_0 = \eta$ (resp. $a_0 = \overline{\eta}$) the final endpoint has to be greater than $\mu/2$ (resp. less than $-\mu/2$). The arguments for a_{-1} are analogous, just reversing time and thus exchanging the role of the endpoints. Finally, Part (3) follows simply by combining (1) and (2). □

3.4. **Hall's argument for the height in any zonal Fuchsian group.** We now have all the elements to conclude the proof of Theorem 1.4 following the scheme of Hall's original proof.

Proof of Theorem 1.4. Let N_0 be given by Proposition 3.2. We will show that

$$[L_0, +\infty] \subset \mathcal{L}(X, \infty), \qquad \text{for any } L_0 > (N_0 + 1)\mu. \tag{3.4}$$

Step one: construction of the bi-infinite word.

By Corollary 3.3 (remark that in particular $L \geq \mu/2$ so we can apply it), there exist $x_1, x_2 \in \mathbb{K}_N$ and $s \geq 1$ such that $L = s\mu + x_2 - x_1$. In particular, write $y = [a_0, a_1, \ldots, a_n, \ldots]_{\partial\mathbb{H}}$ and $x = [b_0, b_1, \ldots, b_n, \ldots]_{\partial\mathbb{H}}$, with both sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ in $\mathbb{K}_N$. Thus,

$$L = s\mu + [a_0, a_1, \ldots, a_n, \ldots]_{\partial\mathbb{H}} - [b_0, b_1, \ldots, b_n, \ldots]_{\partial\mathbb{H}}. \tag{3.5}$$

Let us now construct an infinite word $(w_n)_{n\in\mathbb{Z}}$ that will give the cutting sequence of a geodesic γ such that

$$L_G(\gamma) = 2\,\mathrm{ht}_G(\gamma) = L.$$

We will define blocks of entries W_j, $j \in \mathbb{Z}$, which we will then concatenate to form the word $(w_n)_{n\in\mathbb{Z}}$. Recall that η is the letter such that $p = g_\eta$. Set

$$W_j = \overline{b_{|j|}} \ldots \overline{b_0}\eta^s a_0 \ldots a_{|j|}, \qquad j \in \mathbb{Z},$$

where η^s means that the letter η is repeated s times. We remark that, by definition of the Cantor set $\mathbb{K}_N$, we have that $a_0 \neq \overline{\eta}$ and $\overline{b_0} \neq \overline{\eta}$ (since $b_0 \neq \eta$). Thus W_j satisfies the no-backtracking condition (2.4). Let us choose letters to interpolate between W_j (which ends in a_j) and W_{j+1} (which starts with $\overline{b_{j+1}}$) as follows. Since the alphabet $\mathscr{A}$ has cardinality $2d > 3$, we can pick δ_j such that $\delta_j \neq \overline{a_j}$ and $a_j\delta_j$ is not a cuspidal word and then δ'_j such that $\delta'_j \neq \overline{\delta_j}$, $\delta'_j \neq b_{j+1}$ and $\delta'_j\overline{b_{j+1}}$ is not a cuspidal word. Thus, the word

$$a_0 \ldots a_j\delta_j\delta'_j\overline{b_{j+1}} \ldots \overline{b_0} \tag{3.6}$$

satisfies the no-backtracking condition (2.4). Moreover, as the two infinite words $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are in $\mathbb{B}_N$, and thanks to our choice of δ_j, δ'_j, the word in (3.6) does not contain any parabolic word of length bigger than N. It follows that the infinite word $(w_n)_{n\in\mathbb{Z}}$ obtained juxtaposing the blocks W_j, δ_j, δ'_j in increasing order of $j \in \mathbb{Z}$ satisfies the no-backtracking condition or, in other words, is actually the cutting sequence of some geodesic γ.

In the next two steps we will show that $\mathrm{ht}_G(\gamma) = L/2$ and hence $L_G(\gamma) = 2\,\mathrm{ht}_G(\gamma) = L$. First, in *Step two*, we will check that if we evaluate the lim sup in (3.3) along the subsequence of times where we see the

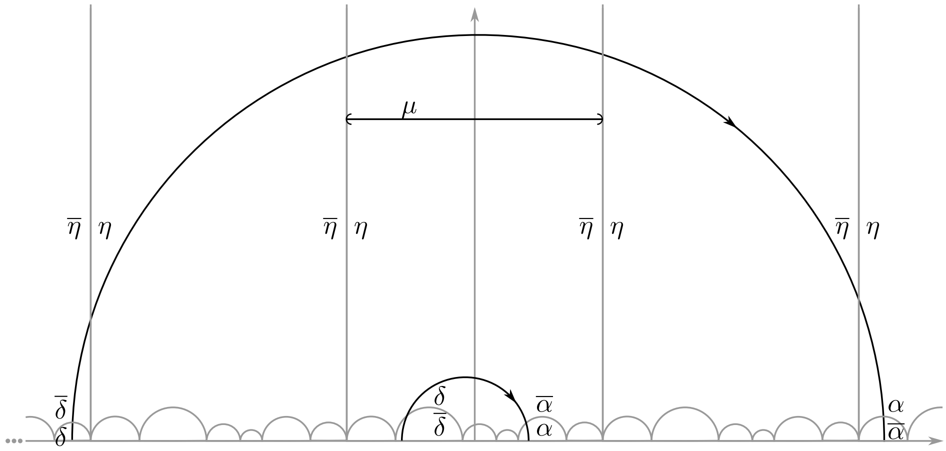

FIGURE 5. In grey, the fundamental domain $\varphi(\mathcal{F})$, and a portion of the tessellation induced by it, with some of the side labels; in black, a geodesic with larger height, coded by $\dots, \overline{\delta}, \eta, \eta, \eta, \eta, \overline{\alpha}, \dots$, and one with smaller height, coded by $\dots, \delta, \alpha, \dots$.

parabolic word η^s we obtain the desired value. Then, in *Step three*, we will show that this subsequence actually realizes the lim sup.

Step two: the infinite word realizes the desired Lagrange value along a subsequence of times.

We remark that η^s is a cuspidal word and, since a_0 and $\overline{b_0}$ are different from η, it is actually a maximal cuspidal word. Let r_k be the subsequence of times where an occurrence of the central word η^s in W_k begins. Thus,

$$[w_{r_k}, w_{r_k+1}, \cdots]_{\partial\mathbb{H}} = [\eta, \eta, \dots, \eta, w_{r_k+s}, w_{r_k+s+1}, \cdots]_{\partial\mathbb{H}} = [\eta, \eta, \dots, \eta, a_0, a_1, \dots]_{\partial\mathbb{H}}.$$

By Lemma 3.5 (or directly by Proposition 2.1), this endpoint is obtained acting by g_η^s on

$$[w_{r_k+s} w_{r_k+s+1} \dots]_{\partial\mathbb{H}} = [a_0, a_1, \dots]_{\partial\mathbb{H}}.$$

Since $g_\eta^s = p^s$ (recall Lemma 3.1), it acts on $\mathbb{H}$ as the translation $z \mapsto z + s\mu$. Thus, evaluating Perron's formula (3.3) along the subsequence r_k, as $|k| \to \infty$, we get

$$\begin{aligned}
\lim_{|k|\to\infty} \frac{\left|[w_{r_k}, w_{r_k+1}, \dots]_{\partial\mathbb{H}} - [w_{r_k-1}, w_{r_k-2}, \dots]_{\overline{\partial\mathbb{H}}}\right|}{2} &= \lim_{|k|\to\infty} \frac{\left|[\eta, \dots, \eta, a_0, \dots]_{\partial\mathbb{H}} - [\overline{b_0}, \overline{b_1}, \dots]_{\overline{\partial\mathbb{H}}}\right|}{2} \\
&= \lim_{|k|\to\infty} \frac{(s\mu + [a_0, a_1, \dots]_{\partial\mathbb{H}} - [b_0, b_1, \dots]_{\partial\mathbb{H}})}{2} \\
&= \frac{L}{2},
\end{aligned} \tag{3.7}$$

where in the last line we used also the definition (3.2) of $[\,\cdot\,]_{\overline{\partial\mathbb{H}}}$,the form of the words W_k and (3.5).

Step three: estimates on the remaining times. We now estimate the value of the lim sup in the formula (3.3) for the other times. For any $j \in \mathbb{Z}$, let γ_j be the j^{th} renormalized geodesic defined in Lemma 3.5, which is a geodesic coded by $(w_{n+j})_{n\in\mathbb{Z}}$ with endpoints $\tilde{x}_j := [w_j, w_{j+1}, \dots]_{\partial\mathbb{H}}$ and $\tilde{y}_j := [w_{j-1}, w_{j-2}, \dots]_{\overline{\partial\mathbb{H}}}$ (see Lemma 3.5).

We begin with the simple remark that if we see a block of k consecutive η's, i.e. $w_n = \cdots = w_{n+k-1} = \eta$, then the naive height value $|\tilde{x}_j - \tilde{y}_j|/2$ of γ_j remains constant for $n \le j \le n+k$. In fact, by Lemma 3.5, γ_j for $n < j \le n+k$ is obtained from γ_n applying a power of $g_\eta^{-1} = p^{-1}$, i.e. rigidly translating to the left by μ the two endpoints of γ_n. In other words, for every occurrence $w_n = \eta$ of η (or similarly for $\overline{\eta}$), we have

$$|[\eta, w_{n+1}, \dots]_{\partial\mathbb{H}} - [w_{n-1}, \dots]_{\overline{\partial\mathbb{H}}}| = |[w_{n+1}, \dots]_{\partial\mathbb{H}} - [\eta, w_{n-1}, \dots]_{\overline{\partial\mathbb{H}}}|.$$

The same remark also holds for a sequence of consecutive $\overline{\eta}$, in this case we act by $g_{\overline{\eta}}^{-1} = g_\eta = p$ and hence we are translating the endpoints to the right by μ.

In particular, by *Step one*, this gives that, for any $k \in \mathbb{Z}$, and any $r_k \le j \le r_k + s$, $\mathrm{ht}(\gamma_j) = \mathrm{ht}(\gamma_{r_k})$ and the argument in Perron's formula (3.3) is constant and equal to $L/2$. We will now evaluate the argument of Perron's formula (3.3) for any j which is not of this form. We will consider four sub-cases.

Case i: If both w_j and w_{j-1} do not belong to $\{\eta, \overline{\eta}\}$, by Lemma 3.6 both the endpoints of γ_j lie inside the interval $[-\frac{\mu}{2}, \frac{\mu}{2}]$, thus $2\,\mathrm{ht}(\gamma_j) \leq \mu$.

Case ii: Suppose now that $w_j \in \{\eta, \overline{\eta}\}$, but $w_{j-1} \notin \{\eta, \overline{\eta}\}$. By our assumption on j, and the structure of the bi-infinite word $(w_n)_{n\in\mathbb{Z}}$, we can have at most N consecutive η or $\overline{\eta}$, beginning with w_j. This implies that the geodesic γ_j crosses at most N vertical lines of the form $k\mu/2$ for $k \in \mathbb{N}$, see Figure 5. Let $N_j \leq N$ be the number of lines actually crossed and assume $w_j = \eta$ (if $w_j = \overline{\eta}$ the argument is analogous). So we have

$$\begin{aligned}
2\,\mathrm{ht}(\gamma_j) &= |[w_j, w_{j+1}, \dots]_{\partial\mathbb{H}} - [w_{j-1}, w_{j-2}, \dots]^{-}_{\partial\mathbb{H}}| \\
&= |[\eta, \dots, \eta, w_{n+N_j}, \dots]_{\partial\mathbb{H}} - [w_{j-1}, w_{j-2}, \dots]^{-}_{\partial\mathbb{H}}| \\
&\leq N_j\mu + |[w_{n+N_j}, w_{n+N_j+1}, \dots]_{\partial\mathbb{H}} - [w_{j-1}, w_{j-2}, \dots]^{-}_{\partial\mathbb{H}}| \\
&< N_j\mu + \left|\frac{\mu}{2} - \left(-\frac{\mu}{2}\right)\right| \leq N\mu + \mu \leq L.
\end{aligned}$$

Case iii: is symmetric to the previous one. If $w_{j-1} \in \{\eta, \overline{\eta}\}$, but $w_j \notin \{\eta, \overline{\eta}\}$, we can repeat the previous argument in the past, i.e. we have that, if $w_{j-1} = \eta$,

$$\begin{aligned}
2\,\mathrm{ht}(\gamma_j) &= |[w_j, w_{j+1}, \dots]_{\partial\mathbb{H}} - [\eta, \dots, \eta, w_{j-N_j-1}, w_{j-N_j-2}, \dots]^{-}_{\partial\mathbb{H}}| \\
&\leq N_j\mu + |[w_j, w_{j+1}, \dots]_{\partial\mathbb{H}} - [w_{j-N_j-1}, w_{j-N_j-2}, \dots]^{-}_{\partial\mathbb{H}}| \\
&< N_j\mu + \left|\frac{\mu}{2} - \left(-\frac{\mu}{2}\right)\right| \leq N\mu + \mu \leq L
\end{aligned}$$

and an analogous estimate holds for $w_{j-1} = \overline{\eta}$.

Case iv: Finally, if both $w_j, w_{j-1} \in \{\eta, \overline{\eta}\}$, since by the no backtracking condition $w_j \neq \overline{w_{j-1}}$, either $w_j = w_{j-1} = \eta$ or $w_j = w_{j-1} = \overline{\eta}$. We claim that in this case we can reduce this case to one of the previous Steps using the remark at the beginning of this step that the naive height of γ_j does not change during a block of consecutive η or $\overline{\eta}$. More precisely, we look for the first $m \leq j$ such that $w_{m-1} \notin \{\eta, \overline{\eta}\}$, that is we choose the time m where the cuspidal word contain w_j begins. If $m = r_k$ for some k, $\mathrm{ht}(\gamma_j) = \mathrm{ht}(\gamma_{r_k}) = L/2$ by *Step two*. Otherwise, the cuspidal word beginning at w_m must be at most of length N, by construction of the word $(w_n)_{n\in\mathbb{Z}}$. In this case, we can use the above estimates for the time $j = m$ to see that $\mathrm{ht}(\gamma_j) = \mathrm{ht}(\gamma_m) < L/2$.

Thus, combining *Step two* and *Step three*, Perron's formula shows that $\mathrm{ht}_G(\gamma) = L/2$ and hence $L_G(\gamma) = 2\,\mathrm{ht}_G(\gamma) = L$. This concludes the proof. □

4. Cantor sets in the boundary

In this section we describe the Cantor set $\mathbb{B}_N$ in the boundary of the disk which correspond to endpoints of geodesics whose excursions in the cusps are *bounded*, in the sense that their boundary expansion contains only cuspidal words of length bounded by N. We first describe their gaps combinatorially, through the symbolic sequences which correspond to them (see Section 4.1), then prove some distortion estimates (see Section 4.2) which will be needed to apply the Stable Hall Theorem 1.12.

Let us first recall the definition of the Cantor set we want to study through the cuspidal acceleration of the boundary expansion. Consider an infinite word $(a_n)_{n\in\mathbb{N}}$ satisfying the no-backtracking condition (2.4) and that is not eventually cuspidal. The cuspidal acceleration described in Section 2.3 provides a sequence of integers $0 =: n(0) < n(1) < n(2) < \dots$ and maximal cuspidal words $C_r := a_{n(r)}, \dots, a_{n(r+1)-1}$ with $r \in \mathbb{N}$, such that for any $r \geq 1$, $n(r+1)$ is the minimal $n > n(r)$ such that $a_{n(r)} \dots a_{n(r+1)}$ is *not* a cuspidal word. Let us write $\ell(a_0 \dots a_{n-1}) = n$ for the length of a word, so that the length $\ell(C_r)$ of the r^{th} cuspidal word is

$$\ell(a_{n(r)}, \dots, a_{n(r+1)-1}) = n(r+1) - n(r).$$

Fix a positive integer N and a letter $\eta \in \mathscr{A}$ and let $\mathbb{B}_N := \mathbb{B}(N, \eta) \subset \partial\mathbb{D}$ be the Cantor set defined in Section 3.2, which consists of the set of points whose boundary expansion $(a_n)_{n\in\mathbb{N}}$ is such that $a_0 \neq \eta, \overline{\eta}$ and the sequence $(a_n)_{n\in\mathbb{N}}$ does not contain any cuspidal word of length $N+1$, that is for any $r \in \mathbb{N}$ we have

$$\ell(C_r) = n(r+1) - n(r) \leq N. \tag{4.1}$$

One can prove that $\mathbb{B}_N$ is indeed a Cantor set. The proof is given in § 7.2 of [1] for some analogous Cantor sets, so we refer the interested reader to it. In the next section, though, we recall a combinatorial description of the gaps of the Cantor set $\mathbb{B}_N$.

4.1. **Combinatorial description of the gaps in the Cantor set.** The union of the gaps of the Cantor set $\mathbb{B}_N$ can be described using *deleted arcs* and the corresponding *forbidden words* as follows. If $\xi \in \mathbb{B}_N$, by definition no cuspidal word of length $N+1$ can appear in its boundary expansion. An *N-forbidden word* of level m, which for short we will call a (N,m)*-forbidden word*, is a finite word which contains a cuspidal word of length $N+1$ after the m^{th} letter, i.e. a word $a_0 \dots a_{m+N}$ of length $m+N+1$ whose cuspidal decomposition

$$a_0 \dots a_{m+N} = C_1 \dots C_r$$

is such that the first $r-1$ terms $C_1, \dots, C_{r-1}$ are cuspidal words of length strictly smaller than $N+1$ (so that the condition (4.1) is satisfied), while for the last term we have $\ell(C_r) = N+1$, that is

$$C_r = a_m \dots a_{m+N}.$$

Arcs $\mathcal{A}[a_0, \dots, a_{m+N}]$ corresponding to (N,m)-forbidden words with $m \in \mathbb{N}$ are called (N,m)-*deleted arcs*. Let $\mathcal{D}_N$ be the family whose elements are all the (N,m)-deleted arcs for $m \in \mathbb{N}$. Elements $\mathcal{A} \in \mathcal{D}_N$ have mutually disjoint interior and are exactly all the arcs which are removed from $\partial\mathbb{D}$ to obtain $\mathbb{B}_N$. Any gap B of $\mathbb{K}_N$ is the countable union $B = \cup_{k\in\mathbb{Z}} \mathcal{A}_k$ of a collection of *adjacent* deleted arcs in $\mathcal{D}_N$, where by adjacent we mean that $\max \mathcal{A}_k = \min \mathcal{A}_{k+1}$ for any $z \in \mathbb{Z}$; the length $|\mathcal{A}_k|_{\partial\mathbb{D}}$ shrinks exponentially as $|k| \to \infty$. An explicit description of all arcs $\mathcal{A}_k$ fitting together in the same gap is given in § 7.2 in [1]. The explicit description is not needed in this paper: we just need to know that if we set $\mathcal{G}_N \subset \partial\mathbb{D}$ to be the union $\mathcal{G}_N := \bigcup_{\mathcal{A}\in\mathcal{D}} \mathcal{A}$ of all deleted arcs $\mathcal{A} \in \mathcal{D}_N$, the set $\mathcal{G}_N$, described as a countable union of closed arcs, is an open set and it coincides precisely with the union of all gaps of $\mathbb{B}_N$, in other words

$$\mathbb{B}_N = \big(\partial\mathbb{D} \setminus (\mathcal{A}[\overline{\eta}] \cup \mathcal{A}[\eta])\big) \setminus \mathcal{G}_N.$$

Let us now describe how to hierarchically produce all the gaps of the Cantor set $\mathbb{B}_N$ through the generators of the boundary expansions. This also gives an ordering of the gaps and the corresponding intervals into *levels*.

The *gaps of level zero* are in one to one correspondence with the $2d-1$ ideal vertices $\xi_1, \dots, \xi_{2d-1}$ of the fundamental domain $\mathcal{F}$. Let $\xi \in \partial\mathbb{D}$ be any such vertex of $\mathcal{F}$. Let us denote by α^l_ξ and α^r_ξ the two letters in $\mathscr{A}$ such that the arcs $\mathcal{A}[\alpha^l_\xi]$ and $\mathcal{A}[\alpha^r_\xi]$ share ξ as an endpoint. As the notation suggest, we assume that $\mathcal{A}[\alpha^r_\xi]$ has ξ as *right* endpoint, while $\mathcal{A}[\alpha^l_\xi]$ has ξ as *left* endpoint. Thus, in the clockwise ordering,

$$\max \mathcal{A}[\alpha^r_\xi] = \xi = \min \mathcal{A}[\alpha^l_\xi].$$

As we observed after Definition 2.2, there are exactly two $(N,0)$-forbidden words $\alpha^l_0 \dots \alpha^l_N$ and $\beta^r_0 \dots \beta^r_N$ with respectively $\alpha^l_0 = \alpha^l_\xi$ and $\beta^r_0 = \alpha^r_\xi$. Moreover $\alpha^l_0 \dots \alpha^l_N$ is left cuspidal and $\beta^r_0 \dots \beta^r_N$ is right cuspidal. The two corresponding $(N,0)$-deleted arcs $\mathcal{A}[\alpha^l_0, \dots, \alpha^l_N]$ and $\mathcal{A}[\beta^r_0, \dots, \beta^r_N]$ share a common endpoint, indeed we have $\max \mathcal{A}[\alpha^r_0, \dots, \alpha^r_N] = \xi = \min \mathcal{A}[\beta^l_0, \dots, \beta^l_N]$.

The gap $B[\xi]$ of level zero corresponding to some ξ is by definition the connected component of $\mathcal{G}$ which contains $\mathcal{A}[\alpha^l_0, \dots, \alpha^l_N] \cup \mathcal{A}[\beta^r_0, \dots, \beta^r_N]$. Thus in particular we have

$$\mathcal{A}[\alpha^l_0, \dots, \alpha^l_N] \cup \mathcal{A}[\beta^r_0, \dots, \beta^r_N] \subset B[\xi].$$

Keeping in mind the geometry of cuspidal arcs, one can also show that

$$B[\xi] \subset \mathcal{A}[\alpha^l_0, \dots, \alpha^l_{N-1}] \cup \mathcal{A}[\beta^r_0, \dots, \beta^r_{N-1}]. \tag{4.2}$$

The level zero gaps are $B[\xi_i]$ for $1 \le i \le 2d-1$. For any $n \ge 1$, to define the *gaps of level* n, we transport the gaps of level zero through the generators as follows. Let $B[\xi]$ be a gap of level zero and $a_0 \dots a_{n-1}$ an admissible word with $a_0 \ne \overline{\eta}, \eta$ and such that both the two words

$$a_0 \dots a_{n-1}\alpha^l_0 \dots \alpha^l_N \qquad \text{and} \qquad a_0 \dots a_{n-1}\beta^l_0 \dots \beta^l_N$$

are admissible and moreover form (N,n)-forbidden words, where $\alpha^l_0 = \alpha^l_\xi$ and $\beta^r_0 = \alpha^r_\xi$. The corresponding gap of level n is the open interval

$$B[a_0, \dots, a_{n-1}; \xi] = g_{a_0} \circ \dots \circ g_{a_{n-1}}(B[\xi]).$$

Correspondingly, it is also convenient to introduce the notion of *intervals of level* n. To do so, for any $n \in \mathbb{N}$ and any letter $\alpha \in \mathscr{A}$, recall that the ideal vertices of the arc $\mathcal{A}[\alpha]$ are ξ^l_α and ξ^r_α. Define the compact arc $K[\alpha]$ as the unique connected component of $\partial\mathbb{D} \setminus \bigcup_{0\le i\le 2d-1} B[\xi_i]$ that shares an endpoint both with $B[\xi^l_\alpha]$ and $B[\xi^r_\alpha]$. This defines the $2d$ *intervals of level zero* $K[\alpha]$, $\alpha \in \mathscr{A}$.

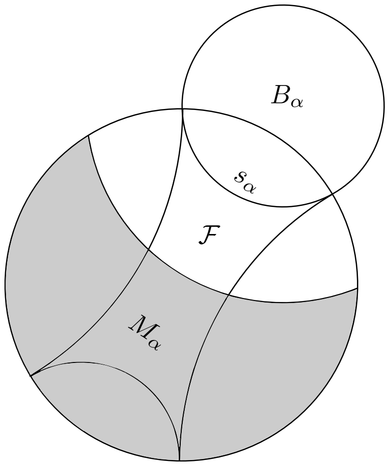

FIGURE 6. The objects defined in Section 4.2.

To define the intervals of other levels, for any admissible word $a_0, \ldots, a_{n-1}$ of length n, define the intervals of level n compatible with $a_0, \ldots, a_{n-1}$ as the intervals

$$K[a_0, \ldots, a_{n-1}; \alpha] := g_{a_0} \circ \cdots \circ g_{a_{n-1}} (K[\alpha]),$$

where α ranges among all the letters with $\alpha \neq \overline{a_{n-1}}$.

4.2. **Distortion estimates.** In the following we consider the Poincaré disc $\mathbb{D}$ as an open subset of the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$; $|\cdot|$ will denote the usual absolute value in $\mathbb{C}$ and by open disc we mean a set of the form $\{z \in \mathbb{C} : |z - z_0| < r\}$ for some $z_0 \in \mathbb{C}$ and radius $r > 0$.

For any letter $\alpha \in \mathscr{A}$, recall that s_α is the side of $\mathcal{F}$ which correspond to the arc $\mathcal{A}[\alpha] \subset \partial\mathbb{D}$. Let B_α be the open disc in $\mathbb{C}$ whose boundary contains the geodesic arc $s_\alpha \subset \mathbb{D}$, see Figure 6. We remark that this is uniquely defined since by our choice of the fundamental domain $\mathcal{F}$ no side is a diameter, see Condition (3.1).

More generally, for any admissible word $a_0 \ldots a_n$ let $B_{a_0 \ldots a_n}$ be the open disc in $\mathbb{C}$ such that

$$g_{a_0} \circ \cdots \circ g_{a_{n-1}} (s_{a_n}) \subset \partial B_{a_0 \ldots a_n}.$$

In order to simplify the notation, for any admissible word $a_0 \ldots a_n$, set

$$g_{a_0 \ldots a_n} := g_{a_0} \circ \cdots \circ g_{a_n}.$$

Throughout this section, we consider $g_{a_0 \ldots a_n}$ as an automorphism of the Riemann sphere $\overline{\mathbb{C}}$ (see the beginning of the proof of Lemma 4.1 for the explicit form of $g_{a_0 \ldots a_n}$ of $\mathrm{Aut}(\overline{\mathbb{C}})$). Recalling that g_α is the isometry which sends the side $s_{\overline{\alpha}}$ onto the side s_α, one can see that the map $g_{a_0 \ldots a_n}$ sends $B_{\overline{a_0}}$ on $\mathbb{C} \setminus B_{a_0 \ldots a_n}$. Let $\zeta_{a_0 \ldots a_n} \in \overline{\mathbb{C}}$ be its pole, that is the point such that $g_{a_0 \ldots a_n}(\zeta_{a_0 \ldots a_n}) = \infty$. We observe that $\zeta_{a_0 \ldots a_n} \in B_{\overline{a_0}}$ (since $B_{a_0 \ldots a_n}$ is a disc in the complex plane $\mathbb{C}$, so that $\infty \in \mathbb{C} \setminus B_{a_0 \ldots a_n}$ and the pole $\zeta_{a_0 \ldots a_n}$, which is the preimage of ∞, belongs to $B_{\overline{a_0}}$).

Thus, the restriction to $\mathbb{C} \setminus B_{\overline{a_0}}$ of $g_{a_0 \ldots a_n} \in \mathrm{Aut}(\overline{\mathbb{C}})$ realizes a bijection, that we still denote by

$$g_{a_0 \ldots a_n} : \mathbb{C} \setminus B_{\overline{a_0}} \to B_{a_0 \ldots a_n}.$$

According to the next Lemma 4.1, the restriction map obtained by any admissible word $a_0 \ldots a_n$ has bounded distortion on a subset $M_{\overline{a_0}} \subset \mathbb{C} \setminus B_{\overline{a_0}}$. More precisely, for any $\alpha \in \mathscr{A}$ let $M_\alpha \subset \overline{\mathbb{D}}$ be the closed set (shaded in grey in the example in Figure 6) of those points $\xi \in \overline{\mathbb{D}}$ with

$$\inf\{|\xi - \xi'| : \xi' \in B_\alpha\} \geq \min_{\beta \in \mathscr{A}} \frac{|\mathcal{A}[\beta]|_{\partial\mathbb{D}}}{2}.$$

The set M_α, as shown in Figure 6, has the shape of a *moon* (thanks to its definition and Condition (3.1)).

Lemma 4.1. *There exists a constant $C > 1$, depending only on Γ and on the choice of the fundamental polygon $\mathcal{F}$ for Γ such that the following holds. Given any admissible word $a_0 \ldots a_n$ and ξ_1, ξ_2, ξ_3 points* *in* $\partial\mathbb{D}\cap M_{\overline{a_0}}$ *we have*

$$\frac{1}{C}\frac{|\xi_1-\xi_2|}{|\xi_1-\xi_3|}\leq\frac{|g_{a_0\dots a_n}(\xi_1)-g_{a_0\dots a_n}(\xi_2)|}{|g_{a_0\dots a_n}(\xi_1)-g_{a_0\dots a_n}(\xi_3)|}\leq C\frac{|\xi_1-\xi_2|}{|\xi_1-\xi_3|}.$$

Proof. The automorphism $g_{a_0\dots a_n}\in\mathrm{Aut}(\overline{\mathbb{C}})$ has the form

$$g_{a_0\dots a_n}(\xi)=\frac{a\xi+b}{c\xi+d},\qquad\text{where } a,b,c,d\in\mathbb{C}\text{ and } ad-bc=1.$$

Since $g_{a_0\dots a_n}$ sends the disk $\mathbb{D}$ into itself (so in particular $g_{a_0\dots a_n}\in\mathrm{Aut}(\mathbb{D})$), we also have that $c=\overline{b}$ and $d=\overline{a}$, but we will not make use of this in what follows.

The pole of $g_{a_0\dots a_n}$ is $\zeta_{a_0\dots a_n}=-d/c$. Observe in particular that $c\neq 0$, otherwise $g_{a_0\dots a_n}$ is complex linear, thus for $\xi\in\mathbb{C}$ let us write

$$g_{a_0\dots a_n}(\xi)=\frac{1}{c}\left(a-\frac{1}{c\xi+d}\right).$$

In particular, for any pair of points ξ_1,ξ_2 in $\mathbb{C}$ we have

$$g_{a_0\dots a_n}(\xi_1)-g_{a_0\dots a_n}(\xi_2)=\frac{\xi_1-\xi_2}{(c\xi_2+d)(c\xi_1+d)}=\frac{1}{c^2}\frac{\xi_1-\xi_2}{(\xi_1-\zeta_{a_0\dots a_n})(\xi_2-\zeta_{a_0\dots a_n})}.$$

Hence, for any three points ξ_1,ξ_2,ξ_3 as in the statement we have

$$\frac{|g_{a_0\dots a_n}(\xi_1)-g_{a_0\dots a_n}(\xi_2)|}{|g_{a_0\dots a_n}(\xi_1)-g_{a_0\dots a_n}(\xi_3)|}=\frac{|\xi_3-\zeta_{a_0\dots a_n}|}{|\xi_2-\zeta_{a_0\dots a_n}|}\cdot\frac{|\xi_1-\xi_2|}{|\xi_1-\xi_3|}.$$

As we observed before the statement of the Lemma, the pole $\zeta_{a_0\dots a_n}\in B_{\overline{a_0}}$. Since on the other hand $\xi_i\in M_{\overline{a_0}}$ for $i=1,2,3$, then it follows that

$$|\xi_3-\zeta_{a_0\dots a_n}|,|\xi_2-\zeta_{a_0\dots a_n}|>\min_{\beta\in\mathscr{A}}\frac{|\mathcal{A}[\beta]|_{\partial\mathbb{D}}}{2}.$$

Moreover any B_α is a disc in the complex plane, thus $\mathrm{diam}(B_\alpha)<+\infty$ (remark that it is here that we crucially use Condition (3.1), since it otherwise B_α could have been a semi-plane or the complement of a disk, hence unbounded). Since $\mathrm{diam}(\mathbb{D})=2$ we have also

$$|\xi_3-\zeta_{a_0\dots a_n}|,|\xi_2-\zeta_{a_0\dots a_n}|<2+\mathrm{diam}(B_{\overline{a_0}}).$$

The Lemma follows with $C>0$ defined by

$$C:=\left(2+\max_{\alpha\in\mathscr{A}}\mathrm{diam}(B_\alpha)\right)\cdot\left(\min_{\beta\in\mathscr{A}}\frac{|\mathcal{A}[\beta]|_{\partial\mathbb{D}}}{2}\right)^{-1}.$$ □

4.3. **Size of gaps for the Cantor set in the boundary.** The next Lemma gives the estimate for the size of gaps of level zero. We refer to the notation introduced in Section 4.1 to give a hierarchical description of gaps.

Lemma 4.2. *Fix $\delta>0$. There exists N_0, depending only on δ, on Γ and on the choice of its fundamental domain $\mathcal{F}\subset\mathbb{D}$, such that any $N\geq N_0$ and for gaps of level zero in $\mathbb{B}_N$, we have*

$$|B[\xi_i]|_{\partial\mathbb{D}}\leq\delta,\qquad 0\leq i\leq 2d-1.$$

Proof. Since each level zero gap $B[\xi]$ is contained in the union of two adjacent arcs of level N by Equation (4.2), the Lemma follows directly from the convergence of the Bowen-Series expansion, which implies that finite cuspidal words $a_0\dots a_{n-1}$ satisfy $|\mathcal{A}[a_0,\dots,a_{n-1}]|_{\partial\mathbb{D}}\to 0$ as n tends to infinity. □

Recall that for any $\alpha\in\mathscr{A}$, $B[\xi^r_\alpha]$ and $B[\xi^l_\alpha]$ (in the notation of Section 4.1) are the two gaps of level zero that share an endpoint with the the zero level interval $K[\alpha]$.

Corollary 4.3. *Fix $\varepsilon\in(0,1)$. There exists N_0, depending only on ε, on Γ and on the choice of its fundamental domain $\mathcal{F}\subset\mathbb{D}$, such that for any $N\geq N_0$, the following estimate holds for holes and intervals of level n in the Cantor set $\mathbb{B}_N$. For any $n\in\mathbb{N}$ and any admissible word $a_0,\dots,a_{n-1}$ of length n and any letter $\alpha\in\mathscr{A}$ with $\alpha\neq\overline{a_{n-1}}$, we have*

$$\frac{|B[a_0,\dots,a_{n-1};\xi^r_\alpha]|_{\partial\mathbb{D}}}{|K[a_0,\dots,a_{n-1};\alpha]|_{\partial\mathbb{D}}}\leq 1-\varepsilon\quad\textit{and}\quad\frac{|B[a_0,\dots,a_{n-1};\xi^l_\alpha]|_{\partial\mathbb{D}}}{|K[a_0,\dots,a_{n-1};\alpha]|_{\partial\mathbb{D}}}\leq 1-\varepsilon.$$

Proof. By definition of holes and intervals of level n (we refer to Section 4.1),

$$\frac{|B[a_0,\dots,a_{n-1};\xi_\alpha^r]|_{\partial\mathbb{D}}}{|K[a_0,\dots,a_{n-1};\alpha]|_{\partial\mathbb{D}}}=\frac{|g_{a_0}\circ\cdots\circ g_{a_{n-1}}\left(B[\xi_\alpha^r]\right)|_{\partial\mathbb{D}}}{|g_{a_0}\circ\cdots\circ g_{a_{n-1}}\left(K[\alpha]\right)|_{\partial\mathbb{D}}}\tag{4.3}$$

and the same expression holds with ξ_α^r replaced by ξ_α^l.

Let us remark now that if $\xi,\xi'\in\partial\mathbb{D}$ bound an arc $\mathcal{A}[\xi,\xi']\subset\partial\mathbb{D}$ of length strictly less than π, then $|\cdot|$ and $|\cdot|_{\partial\mathbb{D}}$ are comparable, i.e.

$$|\xi-\xi'|\le|\mathcal{A}[\xi,\xi']|_{\partial\mathbb{D}}\le\pi\,|\xi-\xi'|\,.$$

Thus, since each hole or gap is contained in $\mathcal{A}[\alpha]$ for some $\alpha\in\mathscr{A}$ and $|\mathcal{A}[\alpha]|_{\partial\mathbb{D}}<\pi$ (see assumption (4.2)), we can apply this remark to (4.3). The proof hence follows from Lemma 4.2 and Lemma 4.1. □

4.4. **Sum of Cantor sets on the real line.** In order to apply results on the sum of Cantor sets, we now consider the image in $\mathbb{R}$ of the Cantor set $\mathbb{B}_N=\mathbb{B}_N^\eta$ (defined in Section 3.2 and described in Section 4.1). Following Section 3.2, let $\mathbb{K}_N=\mathbb{K}_N^\eta:=\varphi(\mathbb{B}_N)$ be its image in $\mathbb{R}$ under the map $\varphi\colon\mathbb{D}\to\mathbb{H}$, where φ is the inverse of the Caley map $\mathscr{C}$ defined in (2.1). Explicitly, $\mathbb{K}_N$ is the set of points $x=[a_0,\dots,a_n,\dots]_{\partial\mathbb{H}}$ corresponding to no-backtracking cutting sequences $(a_n)_{n\in\mathbb{N}}$ not containing any cuspidal word of length $N+1$ and whose first letter satisfies $a_0\notin\{\eta,\overline{\eta}\}$. Remark that this implies in particular that $\mathbb{K}_N$ is contained in $[-\mu/2,\mu/2]$ (see Lemma 3.6).

Let us write for, respectively, minimum, maximum and translates of $\mathbb{K}_N$ by $z\mapsto z+s\mu$, where $s\in\mathbb{Z}$:

$$m_N=\min\mathbb{K}_N,\qquad M_N=\max\mathbb{K}_N,\qquad \mathbb{K}_N^s:=\mathbb{K}_N+s\mu,s\in\mathbb{Z}.$$

We claim that, for any integer s, the Cantor sets $\mathbb{K}_N$ and $\mathbb{K}_N^s$ satisfy the assumptions of the Stable Hall Theorem 1.12, namely the ε-stable gap condition and of the ε-size condition which were defined in Section 1.8.

Lemma 4.4. *There exists an integer $N_0\ge0$ such that for any $N\ge N_0$ and any integer s:*

(1) *the Cantor sets $\mathbb{K}_N$, $-\mathbb{K}_N$ and $\mathbb{K}_N^s$ satisfy the ε-stable gap condition;*
(2) *the pairs $(\mathbb{K}_N,\mathbb{K}_N^s)$ and $(-\mathbb{K}_N,\mathbb{K}_N^s)$ satisfy the ε-size condition.*

Proof. We claim that it is enough to show that there exists $N_0\in\mathbb{N}$ such that for any $N\ge N_0$ the Cantor set $\mathbb{K}_N$ satisfies the ε-stable gap condition. This obviously implies that the same holds also for any of its translated image $\mathbb{K}_N^s$ and also for its reflection $-\mathbb{K}_N$ (since reflecting only inverts the role of left and right intervals K^L and K^R in the definition ε-size condition (1.12)). It is also clear that, for N large enough, the pairs $(\pm\mathbb{K}_N,\mathbb{K}_N^s)$ satisfy the ε-size condition for any s, indeed $|\mathbb{K}_N|\to\mu$ for $N\to\infty$, while the size of the holes shrinks to zero, and the same holds for $-\mathbb{K}_N$ and the translates $\mathbb{K}_N^s$.

Let us hence prove that $\mathbb{K}_N$ satisfies the ε-stable gap condition if N is sufficiently large. Recall that, by the choices made in Section 3.1, the sides s_η and $s_{\overline{\eta}}$ share ξ_0, or, more precisely $\xi_\eta^l=\xi_0=\xi_{\overline{\eta}}^r$ (see Lemma 3.1) and the inverse $\varphi\colon\mathbb{D}\to\mathbb{H}$ of the Cayley map (2.1) is such that

$$\varphi(\xi_0),=\infty\qquad\varphi(\xi_\eta^r)=\frac{\mu}{2}\qquad\text{and}\qquad\varphi(\xi_{\overline{\eta}}^l))=-\frac{\mu}{2}.$$

Hence, the arc $\mathcal{A}[\eta]\cup\mathcal{A}[\overline{\eta}]\subset\partial\mathbb{D}$ is the arc with endpoints $\xi_{\overline{\eta}}^l$ and ξ_η^r which contains in its interior the point $\xi_0=\varphi^{-1}(\infty)$, which is the pole of φ. It follows that there is a constant $\kappa>0$, depending only on the choice of the fundamental domain $\mathcal{F}$ for Γ, such that $|\xi-\varphi^{-1}(\infty)|_{\partial\mathbb{D}}\ge\kappa$ for any $\xi\in\partial\mathbb{D}\setminus\mathcal{A}[\eta]\cup\mathcal{A}[\overline{\eta}]$, and thus the restricted map

$$\varphi\colon\partial\mathbb{D}\setminus(\mathcal{A}[\eta]\cup\mathcal{A}[\overline{\eta}])\to\mathbb{R}$$

has bounded derivative. Since by definition $\mathbb{K}_N\subset\partial\mathbb{D}\setminus(\mathcal{A}[\eta]\cup\mathcal{A}[\overline{\eta}])$, combining the control on the derivative of φ and the estimate in Corollary 4.3, the ε-stable gap condition for $\mathbb{K}_N$ follows. □

The above results together with the Stable Hall Theorem stated in the introduction (and proved in Section 7) allow to conclude the proof of Proposition 3.2 (and hence Theorem 1.4).

Proof of Proposition 3.2. Let us check that, for N sufficiently large, we can apply the Stable Hall theorem to the Cantor sets $\mathbb{K}_N^s$ and $-\mathbb{K}_N$, in the special case in which $S=S_0$ is the sum function and $U=\mathbb{R}^2$.

This is the case, since the Lipschitz norm condition (1.14) is trivially satisfied ($S = S_0$) and the ε-stable gap and ε-size conditions are proved in Lemma 4.4 for $N \geq N_0$. The Stable Hall theorem then gives

$$\begin{aligned}\mathbb{K}_N^s - \mathbb{K}_N &= S_0\big([\min \mathbb{K}_N^s, \max \mathbb{K}_N^s] \times [\min(-\mathbb{K}_N), \max(-\mathbb{K}_N)]\big)\\ &= \big[\min \mathbb{K}_N^s + \min(-\mathbb{K}_N), \max \mathbb{K}_N^s + \max(-\mathbb{K}_N)\big]\\ &= \big[(m_N + s\mu) + (-M_N), (M_N + s\mu) + (-m_N)\big],\end{aligned}$$

which is the desired expression. The form of $\mathbb{K}_N^s + \mathbb{K}_N$ follows analogously. Thus, $|\mathbb{K}_N^s \pm \mathbb{K}_N| = 2(M_N - m_N)$. Since, as $N \to \infty$, $M_N \to \mu/2$ and $m_N \to -\mu/2$ (and hence $M_N - m_N \to \mu$), it is enough to increase N_0 to ensure that $M_N - m_M > \mu/2$ to have also that $|\mathbb{K}_N^s \pm \mathbb{K}_N| \geq \mu$. □

5. Penetration estimates

In this section we bound the height of a geodesic knowing that the cuspidal words of a piece of its cutting sequence are not too long. Let $\gamma\colon \mathbb{R} \to \mathbb{D}$ be a geodesic with cutting sequence $(a_n)_{n\in\mathbb{Z}}$ and let $(t_n)_{n\in\mathbb{Z}}$ be the sequence of times when γ crosses a side of the tessellation of $\mathbb{D}$ given by $\mathcal{F}$, as defined in (2.6). For any $r \in \mathbb{Z}$, the cuspidal words C_r and the integers $n(r)$ such that $C_r := a_{n(r)}, \dots, a_{n(r+1)-1}$ are also defined in Section 2.3.

Lemma 5.1. *For any positive integer $N \geq 1$ there exists a compact $\mathcal{K}_N \subset \mathbb{D}$ such that the following holds. Consider any geodesic $\gamma\colon \mathbb{R} \to \mathbb{D}$ with $\gamma(0) \in \mathcal{F}$. Let $(a_n)_{n\in\mathbb{Z}}$ be its cutting sequence and $(C_r)_{r\in\mathbb{Z}}$ the corresponding decomposition into cuspidal words. Then for any positive integer $r \geq 1$ such that*

$$\ell(C_{r-1}) \leq N, \qquad \ell(C_r) \leq N, \qquad \ell(C_{r+1}) \leq N,$$

and any t with $t_{n(r)} \leq t \leq t_{n(r+1)}$, we have

$$g_{a_{n(r)-1}}^{-1} \circ \cdots \circ g_{a_0}^{-1}\big(\gamma(t)\big) \in \mathcal{K}_N.$$

Proof. For any ideal vertex ξ_i, $i = 0, \dots, 2d-1$ of the fundamental domain $\mathcal{F}$, let $\xi_i^{(-)}$ and $\xi_i^{(+)}$ be the two points in $\partial\mathbb{D}$ at distance $\delta_N > 0$ from ξ_i, where we set

$$\delta_N := \min\{|\mathcal{A}[b_0, \dots, b_N]|_{\partial\mathbb{D}}, b_0 \dots b_N \text{ cuspidal word of length } N+1\ \},$$

and where we recall that $|\mathcal{A}[b_0, \dots, b_N]|_{\partial\mathbb{D}}$ denotes the length of the arc in $\partial\mathbb{D}$. Consider the open disc $\mathbb{D} \subset \mathbb{C}$ as embedded in the complex plane. For any ideal vertex ξ_i, $i = 0, \dots, 2d-1$ of the fundamental domain $\mathcal{F}$, let $B(N, \xi_i) \subset \mathbb{C}$ be the open Euclidean ball whose boundary $\partial B(N, \xi_i)$ intersects $\partial\mathbb{D}$ orthogonally at $\xi_i^{(-)}$ and $\xi_i^{(+)}$. Observe that $\mathbb{D} \setminus B(N, \xi_i)$ is an *hyperbolic half-space*, that is the region of $\mathbb{D}$ delimited by its boundary and a complete geodesics. In particular $\mathbb{D} \setminus B(N, \xi_i)$ is *hyperbolic convex*, that is it contains the entire segment of hyperbolic geodesic connecting any two of its endpoints. Define a compact set $F_N \subset \mathbb{D}$ by

$$F_N := \overline{\mathcal{F}} \setminus \bigcup_{i=0}^{2d-1} B(N, \xi_i),$$

that is the subset of the closure of $\mathcal{F}$ which do not intersects the open balls $B(N, \xi_i)$ defined above, see Figure 7. Since the set F_N is an intersection of hyperbolic half-spaces, then it is also hyperbolic convex. Let $\widetilde{\mathcal{K}}_N$ be the set defined by

$$\widetilde{\mathcal{K}}_N := \bigcup_{b_0 \dots b_{n-1}} g_{b_0} \circ \cdots \circ g_{b_{n-1}} F_N,$$

where $b_0 \dots b_{n-1}$ varies among all cuspidal words with $n \leq N$. In particular, considering the trivial word, it is evident that $F_N \subset \widetilde{\mathcal{K}}_N$. The set $\widetilde{\mathcal{K}}_N$ is compact, since it is the finite union of images of the compact F_N under the continuous maps $g_{b_0} \circ \cdots \circ g_{b_{n-1}}$. On the other hand, it is possible to see that $\widetilde{\mathcal{K}}_N$ is not hyperbolic convex. Thus we define $\mathcal{K}_N$ as the hyperbolic convex hull of $\widetilde{\mathcal{K}}_N$, so that $\mathcal{K}_N$ is hyperbolic convex by definition and it is also compact, since $\widetilde{\mathcal{K}}_N$ is compact.

Fix an integer r as in the statement. Since $\mathcal{K}_N$ is hyperbolic convex, it is enough to prove the statement for $t = t_{n(r)}$ and $t = t_{n(r+1)}$. Consider the normalized geodesics $\gamma_r(t)$ and $\gamma_{r+1}(t)$ in $\mathbb{D}$ (which are simply the $n(r)^{\text{th}}$ and $n(r+1)^{\text{th}}$ normalized geodesics as defined in Lemma 3.5) given by:

$$\begin{aligned}\gamma_r(t) &= g_{a_{n(r)-1}}^{-1} \circ \cdots \circ g_{a_0}^{-1}(\gamma(t)),\\ \gamma_{r+1}(t) &= g_{a_{n(r+1)-1}}^{-1} \circ \cdots \circ g_{a_0}^{-1}(\gamma(t)).\end{aligned}$$

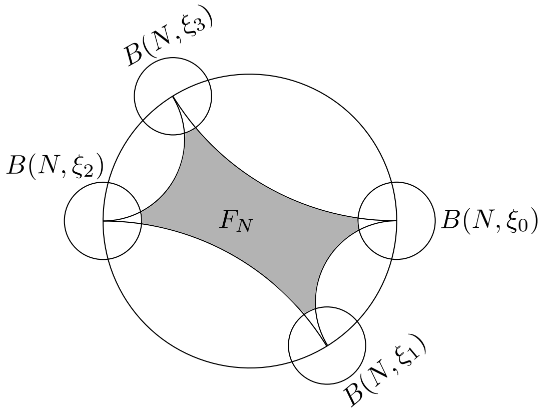

FIGURE 7. A schematic picture of the set F_n, in dark grey inside $\mathcal{F}$.

Using these geodesics, the statement is equivalent to the two conditions

$$\begin{aligned}&\gamma_r(t_{n(r)}) \in \mathcal{K}_N,\\ &\gamma_r(t_{n(r+1)}) = g_{a_{n(r)}} \circ \cdots \circ g_{a_{n(r+1)-1}}(\gamma_{r+1}(t_{n(r+1)})) \in \mathcal{K}_N.\end{aligned}$$

Recall that hyperbolic geodesics intersect in at most one point. Observe also that we have

$$\begin{aligned}\gamma_r(+\infty) &= [a_{n(r)}, a_{n(r)+1}, \ldots]_{\partial\mathbb{D}}, & \gamma_{r+1}(+\infty) &= [a_{n(r+1)}, a_{n(r+1)+1}, \ldots]_{\partial\mathbb{D}},\\ \gamma_r(-\infty) &= [\overline{a_{n(r)-1}}, \overline{a_{n(r)-2}}, \ldots]_{\partial\mathbb{D}}, & \gamma_{r+1}(-\infty) &= [\overline{a_{n(r+1)-1}}, \overline{a_{n(r+1)-2}}, \ldots]_{\partial\mathbb{D}}.\end{aligned}$$

In order to prove $\gamma_r(t_{n(r)}) \in \mathcal{K}_N$ we prove the stronger condition $\gamma_r(t_{n(r)}) \in F_N$. If the latter does not hold, then there is some i for which we either have $\gamma_r(+\infty) \in B(N,\xi_i) \cap \partial\mathbb{D}$ or $\gamma_r(-\infty) \in B(N,\xi_i) \cap \partial\mathbb{D}$. However the first condition would imply that $a_{n(r)} \ldots a_{n(r+1)-1}$ is a cuspidal word of length greater than $N+1$. Similarly, the second condition would imply that $\overline{a_{n(r)-1}} \ldots \overline{a_{n(r-1)}}$ is a cuspidal word of length $N+1$, which is equivalent to the same condition for $a_{n(r-1)} \ldots a_{n(r)-1}$.

In order to prove $\gamma_r(t_{n(r+1)}) \in \mathcal{K}_N$, we prove the stronger condition $\gamma_r(t_{n(r+1)}) \in \widetilde{\mathcal{K}}_N$. To do so, observe that the same argument as above applies to the words $a_{n(r+1)} \ldots a_{n(r+2)-1}$ and $a_{n(r)} \ldots a_{n(r+1)-1}$ and implies $\gamma_{r+1}(t_{n(r+1)}) \in F_N$, that is

$$\gamma_r(t_{n(r+1)}) = g_{a_{n(r)}} \circ \cdots \circ g_{a_{n(r+1)-1}}(\gamma_{r+1}(t_{n(r+1)})) \in g_{a_{n(r)}} \circ \cdots \circ g_{a_{n(r+1)-1}} F_N \subset \widetilde{\mathcal{K}}_N. \qquad \square$$

6. Stable Hall rays for proper functions

In this section we prove Theorem 1.8.

Let G be a Fuchsian group that is a non uniform lattice and such that ∞ is a parabolic fixed point of G. As in Section 3.1, we also let $\Gamma < G$ be its maximal finite index normal subgroup without elliptic elements and we choose a fundamental domain $\mathcal{F}$ for Γ that satisfies the conclusions of Lemma 3.1 and label by $\alpha \in \mathscr{A}$ its sides as in Section 2. Every geodesic boundary expansion $(a_n)_{n\in\mathbb{Z}}$ in this section is with respect to this fundamental domain and is such that $a_n \in \mathscr{A}$.

Finally, let us remark that any $h\colon \mathbb{H} \to \mathbb{R}$ which is G-invariant is in particular Γ-invariant, since $\Gamma < G$ Throughout this section, we will only use Γ-invariance.

6.1. **Naive height as a function of endpoints.**

Let $h\colon \mathbb{H} \to \mathbb{R}_+$ be a Γ-invariant continuous function such that the induced function $h\colon \Gamma\backslash\mathbb{H} \to \mathbb{R}_+$ is proper. Given $l > 0$ define the function $h_l\colon \mathbb{H} \to \mathbb{R}_+$ by

$$h_l(z) = \begin{cases} h(z), & \text{if } \operatorname{Im}(z) > l,\\ 0, & \text{if } \operatorname{Im}(z) \le l.\end{cases}$$

Recall that hyperbolic geodesics are uniquely determined by their endpoints on $\mathbb{R}$ and that, for $x_1, x_2 \in \mathbb{R}$ with $x_1 \neq x_2$ we denote by $\gamma(x_1,x_2)$ the unique geodesic $\gamma(t)$ with $\gamma(-\infty) = x_1$ and $\gamma(+\infty) = x_2$. Denoting by

$$\Delta := \{(x_1,x_2) \in \mathbb{R}^2 : x_1 = x_2\}$$

the diagonal in $\mathbb{R}^2$, we define the function $H\colon \mathbb{R}^2 \setminus \Delta \to \mathbb{R}_+$ by

$$H(x_1, x_2) := \sup_{t\in\mathbb{R}} \{h_l(\gamma(t)),\ \text{where } \gamma = \gamma(x_1, x_2)\}. \tag{6.1}$$

In parallel, consider also the function $H_0\colon \mathbb{R}^2 \setminus \Delta \to \mathbb{R}_+$ given by

$$H_0(x_1, x_2) := \frac{|x_2 - x_1|}{2} = \sup_{t\in\mathbb{R}} \{\mathrm{Im}(\gamma(t)),\ \text{where } \gamma = \gamma(x_1, x_2)\}.$$

Finally, for any $l > 0$ let $U_l \subset \mathbb{R}^2$ be the complement of the $2l$-neighborhood of Δ defined as the set $U_l = \{(x_1, x_2) \in \mathbb{R}^2 : |x_2 - x_1| \geq 2l\}$.

Remark 6.1. Observe that, for any $(x_1, x_2) \in \mathbb{R}^2$, we have $\gamma(x_1, x_2) \cap \mathcal{U}_l \neq \varnothing$ if and only if $(x_1, x_2) \in U_l$.

The regularity of $H\colon \mathbb{R}^2 \to \mathbb{R}_+$ depends on the regularity of $h\colon \mathbb{H} \to \mathbb{R}_+$ via Lemma 6.2 and Proposition 6.3 below. The proof of Lemma 6.2 is an easy estimate and it is left to the reader; Proposition 6.3 is proved in Appendix B.

Lemma 6.2. *For any $l > 0$, we have*

$$\|(H - H_0)|_{U_l}\|_\infty \leq \|(h - \mathrm{Im})|_{\mathcal{U}_l}\|_\infty.$$

The definition of Lipschitz constant $\mathrm{Lip}(\cdot)$ for $H\colon \mathbb{R}^2 \to \mathbb{R}$ corresponds to Equation (1.8) under the standard identification between $\mathbb{C}$ and $\mathbb{R}^2$.

Proposition 6.3. *For any $l > 0$ we have*

$$\mathrm{Lip}\big((H - H_0)|_{U_l}\big) \leq \left(\sqrt{2} + \frac{\sqrt{2}}{l}\right) \cdot \|(h - \mathrm{Im})|_{\mathcal{U}_l}\|_{Lip}.$$

In particular the function $H\colon \mathbb{R}^2 \setminus \Delta \to \mathbb{R}_+$ is continuous.

Remark 6.4. We observe that in general the function $H\colon \mathbb{R}^2 \to \mathbb{R}_+$ is not differentiable even if $h\colon \mathbb{H} \to \mathbb{R}_+$ is. For example, for $h(x, y) := \sqrt{x^2 + y^2}$ one gets

$$H(x_1, x_2) = \max\{|x_1|, |x_2|\}.$$

This is the reason why it is natural to consider the Lipschitz category rather than the category of smooth functions.

6.2. **Intervals from endpoints in Cantor sets.** Let $\mathbb{B}_N = \mathbb{B}_N^\eta$ be the Cantor set defined in Section 3.2 and described combinatorially in Section 4.1. Following Section 3.2, let $\mathbb{K}_N = \mathbb{K}_N^\eta := \varphi(\mathbb{B}_N)$ be its image in $\mathbb{R}$ under the map $\varphi\colon \partial\mathbb{D} \to \partial\mathbb{H}$, where φ is as usual the inverse of the Caley map $\mathscr{C}$ defined in (2.1).

Proposition 6.5. *If there exist $0 < \varepsilon < 1$, $l > 0$ and a function $H\colon \mathbb{R}^2 \to \mathbb{R}$ such that*

$$\frac{1 - 2 \cdot \mathrm{Lip}((H - H_0)|_{U_l})}{1 + 2 \cdot \mathrm{Lip}((H - H_0)|_{U_l})} > 1 - \varepsilon, \tag{6.2}$$

then there exist natural numbers N_0 and s_0 such that, for any $N \geq N_0$ and $s \geq s_0$, $H(\mathbb{K}_N \times \mathbb{K}_N^s)$ is an interval. More precisely, we have

$$|\mathbb{K}_N| = M_N - m_N > \frac{\mu}{2} \qquad \text{and} \qquad H(\mathbb{K}_N \times \mathbb{K}_N^s) = H([m_N, M_N] \times [m_N + s\mu, M_N + s\mu]),$$

where $m_N = \min \mathbb{K}_N$ and $M_N = \max \mathbb{K}_N$.

The proof is simply a reduction of the statement to an application of the Stable Hall Theorem 1.12. We remark that, in order for (6.2) to be satisfied, we must have $\mathrm{Lip}((H - H_0)|_{U_l}) < 1/2$.

Proof. Let us apply a change of variable that allows to reduce the function H_0 (restricted to the set $\{x_2 > x_1\}$) to the sum function S_0. Let

$$(y_1, y_2) = \psi(x_1, x_2) := \left(-\frac{x_1}{2}, \frac{x_2}{2}\right),$$

so that, if $x_2 > x_1$,

$$S_0(y_1, y_2) = -\frac{x_1}{2} + \frac{x_2}{2} = \frac{|x_2 - x_1|}{2} = H_0(x_1, x_2).$$

Then define the function $S\colon \mathbb{R}^2 \to \mathbb{R}$ (which we will also use only on $\{x_2 > x_1\}$) by

$$S(y_1, y_2) := H(\psi^{-1}(y_1, y_2)) = H(-2y_1, 2y_2).$$

Choose $s_0 \geq 1$ sufficiently large so that $\mathbb{K}_N \times \mathbb{K}_N^s \subset U_l$ for any $s \geq s_0$. Fix now any $s \geq s_0$ and consider the Cantor sets $\mathbb{K}' := -(1/2)\mathbb{K}_N$ and $\mathbb{F}' := (1/2)\mathbb{K}_N^s$, so that

$$\mathbb{K}' \times \mathbb{F}' = \psi\left(\mathbb{K}_N \times \mathbb{K}_N^s\right).$$

We showed in Lemma 4.4 that there exists $N_0 \in \mathbb{N}$ such that, for any $N \geq N_0$, $\mathbb{K}_N$ and $\mathbb{K}_N^s$ each satisfy the ε-stable gap condition introduced in Section 1.8 and jointly satisfy as a pair $(\mathbb{K}_N, \mathbb{K}_N^s)$ the ε-size condition (also defined in Section 1.8). It is clear from the definitions that this implies that the Cantor sets $\mathbb{K}'$ and $\mathbb{F}'$, which are images by affine maps of $\mathbb{K}_N$ and $\mathbb{K}_N^s$ respectively, also satisfy such conditions.

Let $U' = \psi(U_l)$ be the image of $U_l \subset \mathbb{R}^2$ (defined in Section 6.1). Since $\mathbb{K}_N \times \mathbb{K}_N^s \subset U_l$ (we fixed $s \geq s_0$), $\mathbb{K}' \times \mathbb{F}' \subset U'$. Moreover, observe that we have $\mathrm{Lip}(S - S_0) = 2 \cdot \mathrm{Lip}(H - H_0)$, so that

$$\frac{1 - \mathrm{Lip}(S - S_0)}{1 + \mathrm{Lip}(S - S_0)} = \frac{1 - 2 \cdot \mathrm{Lip}(H - H_0)}{1 + 2 \cdot \mathrm{Lip}(H - H_0)}.$$

Hence, from the assumption (6.2) on the Lipschitz constant of $(H - H_0)$ restricted to U_l, it follows that $S - S_0$ satisfies the Lipschitz constant assumption (1.14) of Theorem 1.12 on U'.

Thus, we can apply Theorem 1.12. Let us now rephrase its conclusion in terms of H. Notice that for $s_0 \geq 1$, $\min \mathbb{K}_N^s > \max \mathbb{K}_N$, so that $(\mathbb{K}_N, \mathbb{K}_N^s) \subset \{x_2 > x_1\}$. Thus, on the set $\mathbb{K}_N \times \mathbb{K}_N^s$ we have, as seen above, that $S \circ \psi = H$. This gives that

$$S(\mathbb{K}' \times \mathbb{F}') = S \circ \psi(\mathbb{K}_N \times \mathbb{K}_N^s) = H(\mathbb{K}_N \times \mathbb{K}_N^s),$$

and, similarly, that

$$S\big([\min \mathbb{K}', \max \mathbb{K}'] \times [\min \mathbb{F}', \max \mathbb{F}']\big) = H\big([\min \mathbb{K}_N, \max \mathbb{K}_N] \times [\min \mathbb{K}_N^s, \max \mathbb{K}_N^s]\big).$$

This in particular implies that $H(\mathbb{K}_N \times \mathbb{K}_N^s)$ is an interval (see Remark 1.13) and concludes the proof. □

The following Corollary gives the starting point for the existence of a Hall ray. The reader should compare this Corollary (and its proof) with the simpler analogue Corollary 3.3.

Corollary 6.6. *If $H\colon \mathbb{R}^2 \to \mathbb{R}$ is such that there exist $0 < \varepsilon < 1$, $l > 0$ for which (6.2) holds and $\|H - H_0\|_\infty < 1/4$, up to increasing N_0, for any $N \geq N_0$, any $s_1 \geq s_0$ and any*

$$L \geq \inf H(\mathbb{K}_N \times \mathbb{K}_N^{s_1})$$

there exist points $x_1 = [a_0, a_1, \dots]_{\partial\mathbb{H}} \in \mathbb{K}_N$ and $x_2 = [b_0, b_1, \dots]_{\partial\mathbb{H}} \in \mathbb{K}_N$ and an integer $s \geq s_1$ such that

$$L = H(x_1, x_2 + s\mu).$$

Proof. Recall that we are assuming that the Margulis neighborhood starts at $m = 1$, which implies that $\mu \geq 1$. Hence, since $\|H - H_0\|_\infty < 1/4$ and $M_N - m_N \to \mu$ as N tends to infinity, increasing N if needed, we can assume that we have the following *overlapping condition*

$$M_N - m_N = |\mathbb{K}_N| \geq \frac{\mu}{2} + 2\|H - H_0\|_\infty.$$

We will now show that this condition implies that for any $s \geq s_0$ we have

$$H(\mathbb{K}_N \times \mathbb{K}_N^s) \cap H(\mathbb{K}_N \times \mathbb{K}_N^{s+1}) \neq \varnothing. \tag{6.3}$$

Since by Proposition 6.5 we know that both $H(\mathbb{K}_N \times \mathbb{K}_N^s)$ and $H(\mathbb{K}_N \times \mathbb{K}_N^{s+1})$ are intervals, this implies that they overlap and that there is no gap between them. This is then enough to conclude, since, remarking also that $\sup H(\mathbb{K}_N \times \mathbb{K}_N^s)$ tends to $+\infty$ as s grows, it implies that for any $s_1 \geq s_0$,

$$\bigcup_{s \geq s_1} H(\mathbb{K}_N \times \mathbb{K}_N^s) \supset \left(\inf H(\mathbb{K}_N \times \mathbb{K}_N^{s_1}), +\infty\right),$$

so any $L \geq \inf H(\mathbb{K}_N \times \mathbb{K}_N^{s_1})$ belongs to $H(\mathbb{K}_N \times \mathbb{K}_N^s)$ for some $s \geq s_1$ and hence can be written as $H(x_1, x_2 + s\mu)$ for some $x_1 \in \mathbb{K}_N$ and $x_2 + s\mu \in K_N^s$.

It remains to show (6.3). On one hand we have that, for $s \geq s_0$,

$$\begin{aligned}\sup H(\mathbb{K}_N \times \mathbb{K}_N^s) \geq H(m_N, M_N + s\mu) &> H_0(m_N, M_N + s\mu) - \|H - H_0\|_\infty \\ &= \frac{M_N + s\mu - m_N}{2} - \|H - H_0\|_\infty;\end{aligned}$$

while, for $s+1$, we have

$$\begin{aligned}\inf H(\mathbb{K}_N \times \mathbb{K}_N^{s+1}) \leq H(M_N, m_N+(s+1)\mu) &< H_0(M_N, m_N+(s+1)\mu) + \|H-H_0\|_\infty \\ &= \frac{m_N+(s+1)\mu-M_N}{2} + \|H-H_0\|_\infty,\end{aligned}$$

where to remove the absolute value in H_0 we used that $m_N+(s+1)\mu \geq M_N$ for any $s \geq 0$. Since one can check that the overlap condition implies that

$$\frac{M_N+s\mu-m_N}{2} - \|H-H_0\|_\infty \geq \frac{m_N+(s+1)\mu-M_N}{2} + \|H-H_0\|_\infty,$$

the combination of the last three inequalities shows that $\sup H(\mathbb{K}_N \times \mathbb{K}_N^s) > \inf H(\mathbb{K}_N \times \mathbb{K}_N^{s+1})$ and hence (6.3) holds. As explained above, this concludes the proof. □

6.3. **A Perron-like formula to produce values in the Hall ray.** The next result is the key step for the proof of Theorem 1.8 on the existence of a Hall ray for proper functions. It provides a formula which will allows us to verify that certain geodesics realize values of the Lagrange spectrum. The formula resembles the generalized Perron formula in Section 3.3 (see (3.3)). However, since we have the additional difficulty of controlling the values of the proper function h in the other cusps, we can only prove it for sequences of a special form, which we will use to prove the existence of the Hall ray in Section 6.4.

Let $h\colon \mathbb{H} \to \mathbb{R}_+$ be a function satisfying the assumptions of Theorem 1.8 and let H denote the function defined in Equation (6.1). Recall also that we are assuming that the Γ contains the parabolic element $p = \left(\begin{smallmatrix} 1 & \mu \\ 0 & 1 \end{smallmatrix}\right)$ which acts on $\mathbb{H}$ as $p(z) = z+\mu$ and that $p = g_\eta$ for some $\eta \in \mathscr{A}$.

Proposition 6.7. *For any integer $N \geq 1$ there exists an integer $M = M(l_0, h, N) \geq N$ such that the following holds. Let $\gamma\colon \mathbb{R} \to \mathbb{H}$ be a hyperbolic geodesic and let $(a_n)_{n\in\mathbb{Z}}$ be its cutting sequence. Assume that the cuspidal words $(C_r)_{r\in\mathbb{N}}$ in the cuspidal decomposition of the positive half sequence $(a_n)_{n\in\mathbb{N}}$ satisfy the properties below.*

1. *There exists an increasing subsequence $(r_k)_{k\in\mathbb{N}}$ such that, for any $k \in \mathbb{N}$, C_{r_k} is a cuspidal word obtained concatenating $M \geq M_0$ times the letter η corresponding to the parabolic element $p = g_\eta$, i.e.*
$$C_{r_k} = \eta^M = \underbrace{\eta \dots \eta}_{M \text{ times}}, \qquad M \geq M_0.$$
Moreover, we eventually have $r_k - r_{k-1} > 3$.
2. *For any $r \neq r_k$, $k \in \mathbb{Z}$, we have $\ell(C_r) \leq N$.*

Then, we have

$$\limsup_{t\to+\infty} h(\gamma) = \limsup_{k\to\infty} H([a_{n(r_k)-1}, a_{n(r_k)-2}, \dots]^-_{\partial\mathbb{H}}, [a_{n(r_k)}, a_{n(r_k)+1}, \dots]_{\partial\mathbb{H}},)$$

where the notation $[\,\cdot\,]^-_{\partial\mathbb{H}}$ was defined in (3.2).

The latter formula, in the special case of $h = \mathrm{Im}$ and $H_0(x,y) = |x-y|/2$, should be compared to the generalization (3.3) of Perron's formula (1.11).

Before starting the proof of the Proposition, let us explain the strategy of the proof.

Outline of the Proof of Proposition 6.7. Let $(t_n)_{n\in\mathbb{Z}}$ be the sequence of hitting times of γ with sides of the ideal tessellation of $\mathbb{H}$ induced by $\mathcal{F}$ (see (2.6)) and let $\gamma([t_{n(r)}, t_{n(r+1)}))$ be the segment of γ encoded by the r^{th} cuspidal word $C_r = a_{n(r)}, \dots, a_{n(r+1)-1}$ (we refer to Section 2.3 for definitions).

We will split the integers r indexing cuspidal words into two groups and consider two different scenarios. First, consider any r which is not equal to any of r_k, r_k-1 or r_k+1, for some $k \in \mathbb{N}$. We will call these indexes r *intermediate times*. For these r, the parabolic word C_r, as well as the preceding and following parabolic words C_{r-1} and C_{r+1} all have by assumption length at most N. Thus, using Lemma 5.1, we will establish for any such r a uniform bound $C_1(N,h) > 0$ for the supremum of h along the r^{th} segment of γ, namely we will show (in *Step* 1 of the proof) that

$$\sup_{t_{n(r)} \leq t < t_{n(r+1)}} h\big(\gamma(t)\big) \leq C_1(N,h), \qquad \forall\, r \notin \cup_{k\in\mathbb{Z}}\{r_k-1, r_k, r_k+1\}.$$

Then, we will consider the parabolic words C_{r_k}, coupled together with the precedent and following parabolic words, respectively C_{r_k-1} and C_{r_k+1}, and consider the segments of γ corresponding to the triple $C_{r_{k-1}}C_{r_k}C_{r_{k+1}}$, that we will denote $\gamma^{(k)}$, in other words we set

$$\gamma^{(k)} := \{\gamma(t),\ t_{n(r_k-1)} \leq t \leq t_{n(r_k+2)}\}, \qquad k \in \mathbb{N}.$$

We show (in *Step* 2 and *Step* 3 of the proof) that the supremum of h along $\gamma^{(k)}$ is bigger than $C_1(N,h)$. In order to do this, since the function h is *proper* on $X = \Gamma\backslash\mathbb{H}$, so that $h(z)$ diverges as z get closer to $\partial\mathbb{H} = \overline{\mathbb{R}}$, we first need to establish a lower bound on $\mathrm{Im}(z)$ for $z \in \gamma^{(k)}$ (this is done in *Step* 2). This then allows to guarantee that, for sufficiently large excursions into the cusp at infinity (i.e. when M is sufficiently large), the supremum of h along $\gamma^{(k)}$ is taken for z in the central part of the segment.

Finally, in *Step* 3, we will show that, when we *normalize* the geodesic segment $\gamma^{(k)}$ to bring it back to the fundamental domain (so that it crosses the center of the disk in the time interval $[t_{n(r_k-1)}, t_{n(r_k+2)}]$), then the maximum of h is taken inside the region $\mathcal{U}_l$. The Proposition then follows because this last property enables to express the supremum of h along the central segment as the value of $H(\cdot,\cdot)$ at the two endpoints of the renormalized geodesic, which leads to the desired formula.

Proof of Proposition 6.7. We begin by introducing some auxiliary notation for the proof. Recall that $\gamma(0) \in \varphi(\mathcal{F})$. For any $r \geq 1$ it is convenient to introduce the group element associated to the r^{th} word C_r in the parabolic decomposition, i.e.

$$G_r := g_{a_{n(r)}} \circ g_{a_{n(r)-1}} \circ \cdots \circ g_{a_{n(r+1)-1}} \in \Gamma.$$

We also define the r^{th} *normalizing element* F_r to be the product:

$$F_r := G_{r-1}^{-1} \cdot G_{r-2}^{-1} \cdots \circ G_0^{-1} = g_{a_{n(r)-1}}^{-1} \circ \cdots \circ g_{a_0}^{-1} \in \Gamma.$$

This is the element of Γ that *renormalizes* the geodesic at time $t_{n(r)}$, in the sense that the geodesic $F_r(\gamma(t))$ passes through the fundamental domain $\varphi(\mathcal{F})$ for some portion of the time $[t_{n(r)}, t_{n(r+1)})$.

We remark that, for any $r \geq 1$, we have

$$F_r = G_{r-1}^{-1}F_{r-1} = g_{a_{n(r)-1}}^{-1} \circ \ldots g_{a_{n(r-1)}}^{-1} \circ F_{r-1}. \tag{6.4}$$

Step 0. Reformulation of the lim sup.

Let us first express the lim sup of h along γ as the lim sup over r of the supremum of h along the r^{th} segment of γ, and use the invariance of h under F_r, to get

$$\begin{aligned}\limsup_{t\to+\infty} h\big(\gamma(t)\big) &= \limsup_{r\to+\infty} \sup_{t_{n(r)}\leq t<t_{n(r+1)}} h\big(\gamma(t)\big)\\ &= \limsup_{r\to+\infty} \sup_{t_{n(r)}\leq t<t_{n(r+1)}} h\big(F_r \circ \gamma(t)\big).\end{aligned}$$

Step 1. Upper bound on intermediate times.

Fix $r \in \mathbb{N}$. We first establish a uniform bound for the value of h along any segment $\gamma\big([t_{n(r)}, t_{n(r+1)}]\big)$ when r is not equal to $r_k, r_k - 1$ or $r_k + 1$ for any $k \in \mathbb{Z}$. Fix an integer $N \geq 1$, let $\mathcal{K}_N \subset \mathbb{D}$ be the compact set provided by Lemma 5.1 and consider its image $\varphi(\mathcal{K}_N) \subset \mathbb{H}$ in the upper half plane, which is compact too. Since h is continuous, set

$$C_1 = C_1(N,h) := \max_{z\in\varphi(\mathcal{K}_N)} h(z) < +\infty.$$

By assumption $\ell(C_r) \leq N$ for any $r \neq r_k$ for any $k \in \mathbb{Z}$. Therefore, for all but at most finitely many intermediate r, i.e. any r different than any of r_k, $r_k + 1$ and $r_k - 1$, C_r is preceded and followed by cuspidal words with length less or equal to N and hence Lemma 5.1 implies that, for any such r and any t with $t_{n(r)} \leq t \leq t_{n(r+1)}$, we have

$$F_r(\gamma(t)) \in \varphi(\mathcal{K}_N).$$

Thus, for any intermediate r we have

$$\sup_{t_{n(r)}\leq t<t_{n(r+1)}} h\big(F_r(\gamma(t))\big) \leq C_1(N,h). \tag{6.5}$$

Step 2. Lower bound for the imaginary part along special segments.

Now we consider one of the special segments $\gamma^{(k)}$ which corresponds the block $C_{r_k-1}C_{r_k}C_{r_k+1}$ and we establish a lower bound on $\mathrm{Im}(z)$ for $z \in \gamma^{(k)}$, in order to guarantee that the supremum of h along $\gamma^{(k)}$ is taken for z in the central part of the segment.

Observe first that for any hyperbolic geodesic $t \mapsto \gamma'(t)$ in $\mathbb{H}$, for any a, b in $\mathbb{R}$ with $a < b$ and for any $t \in [a, b]$ one has $\operatorname{Im}\left(\gamma'(t)\right) \geq \min\left\{\operatorname{Im}\left(\gamma'(a)\right), \operatorname{Im}\left(\gamma'(b)\right)\right\}$. Thus, it is enough to control the imaginary part at the endpoints of the geodesic segment $\gamma^{(k)}$, which correspond to $t_{n(r_k-1)}$ and $t_{n(r_k+2)}$ respectively.

To this end, since for all k's large enough, $r_k - r_{k-1} \geq 4$, C_{r_k-2} and C_{r_k+2} are both preceded and followed by cuspidal words of length less than N, using again Lemma 5.1 with $r = r_k - 2$ and $t = t_{n(r_k-1)}$ and, respectively, $r = r_k + 2$ and $t = t_{n(r_k+2)}$ we have

$$F_{r_k-2}\left(\gamma\left(t_{n(r_k-1)}\right)\right) \in \varphi(\mathcal{K}_N) \quad \text{ and } \quad F_{r_k+2}\left(\gamma\left(t_{n(r_k+2)}\right)\right) \in \varphi(\mathcal{K}_N).$$

Therefore, recalling Equation (6.4), we have

$$\begin{aligned} F_{r_k}\left(\gamma\left(t_{n(r_k-1)}\right)\right) &= G_{r_k-1}^{-1} G_{r_k-2}^{-1} F_{r_k-2}\left(\gamma\left(t_{n(r_k-1)}\right)\right) \\ &\in G_{r_k-1}^{-1} G_{r_k-2}^{-1}\left(\varphi\left(\mathcal{K}_N\right)\right), \end{aligned}$$

and, similarly,

$$\begin{aligned} F_{r_k}\left(\gamma\left(t_{n(r_k+2)}\right)\right) &= G_{r_k} G_{r_{k+1}} F_{r_k+2}\left(\gamma\left(t_{n(r_k+2)}\right)\right) \\ &\in G_{r_k} G_{r_{k+1}}\left(\varphi\left(\mathcal{K}_N\right)\right) = G_{r_{k+1}}\left(\varphi(\mathcal{K}_N)\right) + M\mu, \end{aligned}$$

where the last line follows observing that that, by assumption (1) $C_{r_k} = \eta^M$, $G_{r_k}(z) = z + M\mu$.

Now, since by assumptions, the cuspidal words C_{r_k-2}, C_{r_k-1} and C_{r_k+1} all have length at most N, they correspond to elements of Γ whose norm is uniformly bounded. Moreover the image of the compact set $\varphi(\mathcal{K}_N)$ under such elements of Γ is contained in a bigger compact subset of $\mathbb{H}$, whose size depends only on N. It follows that there exists $\varepsilon_N > 0$, depending only on N, such that for any $k \geq 1$ we have

$$\operatorname{Im}\left(F_{r_k}\left(\gamma(t_{n(r_k-1)})\right)\right) \geq \varepsilon_N \quad \text{ and } \quad \operatorname{Im}\left(F_{r_k}\left(\gamma(t_{n(r_k+2)})\right)\right) \geq \varepsilon_N.$$

By applying the observation at the beginning of this step to $\gamma' := F_{r_k} \circ \gamma$ and $a = t_{n(r_k-1)}$ and $b = t_{n(r_k+2)}$, this shows that

$$\operatorname{Im}\big(F_{r_k}(\gamma(t))\big) \geq \varepsilon_N \quad \text{ for any } \quad t_{n(r_k-1)} \leq t \leq t_{n(r_k+2)}.$$

Recalling that h is Γ-periodic, and thus in particular periodic under the translation $z \mapsto z + \mu$, set

$$C_2 = C_2(N, h, l) := \max_{\varepsilon_N \leq \operatorname{Im}(z) \leq l} h(z) < +\infty.$$

Step 3. Lower bound on the supremum on special segments.

Let us recall that the block $C_{r_k-1} C_{r_k} C_{r_k+1}$ codes the geodesic γ from time $t = t_{n(r_k-1)}$ up to time $t = t_{n(r_k+2)}$. Moreover the central word C_{r_k} corresponds to M iterations of the parabolic transformation $p(z) = z + \mu$. Hence, as in the proof of Theorem 1.4, we see that the renormalized geodesic $F_{r_k} \circ \gamma$ crosses exactly M vertical lines of the form $\mathcal{V}_j := \{z \in \mathbb{H} : \operatorname{Re}(z) = j\frac{\mu}{2}, j \in \mathbb{Z}\}$ in the upper half plane, see Figure 5. It follows that

$$\sup_{t_{n(r_k-1)} \leq t \leq t_{n(r_k+2)}} \operatorname{Im}\big(F_{r_k}(\gamma(t))\big) = \sup_{t \in \mathbb{R}} \operatorname{Im}\big(F_{r_k}(\gamma(t))\big) \geq (M-1)\frac{\mu}{2}.$$

We now choose M such that

$$(M-1)\frac{\mu}{2} \geq \max\left\{l, C_1(N,h) + \delta, C_2(N,h,l) + \delta\right\},$$

with $\delta = \delta_G$, defined in Theorem 1.8. It follows that there exists some $t^{(k)}$ with $t_{n(r_k-1)} \leq t^{(k)} \leq t_{n(r_k+2)}$ such that $F_{r_k}(\gamma(t^{(k)})) \in \mathcal{U}_l$ and, moreover, $\operatorname{Im}\big(F_{r_k}(\gamma(t^{(k)}))\big) \geq (M-1)\mu/2$. Since $|h(z) - \operatorname{Im}(z)| < \delta$ for any $z \in \mathcal{U}_l$, then for such $t^{(k)}$ we have

$$h\big(F_{r_k}(\gamma(t^{(k)}))\big) \geq \operatorname{Im}\big(F_{r_k}(\gamma(t^{(k)}))\big) - \delta \geq (M-1)\frac{\mu}{2} - \delta \geq \max\left\{C_1(N,h), C_2(N,h,l)\right\}. \tag{6.6}$$

Step 4. Final arguments. We can now conclude the proof. From Equation (6.5) and (6.6) it follows that

$$\limsup_{t \to +\infty} h\big(\gamma(t)\big) = \limsup_{r \to +\infty} \sup_{t_{n(r)} \leq t \leq t_{n(r+1)}} h\big(F_r(\gamma(t))\big).$$

Now, Equation (6.6) also implies that the large values of $h\big(F_{r_k}(\gamma(t))\big)$ are always taken when $F_{r_k}(\gamma(t)) \in \mathcal{U}_l$. Moreover, we claim that $h_l\big(F_{r_k}(\gamma(t))\big) = 0$ for $t \notin [t_{n(r_k-1)}, t_{n(r_k+2)}]$. In fact, we recall that the fundamental horodisk $\mathcal{U}_l$ is precisely invariant, meaning that for each $g \in G$ we either have $g(\mathcal{U}_l) = \mathcal{U}_l$ or $\mathcal{U}_l \cap g(\mathcal{U}_l) = \varnothing$, and the former happens only if g is a power of p. Hence, for any $t \notin [t_{n(r_k-1)}, t_{n(r_k+2)}]$, as we can assume without loss of generality that $l \geq 1$, there exists some non parabolic $g \in \Gamma$ with $F_{r_k}(\gamma(t)) \in g(\mathcal{U}_l)$ and $\mathcal{U}_l \cap g(\mathcal{U}_l) = \varnothing$. Thus, we get that

$$\sup_{t_{n(r)} \leq t \leq t_{n(r+1)}} h\big(F_r(\gamma(t))\big) = \sup_{t_{n(r_k-1)} \leq t \leq t_{n(r_k+2)}} h_l\left(F_{r_k}(\gamma(t))\right) = \sup_{t\in\mathbb{R}} h_l\big(F_{r_k}(\gamma(t))\big).$$

Finally, recalling the definition of the function $H(\cdot,\cdot)$ and the expression of the endpoints of the geodesic $F_{r_k} \circ \gamma$, we get

$$\begin{aligned}\sup_{t\in\mathbb{R}} h_l\big(F_{r_k}(\gamma(t))\big) &= H\left(F_{r_k}(\gamma(-\infty)), F_{r_k}(\gamma(+\infty))\right)\\ &= H\left(\big[a_{n(r_k)-1}, a_{n(r_k)-2}, \dots\big]^{-}_{\partial\mathbb{H}}, \big[a_{n(r_k)}, a_{n(r_k)+1}, \dots\big]_{\partial\mathbb{H}}\right).\end{aligned}$$

Combining all the last series of equalities together, the proof is hence concluded. □

6.4. **Proof of Theorem 1.8.** We have all the ingredients in order to give the proof of Theorem 1.8 following the same scheme of the proof of Theorem 1.4.

Proof of Theorem 1.8. Let us first assume that $m = 1$ and prove the result under the Lipschitz condition (1.10) in Remark 1.9. We will then show at the end how to deduce the result for other values of m from this special case. Let us first verify that we can apply Proposition 6.5 and Corollary 6.6. Assume without loss of generality that l_0 in the assumptions of Theorem 1.8 is greater than 1. Thus, from the assumption (1.10) on the Lipschitz control of the perturbation on $\mathcal{U}_{l_0}$, by Proposition 6.3,

$$\mathrm{Lip}\big((H - H_0)|_{U_{l_0}}\big) \leq \left(\sqrt{2} + \frac{\sqrt{2}}{l_0}\right) \cdot \|(h - \mathrm{Im})|_{\mathcal{U}_{l_0}}\|_{\mathrm{Lip}} \leq \frac{2\sqrt{2}}{4\sqrt{2}} = \frac{1}{2}.$$

Moreover, since by Lemma 6.2 $\|(H - H_0)|_{U_{l_0}}\|_\infty \leq \|(h - \mathrm{Im})|_{\mathcal{U}_{l_0}}\|_\infty \leq \|(h - \mathrm{Im})|_{\mathcal{U}_{l_0}}\|_{\mathrm{Lip}} < 1/4$. Thus, there exists $0 < \varepsilon < 1$ such that the assumptions of Proposition 6.5 and Corollary 6.6 hold for the function H corresponding to h and $l = l_0$.

Let s_0 be given by Proposition 6.5 and let N_0 be as in Corollary 6.6. Let also $M_0 := M(l_0, h, N_0)$ be given by Proposition 6.7 in correspondence to N_0. We will show that

$$[L_0, +\infty) \subset \mathcal{L}(X, h), \qquad \text{where } L_0 := \max\{\inf H(\mathbb{K}_N \times \mathbb{K}_N^{s_0}), \inf H(\mathbb{K}_N \times \mathbb{K}_N^{M_0})\}.$$

Take any $L \geq L_0$. By Corollary 6.6 (with $s_1 = \max\{M_0, s_0\}$), there exist points $x_2 = [a_0, a_1, \dots]_{\partial\mathbb{H}} \in \mathbb{K}_N$ and $x_1 = [b_0, b_1, \dots]_{\partial\mathbb{H}} \in \mathbb{K}_N$ and an integer s such that

$$L = H(x_1, x_2 + s\mu), \qquad s \geq s_1. \tag{6.7}$$

We now construct a geodesic γ such that $L(\gamma, h) = L$, by prescribing its symbolic coding $(w_n)_{n\in\mathbb{Z}}$. The construction is the same that the one in the proof of Theorem 1.4, so we only sketch it to avoid unnecessary repetitions. We construct the word $(w_n)_{n\in\mathbb{Z}}$ by concatenating blocks of the form $W_j = \overline{b_{|j|}} \dots \overline{b_0} \eta^s a_0 \dots a_{|j|}$, $j \in \mathbb{Z}$, interpolated via letters δ_j, δ'_j chosen exactly as in the proof of Theorem 1.4, so that in particular $(w_n)_{n\in\mathbb{Z}}$ satisfies the non-backtracking condition (2.4) and hence is the cutting sequence of a geodesic γ. Recall also that the central block η^s in the word W_k is, by construction, a single parabolic word.

The assumptions of Proposition 6.7 apply by construction to the geodesic γ, by letting $(r_k)_{k\in\mathbb{Z}}$ the sequence such that the parabolic word $C_{r_k} = \eta^s$ is the central block of W_k. In fact, Condition (1) is obvious since $s \geq M_0$ by construction, and the distance between r_k and r_{k-1} grows linearly; Condition (2) follows since $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ code points in $\mathbb{K}_N$ and by choice of the interpolating letters, we refer to the proof of Theorem 1.4 for details.

Thus, Proposition 6.7, the form of the word $(w_n)_{n\in\mathbb{Z}}$ and (6.7), give that $L(h, \gamma) = L$. This concludes the proof under the assumptiont that $m = 1$.

Finally, let us deal with the case when $m \neq 1$. We can conjugate the group G with an element $g \in \mathrm{PSL}(2, \mathbb{R})$ that fixes infinity and that normalizes m to 1. In particular we can choose $g(z) = z/m$. Denote by

$$h' := h \circ g^{-1}, \qquad \mathrm{Im}' = \mathrm{Im} \circ g^{-1} = m \cdot \mathrm{Im}\,.$$

If we set $G' := gGg^{-1}$, there is a one to one correspondence between geodesics γ on $X = G\backslash\mathbb{H}$ and geodesics γ' on $X' = G'\backslash\mathbb{H}$, given by $\gamma'(t) = g(\gamma(t))$. Using this observation and recalling the definition (1.6) of Lagrange values, we have that

$$L_{G'}(h', \gamma') = \limsup_{t\to\infty} h'(\gamma'(t)) = \limsup_{t\to\infty} h \circ g^{-1}(g(\gamma(t))) = \limsup_{t\to\infty} h(\gamma(t)) = L_G(h, \gamma). \tag{6.8}$$

The formula implies that the two corresponding spectra coincide, that is, $\mathcal{L}(X, h) = \mathcal{L}(X', h')$. Thus, it is now enough to show that the group G' and the function h' satisfy the assumptions of Theorem 1.8 with $m = 1$. We begin by observing that a point $z \in \mathcal{U}_l$ if and only if $g^{-1}(z) \in \mathcal{U}_{lm}$. This implies that, for any $l > 0$,

$$\|(h' - \mathrm{Im}')|_{\mathcal{U}_l}\|_\infty = \sup_{x\in\mathcal{U}_l} |((h - \mathrm{Im}) \circ g^{-1})(x)| = \sup_{y\in\mathcal{U}_{lm}} |((h - \mathrm{Im}))(y)| = \|(h - \mathrm{Im})|_{\mathcal{U}_{lm}}\|_\infty.$$

Similarly we have that

$$\begin{aligned}
\mathrm{Lip}((h' - \mathrm{Im}')|_{\mathcal{U}_l}) &= \sup_{x,y\in\mathcal{U}_l} \frac{|(h' - \mathrm{Im}')(x) - (h' - \mathrm{Im}')(y)|}{|x - y|} \\
&= \sup_{x,y\in\mathcal{U}_l} \frac{|((h - \mathrm{Im}) \circ g^{-1})(x) - ((h - \mathrm{Im}) \circ g^{-1})(y)|}{|x - y|} \\
&= \sup_{x',y'\in\mathcal{U}_{lm}} \frac{|(h - \mathrm{Im})(x') - (h - \mathrm{Im})(y')|}{|g(x') - g(y')|} \\
&= \sup_{x',y'\in\mathcal{U}_{lm}} \frac{|(h - \mathrm{Im})(x') - (h - \mathrm{Im})(y')|}{\frac{1}{m}|x' - y'|} = m \cdot \mathrm{Lip}((h - \mathrm{Im})|_{\mathcal{U}_{lm}}).
\end{aligned}$$

Thus, the computations above show that

$$\|(h' - \mathrm{Im}')|_{\mathcal{U}_l}\|_{\mathrm{Lip}} = \|(h - \mathrm{Im})|_{\mathcal{U}_{lm}}\|_\infty + m\,\mathrm{Lip}((h - \mathrm{Im})|_{\mathcal{U}_{lm}}) \leq \max\{1, m\} \cdot \|(h - \mathrm{Im})|_{\mathcal{U}_{lm}}\|_{\mathrm{Lip}}.$$

Let us now assume that, for some $l_0 \geq m > 0$, we have

$$\|(h - \mathrm{Im})|_{\mathcal{U}_{l_0}}\|_{\mathrm{Lip}} < \delta_G := \min\left\{\frac{1}{4m\sqrt{2}}, \frac{1}{4\sqrt{2}}\right\},$$

which yields

$$\|(h' - \mathrm{Im}')|_{\mathcal{U}_{\frac{l_0}{m}}}\|_{\mathrm{Lip}} \leq \max\{1, m\} \cdot \|(h - \mathrm{Im})|_{\mathcal{U}_{l_0}}\|_{\mathrm{Lip}} < \max\{1, m\} \cdot \delta_G = \frac{1}{4\sqrt{2}},$$

where the last equality follows considering separately the cases $m \leq 1$ and $m \geq 1$. Thus, the first part of the proof implies that $\mathcal{L}(X', h')$ contains a Hall ray. Thanks to (6.8), this shows also that $\mathcal{L}(X, h)$ contains a Hall ray and hence concludes the proof in the general case. □

7. Proof of the Stable Hall theorem

In this section we prove Theorem 1.12 stated in the introduction (see Section 1.8), which generalizes Hall's theorem on the sum of Cantor sets to Lipschitz perturbations of the sum. Throughout the section we use the notation introduced in Section 1.8.

Consider any Cantor set $\mathbb{K}$ and let $(\mathbb{K}(n))_{n\in\mathbb{N}}$ be a slow subdivision for $\mathbb{K}$. By Remark 1.11, the collection of holes of $\mathbb{K}$ inherits from the subdivision an ordering. We will denote by B_n the n^{th} hole, so that $(B_n)_{n\in\mathbb{N}}$ is the collection of holes of $\mathbb{K}$ ordered according to $(\mathbb{K}(n))_{n\in\mathbb{N}}$. We say that $(\mathbb{K}(n))_{n\in\mathbb{N}}$ is a *monotone slow subdivision* for $\mathbb{K}$ if the ordered sequence of holes $(B_n)_{n\in\mathbb{N}}$ satisfies

$$|B_{n+1}| \leq |B_n| \quad \text{for any} \quad n \in \mathbb{N}.$$

It is clear that monotone slow subdivisions always exist.

Let us now state two preliminary Lemmas which will be used in the proof of Theorem 1.12

Lemma 7.1. *Let $(\mathbb{K}(n))_{n\in\mathbb{N}}$ be a monotone slow subdivision for the Cantor set $\mathbb{K}$. If $\mathbb{K}$ admits another slow subdivision $(\widetilde{\mathbb{K}}(n))_{n\in\mathbb{N}}$ which satisfies the ε-stable gap condition (1.12), then the same is true for the monotone slow subdivision $(\mathbb{K}(n))_{n\in\mathbb{N}}$.*

This Lemma was proved as Lemma A.1 in the Appendix of [1] (see also [8]).

Remark 7.2. Observe that if K and F are closed intervals then $K \times F$ is connected and since $S\colon U \to \mathbb{R}$ is continuous then the image $S(K \times F)$ is connected too, that is it is an interval. Moreover, for the same reason, if K and F are closed intervals, then $K \times F$ is compact and thus its image $S(K \times F)$ is compact, and thus closed.

The next Lemma provides the key step to prove the Stable Hall theorem.

Lemma 7.3. *Let $S\colon U \to \mathbb{R}$ be a function satisfying Condition (1.14). Let K and F be two compact intervals with $K \times F \subset U$. Let B be an open interval contained in K such that $|B| < (1-\varepsilon)|F|$. Then we have*

$$S(K \times F) = S(K^L \times F) \cup S(K^R \times F).$$

Similarly, if C is an open interval contained in F with $|C| \leq (1-\varepsilon)|K|$ then we have

$$S(K \times F) = S(K \times F^L) \cup S(K \times F^R).$$

Proof. We only prove the first statement, the argument for the second being the same. Set

$$G := S - S_0.$$

Let $K = [a,b]$, $F = [c,d]$ and $B = (e,f) \subset K$ for real numbers

$$a < e < f < b, \qquad c < d.$$

Let us first show that

$$\inf S(K \times F) = \inf S(K^L \times F). \tag{7.1}$$

Since the inequality $\leq$ is obvious, it is enough to prove the inequality $\geq$. Moreover, since $K = K^L \cup [e,b]$, it is enough to show that

$$\inf S([e,b] \times F) \geq \inf S(K^L \times F). \tag{7.2}$$

To prove this, consider any (x_1, x_2) with $e \leq x_1 \leq b$ and $x_2 \in F$, i.e. $c \leq x_2 \leq d$ and (a,c), which belongs to $K^L \times F$. Recalling the definition of Lipschitz constant (1.8) and using that $\mathrm{Lip}(G) < 1$ and $x_1 \geq a$, $x_2 \geq c$, we have that

$$\begin{aligned}
S(x_1,x_2) - S(a,c) &= S_0(x_1,x_2) - S_0(a,c) + G(x_1,x_2) - G(a,c) \\
&= (x_1 + x_2) - (a + c) + G(x_1,x_2) - G(a,x_2) + G(a,x_2) - G(a,c) \\
&= |x_1 - a| + |x_2 - c| + \frac{G(x_1,x_2) - G(a,x_2)}{x_1 - a}(|x_1 - a|) + \frac{G(a,x_2) - G(a,c)}{x_2 - c}(|x_2 - c|) \\
&\geq |x_1 - a|\big(1 - \mathrm{Lip}(G)\big) + |x_2 - c|\big(1 - \mathrm{Lip}(G)\big) > 0.
\end{aligned}$$

This proves Equation (7.2) and hence concludes the proof of Equation (7.1). With a similar argument, we get also that

$$\sup S(K \times F) = \sup S(K^R \times F).$$

Since $S(K \times F)$, $S(K^L \times F)$ and $S(K^R \times F)$ are three intervals (see Remark 7.2), it is hence enough to prove that

$$\sup S(K^L \times F) \geq \inf S(K^R \times F).$$

To show this, we will show that $S(e,d) > S(f,c)$ (remark that $(e,d) \in K^L \times F$ and $(f,c) \in K^R \times F$). Reasoning as before, we have

$$\begin{aligned}
S(e,d) - S(f,c) &= S_0(e,d) - S_0(f,c) + G(e,d) - G(f,c) \\
&= (d - c) - (f - e) + G(e,d) - G(e,c) + G(e,c) - G(f,c) \\
&\geq |d - c|\big(1 - \mathrm{Lip}(G)\big) - |f - e|\big(1 + \mathrm{Lip}(G)\big)
\end{aligned}$$

so that $S(e,d) > S(f,c)$ is implied by

$$|B| = |f - e| < \frac{1 - \mathrm{Lip}(G)}{1 + \mathrm{Lip}(G)} \cdot |d - c| = \frac{1 - \mathrm{Lip}(G)}{1 + \mathrm{Lip}(G)} \cdot |F|,$$

which is satisfied according to Condition (1.14), because $|B| < (1-\varepsilon)|F|$ by assumption. $\square$

We can now give the Proof of Theorem 1.12.

Proof of Theorem 1.12. Let $\big(\mathbb{K}(n)\big)_{n\in\mathbb{N}}$ and $\big(\mathbb{F}(n)\big)_{n\in\mathbb{N}}$ be slow monotone subdivisions respectively for $\mathbb{K}$ and $\mathbb{F}$. Since by assumption $\mathbb{K}$ and $\mathbb{F}$ admit a slow subdivision which satisfy Condition (1.12), then by Lemma 7.1 the same is true for the subdivisions $\big(\mathbb{K}(n)\big)_{n\in\mathbb{N}}$ and $\big(\mathbb{F}(n)\big)_{n\in\mathbb{N}}$. Set $K_0 := [\min\mathbb{K}, \max\mathbb{K}]$ and $F_0 := [\min\mathbb{F}, \max\mathbb{F}]$ and fix any point $x \in S(K_0 \times F_0)$.

The Theorem follows if we show that we can construct two sequences $(n_i)_{i\in\mathbb{N}}$ and $(m_i)_{i\in\mathbb{N}}$ such that $m_{i+1} \geq m_i$ and $n_{i+1} \geq n_i$ for any $i \in \mathbb{N}$ and $n_i \to \infty$, $m_i \to \infty$, and two sequences of nested closed intervals $(K_i)_{i\in\mathbb{N}}$ and $(F_i)_{i\in\mathbb{N}}$, where K_i is an interval of the level $\mathbb{K}(n_i)$ and F_i is an interval of the level $\mathbb{F}(m_i)$, such that for any $i \in \mathbb{N}$ we have

$$x \in S(K_i \times F_i).$$

Indeed setting $k := \bigcap_{i\in\mathbb{N}} K_i$ and $f := \bigcap_{j\in\mathbb{N}} F_j$ continuity of S implies $x = S(k, f)$, where $k \in \mathbb{K}$ and $f \in \mathbb{F}$. Observe that we require $n_i \to \infty$, but steps i for which $n_{i+1} = n_i$ are allowed, and similarly for the integers m_i.

We will construct the sequences $(n_i)_{i\in\mathbb{N}}$ and $(m_i)_{i\in\mathbb{N}}$ and the two families of nested intervals by induction on i in $\mathbb{N}$. Fix $i \in \mathbb{N}$ and assume that respectively the first $i+1$ nested intervals $K_0 \supset K_1 \supset \cdots \supset K_i$ and the first $i+1$ nested intervals $F_0 \supset F_1 \supset \cdots \supset F_i$ are defined. Let $n(K_i)$ be the minimum $n \in \mathbb{N}$ such that $K_i \cap \mathbb{K}(n) \neq K_i$ and let B_i be the hole in K_i, i.e. the open subinterval $B_i \subset K_i$ such that $\mathbb{K}(n(K_i)) \cap K_i = K_i \setminus B_i$. Similarly, let $m(F_i)$ be the minimum $m \in \mathbb{N}$ such that $F_i \cap \mathbb{F}(m) \neq F_i$ and let C_i be the hole in F_i, i.e. the open subinterval $C_i \subset F_i$ such that $\mathbb{F}(m(F_i)) \cap F_i = F_i \setminus C_i$.

During the inductive construction, we will also prove that for every i the intervals (K_i, F_i) and the holes $B_i \subset K_i$, $C_i \subset F_i$ in our construction satisfy the following *balanced gap* condition:

$$|B_i| < (1-\varepsilon)|F_i| \quad \text{and} \quad |C_i| < (1-\varepsilon)|K_i|. \tag{7.3}$$

Observe that for $i = 0$ the condition is true according to the ε-size condition (1.13). Assume that balanced gap condition (7.3) is satisfied for $i \geq 0$. To define the intervals at level $i+1$, we subdivide the interval having the bigger hole. Assume that $|B_i| \geq |C_i|$, the other case being the same. Since $|B_i| < (1-\varepsilon)|F_i|$ then Lemma 7.3 implies

$$S(K_i \times F_i) = S(K_i^L \times F_i) \cup S(K_i^R \times F_i).$$

If $x \in S(K_i^L \times F_i)$ (respectively $x \in S(K_i^R \times F_i)$), set $K_{i+1} := K_i^L$ (respectively $K_{i+1} := K_i^R$) and $n_{i+1} = n(K_i)$, so that $K_{i+1} \in \mathbb{K}(n_{i+1})$. Set also $F_{i+1} = F_i$ and $m_{i+1} = m_i$, so that $F_{i+1} \in \mathbb{F}(m_{i+1})$ holds trivially. By the property of a monotone slow subdivision, the hole $B_{i+1} \subset K_{i+1}$ satisfies $|B_{i+1}| \leq |B_i|$ and therefore by inductive assumption we get

$$|B_{i+1}| \leq |B_i| < (1-\varepsilon)|F_i| = (1-\varepsilon)|F_{i+1}|.$$

On the other hand Condition (1.12) implies $|B_i| < (1-\varepsilon)|K_i^L| = (1-\varepsilon)|K_{i+1}|$ and therefore, since C_i is by choice the smaller of the two holes, we get

$$|C_i| \leq |B_i| < (1-\varepsilon)|K_{i+1}|.$$

Thus, the pair of intervals (K_{i+1}, F_{i+1}), with holes B_{i+1} and $C_{i+1} = C_i$ satisfies balanced gap condition (7.3), and moreover we have $x \in S(K_{i+1} \times F_{i+1})$. The inductive step is complete.

Finally, since there are only finitely many holes which are longer than a given positive constant, it is clear that both $(n_i)_{i\in\mathbb{N}}$ and $(m_i)_{i\in\mathbb{N}}$ satisfy $n_i \to \infty$ and $m_i \to \infty$. Theorem 1.12 is proved. □

Appendix A. Proofs of Lemmas on parabolic words

We present here the simple proofs of Lemma 2.3 and Lemma 2.4 on parabolic words in Section 2.4. With a slightly different notation, the proofs were essentially contained in [1].

Proof of Lemma 2.3. In order to prove point (1), observe that for any $k = 0, \dots, n-1$ we have

$$\mathcal{A}[a_0, \dots, a_k] = g_{a_0} \circ \cdots \circ g_{\alpha_{k-1}} \mathcal{A}[a_k],$$
$$\mathcal{A}[a_0, \dots, a_k, a_{k+1}] = g_{a_0} \circ \cdots \circ g_{\alpha_k} \mathcal{A}[a_{k+1}].$$

Applying $\big(g_{a_0} \circ \cdots \circ g_{a_{k-1}}\big)^{-1}$ and recalling that elements of Γ preserve the orientation on $\partial\mathbb{D}$, we see that $\mathcal{A}[a_0, \dots, a_k]$ and $\mathcal{A}[a_0, \dots, a_k, a_{k+1}]$ share the same left endpoint if and only if

$$\xi^l_{a_k} = \inf \mathcal{A}[a_k] = g_{a_k}\big(\inf \mathcal{A}[a_{k+1}]\big) = g_{a_k}(\xi^l_{a_{k+1}}).$$

Point (2) for right endpoints follows with the same argument. In order to prove point (3), recall that, for any letter α, $\xi^l_\alpha = g_\alpha(\xi^r_{\overline{\alpha}})$ and $g_\alpha^{-1} = g_{\overline{\alpha}}$. Therefore, according to first two points, $a_0 \dots a_n$ is left cuspidal if and only if for any $k = 0, \dots, n-1$ we have

$$g_{a_k}(\xi^l_{a_{k+1}}) = \xi^l_{a_k} \iff g_{a_k} g_{a_{k+1}}(\xi^r_{\overline{a_{k+1}}}) = g_{a_k}(\xi^r_{\overline{a_k}}) \iff \xi^r_{\overline{a_{k+1}}} = g^{-1}_{a_{k+1}}(\xi^r_{\overline{a_k}}) = g_{\overline{a_{k+1}}}(\xi^r_{\overline{a_k}}),$$

that is $\overline{a_n} \dots \overline{a_0}$ is right cuspidal. □

Proof of Lemma 2.4. Point (1) of Lemma 2.3 implies $\xi^l_{a_k} = g_{a_k}(\xi^l_{a_{k+1}})$ for any $k = 0, \dots, n-1$, thus

$$\xi^l_{a_0} = g_{a_0} \circ \cdots \circ g_{a_k}(\xi^l_{a_{k+1}}).$$

If $a_0 \dots a_n a_0$ is also cuspidal then the condition above holds with $k = n$ and therefore

$$g(\xi^l_{a_0}) = \xi^l_{a_0}.$$

Since $a_0 \dots a_n a_0$ is left cuspidal, then also $a_n a_0 \dots a_n a_0$ is so, and finally $a_n a_0 \dots a_n$ is left cuspidal too. Hence, we have

$$\xi^l_{a_0} = g^{-1}_{a_n}(\xi^l_{a_n}) = g^{-1}_{a_n} \circ g_{a_n}(\xi^r_{\overline{a_n}}) = \xi^r_{\overline{a_n}}.$$

According to points (2) and (3) of Lemma 2.3, the word $\overline{a_n} \dots \overline{a_0}\,\overline{a_n}$ is right cuspidal and reasoning as above we have that

$$g^{-1}(\xi^r_{\overline{a_n}}) = g_{\overline{a_n}} \circ \cdots \circ g_{\overline{a_0}}(\xi^r_{\overline{a_n}}) = \xi^r_{\overline{a_n}}.$$

Observe also that

$$g\mathcal{A}[a_0] = g_{a_0} \circ \dots g_{a_n} \mathcal{A}[a_0] = \mathcal{A}[a_0, \dots, a_n, a_0] \subset \mathcal{A}[a_0],$$
$$g^{-1}\mathcal{A}[\overline{a_n}] = g_{\overline{a_n}} \circ \dots g_{\overline{a_0}} \mathcal{A}[\overline{a_n}] = \mathcal{A}[\overline{a_n}, \dots, \overline{a_0}, \overline{a_n}] \subset \mathcal{A}[\overline{a_n}].$$

Thus $\xi^l_{a_0}$ is a fixed point of g and $\mathcal{A}[a_0]$ is a right neighborhood of it where g is contracting. On the other hand $\mathcal{A}[\overline{a_n}]$ is a left neighborhood of $\xi^l_{a_0}$ where g^{-1} acts contracting, thus g is expanding. It follows that $\xi^l_{a_0}$ is not hyperbolic and is thus the unique fixed point of g. Thus g is a parabolic element of Γ. □

Appendix B. Lipschitz norm estimates

In this Appendix, we present the proof of Proposition 6.3.

We use in this Appendix the notation introduced in the beginning of Section 6. We recall, in particular that, for $l > 0$, $\mathcal{U}_l \subset \mathbb{H}$ is a horocyclic neighborhood and $U_l \subset \mathbb{R}^2$ a the diagonal neighborhood. In order to simplify the notation, we write simply H and H_0 instead of $H|_{U_l}$ and $H_0|_{U_l}$ respectively and simply h and Im instead of $h|_{\mathcal{U}_l}$ and $\mathrm{Im}\,|_{\mathcal{U}_l}$ respectively. For x_1, x_2 in $\mathbb{R}$ let $\gamma(x_1, x_2, \cdot) \colon \mathbb{R} \to \mathbb{H}$, $t \mapsto \gamma(x_1, x_2, t)$ be the geodesic parametrization of the hyperbolic geodesic $\gamma(x_1, x_2)$ in $\mathbb{H}$ with endpoints x_1, x_2, such that $\gamma(x_1, x_2, 0)$ is its highest point, i.e.

$$\text{(B.1)} \qquad \lim_{t\to-\infty} \gamma(x_1, x_2, t) = x_1, \qquad \lim_{t\to+\infty} \gamma(x_1, x_2, t) = x_2, \qquad \gamma(x_1, x_2, 0) = \frac{x_1 + x_2}{2} + i\frac{x_2 - x_1}{2}.$$

Finally set $\delta := \|h - \mathrm{Im}\,\|_{\mathrm{Lip}}$ and recall that we have

$$\sup_{z\in\mathcal{U}_l} |h(z) - \mathrm{Im}(z)| \le \delta \quad \text{ and } \quad \sup_{z,z'\in\mathcal{U}_l} \frac{\left|\big(h(z) - \mathrm{Im}(z)\big) - \big(h(z') - \mathrm{Im}(z')\big)\right|}{|z - z'|} \le \delta.$$

We recall that the Lipschitz constant of any $G \colon U \subset \mathbb{R}^2 \to R$ is given by

$$\sup_{(x_1,x_2),(x'_1,x'_2)\in U} \frac{|G(x_1, x_2) - G(x'_1, x'_2)|}{|(x_1, x_2) - (x'_1, x'_2)|},$$

where $|(y_1, y_2)|$ denotes the Euclidean norm of $|(y_1, y_2)| \in \mathbb{R}^2$ (and hence corresponds to the absolute value $|y_1 + iy_2|$ in $\mathbb{C}$).

The idea of the proof is to first estimate the Lipschitz constant of H in two directions which are geometrically meaningful and hence easier to control. We remark indeed that if we consider a point (x'_1, x'_2) of the form $(x_1, x_2) + s(1, 1) = (x_1 + s, x_2 + s)$, where $s \in \mathbb{R}$, the geodesic $\gamma(x'_1, x'_2)$ is obtaining by *sliding* horizontally the endpoints of $\gamma(x_1, x_2)$; in particular, the geodesics are rigidly translated. On the other hand, if we consider a point (x'_1, x'_2) of the form $(x_1, x_2) + s(-1, 1) = (x_1 - s, x_2 + s)$, for small values of $s \in \mathbb{R}$, the geodesic $\gamma(x'_1, x'_2)$, as a Euclidean semi-circle, is *concentric* to $\gamma(x_1, x_2)$, while the radii differ by s, see Figure 8.

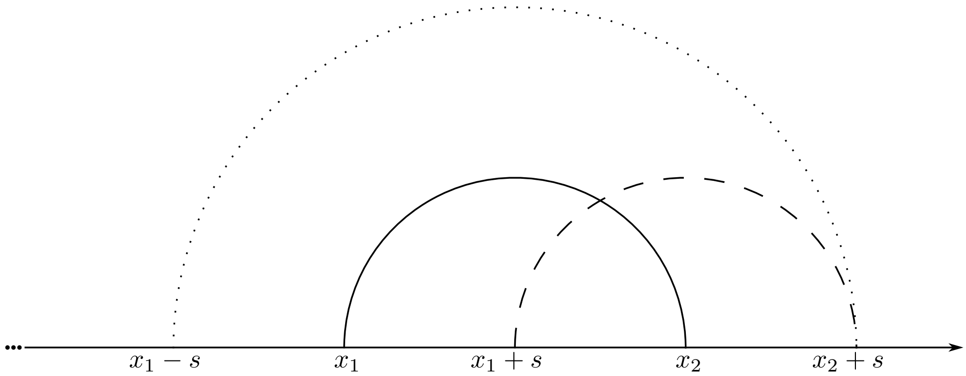

FIGURE 8. The geometric meaning of the vectors $(1,1)$ and $(-1,1)$ considered in the proof of Proposition 6.3.

Proof of Proposition 6.3. Set $G := H - H_0$. Consider the vectors $v := (1,1)$ and $w := (-1,1)$ (for the motivation explained before the proof). We will prove that, for any line $\mathcal{V}$ parallel to v and any line $\mathcal{W}$ parallel to w, we have

$$\mathrm{Lip}(G|_{\mathcal{V}}) \leq \frac{\delta}{\sqrt{2}} \quad \text{and} \quad \mathrm{Lip}(G|_{\mathcal{W}}) \leq \left(\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{r}\right)\delta,$$

where $G|_{\mathcal{V}}$ and $G|_{\mathcal{W}}$ denote respectively the restriction of G to the lines $\mathcal{V}$ and $\mathcal{W}$. We claim that this is enough to conclude, since for any (x_1,x_2) and (x_1',x_2') in $\mathbb{R}^2$, there exists λ,μ in $\mathbb{R}$ and $(x_1^*,x_2^*) \in \mathbb{R}^2$ such that

$$(x_1,x_2) - (x_1^*,x_2^*) = \lambda v \quad \text{and} \quad (x_1^*,x_2^*) - (x_1',x_2') = \mu w,$$

so that, remarking that v and w are orthogonal and hence one can use Pythagoras' theorem,

$$\begin{aligned}|G(x_1,x_2) - G(x_1',x_2')| &\leq |G(x_1,x_2) - G(x_1^*,x_2^*)| + |G(x_1^*,x_2^*) - G(x_1',x_2')| \\ &\leq \frac{\delta}{\sqrt{2}} \cdot |(x_1,x_2) - (x_1^*,x_2^*)| + \left(\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{r}\right)\delta \cdot |(x_1^*,x_2^*) - (x_1',x_2')| \\ &\leq \left(\sqrt{2} + \frac{\sqrt{2}}{r}\right)\delta \cdot |(x_1,x_2) - (x_1',x_2')|.\end{aligned}$$

We will consider separately the estimate for $G|_{\mathcal{V}}$ and the one for $G|_{\mathcal{W}}$. Before doing that, for any pair of points (x_1,x_2) and (x_1',x_2') in $\mathbb{R}^2$, let

$$\gamma = \{\gamma(t) := \gamma(x_1,x_2,t),\ t \in \mathbb{R}\}, \qquad \gamma' := \{\gamma'(t) := \gamma(x_1',x_2',t),\ t \in \mathbb{R}\},$$

be the time parametrizations of the geodesics with respective endpoints x_1,x_2 and x_1',x_2' described in (B.1). Let also t_0,t_0' in $\mathbb{R}$ be such that

$$H(x_1,x_2) = h(\gamma(t_0)) \quad \text{and} \quad H(x_1',x_2') = h(\gamma'(t_0')).$$

Estimate for $G|_{\mathcal{V}}$. In order to prove the estimate for $G|_{\mathcal{V}}$, where $\mathcal{V}$ is any line parallel to v, consider any (x_1,x_2) and (x_1',x_2') in such line $\mathcal{V}$ and let

$$s := x_2' - x_2 = x_1' - x_1, \qquad \text{so} \quad (x_1',x_2') = (x_1,x_2) + s(1,1).$$

As we remarked before the proof, the geodesic γ' is hence obtained by sliding horizontally γ by s, i.e. for every $t \in \mathbb{R}$ we have $\gamma'(t) = \gamma(t) + s$. In particular, $\gamma(t_0) + s \in \gamma'$. Moreover, remark that since the function $z \mapsto \mathrm{Im}(z)$ is constant along horizontal lines, we have that $\mathrm{Lip}(h|_{\mathcal{V}}) = \mathrm{Lip}((h - \mathrm{Im})|_{\mathcal{V}}) \leq \delta$. It follows from these two remarks that

$$H(x_1',x_2') = h\big(\gamma'(t_0')\big) \geq h\big(\gamma(t_0) + s\big) \geq h\big(\gamma(t_0)\big) - \delta \cdot |s| = H(x_1,x_2) - \delta \cdot |s|.$$

Similarly, using this time that $\gamma'(t_0') - s \in \gamma$,

$$H(x_1,x_2) = h\big(\gamma(t_0)\big) \geq h\big(\gamma'(t_0') - s\big) \geq h\big(\gamma'(t_0')\big) - \delta \cdot |s| = H(x_1',x_2') - \delta \cdot |s|.$$

Observing that $H_0(x_1',x_2')=H_0(x_1,x_2)$, it follows that

$$|G(x_1',x_2')-G(x_1,x_2)|=|H(x_1',x_2')-H(x_1,x_2)|\leq\delta\cdot|s|=\frac{\|(x_1'-x_1,x_2'-x_2)\|}{\sqrt{2}}\cdot\delta.$$

Estimate for $G|_{\mathcal{W}}$. In order to prove the estimate for $G|_{\mathcal{W}}$, where $\mathcal{W}$ is any line parallel to w, consider any (x_1,x_2) and (x_1',x_2') in such line $\mathcal{W}$ and set

$$s:=x_2'-x_2=-(x_1'-x_1),\qquad\text{so}\quad(x_1',x_2')=(x_1,x_2)+s(-1,1).$$

Observe that for any $z',z\in\mathbb{H}$ we have

$$\big|\big(h(z')-\mathrm{Im}(z')\big)-\big(h(z)-\mathrm{Im}(z)\big)\big|\leq\mathrm{Lip}(h-\mathrm{Im})\cdot|z'-z|\leq\delta\cdot|z'-z|$$

and thus $h(z')-\mathrm{Im}(z')\geq h(z)-\mathrm{Im}(z)-\delta\cdot|z'-z|$, which implies

$$h(z')\geq h(z)+\mathrm{Im}(z'-z)-\delta\cdot|z'-z|. \tag{B.2}$$

Consider now a parametrization in polar coordinates of the semicircles described by the geodesics γ and γ', i.e. for any $t\in\mathbb{R}$ let $\theta(t)\in[0,\pi]$ be the angle such that

$$\gamma(t)=\frac{x_1+x_2}{2}+\frac{x_2-x_1}{2}e^{i\theta(t)},$$
$$\gamma'(t)=\frac{x_1'+x_2'}{2}+\frac{x_2'-x_1'}{2}e^{i\theta(t)}.$$

Since, as remarked before the proof, γ' has the same center of γ but radius increased by s, if we set $\theta_0:=\theta(t_0)$ and $\theta_0':=\theta(t_0')$, we have that

$$\gamma'(t_0')-s\cdot e^{i\theta_0'}\in\gamma,\quad\text{and}\quad\gamma(t_0)+s\cdot e^{i\theta_0}\in\gamma'.$$

We claim that we must have

$$\sin\theta_0>1-\frac{2\delta}{l}\quad\text{and}\quad\sin\theta_0'>1-\frac{2\delta}{l}. \tag{B.3}$$

Indeed, recalling that $\|h-\mathrm{Im}\|_\infty\leq\delta$, we have

$$\begin{aligned}H(x_1,x_2)&=\max_{0<\theta<\pi}h\left(\frac{x_1+x_2}{2}+\frac{x_2-x_1}{2}e^{i\theta}\right)\\&\geq\max_{0<\theta<\pi}\mathrm{Im}\left(\frac{x_1+x_2}{2}+\frac{x_2-x_1}{2}e^{i\theta}\right)-\delta=\frac{x_2-x_1}{2}-\delta,\end{aligned}$$

but, if the first half of (B.3) fails, since $(x_1,x_2)\in\mathcal{U}_l$ and hence $x_2-x_1>l$, we have

$$\sin\theta_0\leq1-\frac{2\delta}{l}\leq1-\frac{4\delta}{x_2-x_1},$$

so that

$$\begin{aligned}H(x_1,x_2)&=h\left(\frac{x_1+x_2}{2}+\frac{x_2-x_1}{2}e^{i\theta_0}\right)\leq\mathrm{Im}\left(\frac{x_1+x_2}{2}+\frac{x_2-x_1}{2}e^{i\theta_0}\right)+\delta\\&=\frac{x_2-x_1}{2}\sin\theta_0+\delta\leq\frac{x_2-x_1}{2}-\delta,\end{aligned}$$

which is absurd. The same argument holds for θ_0' and proves the second part of (B.3).

Combining (B.2) and (B.3) we get

$$\begin{aligned}H(x_1',x_2')&=h\left(\gamma'(t_0')\right)\geq h\left(\gamma(t_0)+s\cdot e^{i\theta_0}\right)\\&\geq h\left(\gamma(t_0)\right)+\mathrm{Im}(se^{i\theta_0})-\delta\cdot|s|\\&\geq h\left(\gamma(x_1,x_2,t_0)\right)+s\cdot\left(1-\frac{2\delta}{l}\right)-\delta\cdot|s|\\&\geq H(x_1,x_2)+s-\left(1+\frac{2}{l}\right)\delta\cdot|s|.\end{aligned}$$

Similarly, one can also get

$$H(x_1,x_2)\geq H(x_1',x_2')-s-\left(1+\frac{2}{l}\right)\delta\cdot|s|.$$

Therefore, observing that $H_0(x_1', x_2') - H_0(x_1, x_2) = s$ it follows that

$$\begin{aligned}|G(x_1', x_2') - G(x_1, x_2)| &= \left|\big(H(x_1', x_2') - H(x_1, x_2)\big) - \big(H_0(x_1', x_2') - H_0(x_1, x_2)\big)\right| \\ &= \left|H(x_1', x_2') - \big(H(x_1, x_2) + s\big)\right| \\ &\le \left(1 + \frac{2}{l}\right)\delta \cdot |s| = \left(\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{l}\right)\delta \cdot |(x_1' - x_1, x_2' - x_2)|.\end{aligned}$$

This concludes the proof. □

Acknowledgements. This collaboration was made possible by the support of the ERC Starting Grant ChaParDyn. M. A. is supported by Unicredit Bank. C. U. is also supported by the Leverhulme Trust through a Leverhulme Prize. The research leading to these results has received funding from the European Research Council under the European Union Seventh Framework Program (FP/2007-2013) / ERC Grant Agreement n. 335989. L. M. is grateful to French CNRS and De Giorgi Center for financial and logistic support during his stay in Pisa.

References

[1] M. Artigiani, L. Marchese, and C. Ulcigrai, *The Lagrange spectrum of a Veech surface has a Hall ray*, Groups, Geometry, and Dynamics, **10** (2016), 1287–1337.
[2] S. Astels, *Cantor sets and numbers with restricted partial quotients*, Transactions of the American Mathematical Society, **352** (2000), 133–170.
[3] A. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, 91, Springer-Verlag, New York, 1995.
[4] M. Boshernitzan, *A condition for minimal interval exchange maps to be uniquely ergodic*, Duke Mathematical Journal, **52** (1985), 723–752
[5] M. Boshernitzan, and V. Delecroix, *From a packing problem to quantitative recurrence in* $[0, 1]$ *and the Lagrange spectrum of interval exchanges*, Discrete Analysis, **10** (2017), 1–25.
[6] R. Bowen, and C. Series, *Markov maps associated with Fuchsian groups*, Institut des Hautes Études Scientifiques. Publications Mathématiques, **50** (1979), 153–170.
[7] A. Cerqueira, C. Matheus, and C. G. Moreira, *Continuity of Hausdorff dimension across generic dynamical Lagrange and Markov spectra*, Journal of Modern Dynamics, **12** (2018), 151–174.
[8] T. W. Cusick, and M. E. Flahive, *The Markoff and Lagrange Spectra*, Mathematical Surveys and Monographs, 30, American Mathematical Society, Providence, RI, USA, 1989.
[9] A. Edmonds, J. Ewing, R. Kulkarni, *Torsion free subgroups of Fuchsian groups and tessellations of surfaces*, Inventiones Mathematicae, **69** (1982), 331–346.
[10] S. Ferenczi, *Dynamical generalizations of the Lagrange spectrum*, Journal d'Analyse Mathématique, **118** (2012), 19–53.
[11] L. Ford, *Fractions*, American Mathematical Monthly, **45** (1938), 586–601.
[12] R. Fox, *On Fenchel's Conjecture about F-Groups*, Matematisk Tidsskrift. B (1952), 61–65.
[13] G.A. Freiman, *Diophantine approximation and geometry of numbers (The Markoff spectrum)*, Kalininskii Gosudarstvennyi Universitet, Moscow, 1973.
[14] A. Haas, *Diophantine approximation on hyperbolic Riemann surfaces*, Acta Mathematica, **156** (1986), 33–82.
[15] A. Haas, and C. Series *The Hurwitz constant and Diophantine approximation on Hecke groups* Journal of the London Mathematical Society, **34** (1986), 219–334.
[16] M. Hall, *On the sum and products of continued fractions*, Annals of Mathematics (2), **48** (1947), 966–993.
[17] S. Hersonsky, and F. Paulin, *Diophantine approximation for negatively curved manifolds*, Mathematische Zeitschrift, **241** (2002), 181–226.
[18] S. Hersonsky, and F. Paulin, *On the almost sure spiraling of geodesics in negatively curved manifolds*, Journal of Differential Geometry, **85** (2010), 271–314.
[19] P. Hubert, S. Lelièvre, L. Marchese, and C. Ulcigrai, *The Lagrange spectrum of some square-tiled surfaces*, Israel Journal of Mathematics, **225** (2018), 553–607.
[20] P. Hubert, L. Marchese and C. Ulcigrai, *Lagrange spectra in Teichmüller dynamics via renormalization*, Geometric and Functional Analysis, **25** (2015), 180–255.
[21] A. Hurwitz, *Über die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche*, Mathematische Annalen, **39** (1891), 279–284.
[22] S. Katok, *Fuchsian groups*,Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992.
[23] J. Lehner, *A diophantine property of the Fuchsian groups*, Pacific Journal of Mathematics, **2** (1952), 327–333.
[24] J. Lehner, *A short course in automorphic functions*, Holt, Rinehart and Winston, New York, 1966.
[25] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Mathematische Annalen, **17** (1880), 379–399.
[26] C. Matheus, *The Lagrange and Markov spectra from the dynamical point of view*, preprint, `https://arxiv.org/abs/1703.01748`.
[27] F. Maucourant, *Sur les spectres de Lagrange et de Markoff des corps imaginaires quadratiques* Ergodic theory and Dynamical Systems, **23** (2003), 193–205.
[28] C. G. Moreira, *Geometric properties of the Markov and Lagrange spectra*, preprint, `http://preprint.impa.br/visualizar?id=5807`.

[29] C. G. Moreira, and J.C. Yoccoz, *Stable intersections of regular Cantor sets with large Hausdorff dimensions*, Annals of Mathematics, **54** (2001), 45–96.

[30] S. E. Newhouse, *The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms*, Institut des Hautes Études Scientifiques. Publications Mathématiques, **50** (1979), 101–151.

[31] J. Parkkonen, F. Paulin, *Diophantine approximation on hyperbolic surfaces*, Appendix to "Diophantine approximation in negatively curved manifolds and in the Heisenberg group", in Rigidity in dynamics and geometry, edited by M. Burger and A. Iozzi, Springer-Verlag, Berlin, 2002.

[32] J. Parkkonen, F. Paulin, *Prescribing the behaviour of geodesics in negative curvature*, Geometry & Topology, **14** (2010), 277–392.

[33] O. Perron, *Die Lehre von den Kettenbrüchen*, Chelsea, New York, 1950.

[34] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Mathematics, 149, Springer-Verlag, New York, 2006.

[35] T. A. Schmidt, and M. Sheingorn, *Riemann surfaces have Hall rays at each cusp*, Illinois Journal of Mathematics, **41** (1997), 378–397.

[36] C. Series, *The modular surface and continued fractions*, Journal of the London Mathematical Society, **31** (1985), 69–80.

[37] C. Series, *Geometrical methods of symbolic coding*, in "Ergodic theory, symbolic dynamics, and hyperbolic spaces", edited by T. Bedford, M. Keane, and C. Series, Oxford University Press, New York, 1991, 121–151.

[38] H. Shimizu, *On discontinuous groups operating on the product of upper half planes*, Annals of Mathematics, **77** (1963), 33–71.

[39] P. Tukia, *On discrete groups of the unit disk and their isomorphisms*, Annales AcademiæScientiarum Fennicæ, Series A I, **504** (1972), 1–45.

[40] L. Ya. Vulakh, *Diophantine approximation in* $\mathbb{R}^n$, Transactions of the American Mathematical Society, **347** (1995), 573–585.

[41] L. Ya. Vulakh, *Diophantine approximation on Bianchi groups*, Journal of Number Theory, **54** (1995), 73–80.

[42] L. Ya. Vulakh, *On Hurwitz constants for Fuchsian groups*, Canadian Journal of Mathematics, **49** (1997), 405–416.

[43] L. Ya. Vulakh, *The Markov spectra for triangle groups*, Journal of Number Theory, **67** (1997), 11–28.

[44] L. Ya. Vulakh, *The Markov spectra for Fuchsian groups*, Transactions of the American Mathematical Society, **352** (2000), 4067–4094.

Universidad de los Andes, Cra 1 # 18A - 12, Bogotá, 111711, Colombia.
E-mail address: m.artigiani@uniandes.edu.co

Université Paris 13, Sorbonne Paris Cité, LAGA, UMR 7539, 99 Avenue Jean-Baptiste Clément, 93430 Villetaneuse, France.
E-mail address: marchese@math.univ-paris13.fr

School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, United Kingdom.
E-mail address: corinna.ulcigrai@bristol.ac.uk