

ON THE EVOLUTION OF CONTINUED FRACTIONS IN A FIXED QUADRATIC FIELD

MENNY AKA AND URI SHAPIRA

ABSTRACT. We prove that the statistics of the period of the continued fraction expansion of certain sequences of quadratic irrationals from a fixed quadratic field approach the ‘normal’ statistics given by the Gauss-Kuzmin measure. As a by-product, the growth rate of the period is analyzed and, for example, it is shown that for a fixed integer k and a quadratic irrational α , the length of the period of the continued fraction expansion of $k^n \alpha$ equals $ck^n + o(k^{(1-\frac{1}{16})n})$ for some positive constant c . This improves results of Cohn, Lagarias, and Grisel, and settles a conjecture of Hickerson. The results are derived from the main theorem of the paper, which establishes an equidistribution result regarding single periodic geodesics along certain paths in the Hecke graph. The results are effective and give rates of convergence and the main tools are spectral gap (effective decay of matrix coefficients) and dynamical analysis on S -arithmetic homogeneous spaces.

1. INTRODUCTION

1.1. Continued fractions. The elementary theory of continued fractions starts by assigning to each real number $x \in [0, 1] \setminus \mathbb{Q}$ an infinite sequence of positive integers¹² referred to as the *continued fraction expansion* of x (abbreviated hereafter by c.f.e). Namely, to each number x corresponds a sequence $a_n = a_n(x)$, $n = 1, 2, \dots$ which is characterized by the requirement $x = \lim_{n \rightarrow \infty} \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_n}}}$. We refer to the numbers $a_n(x)$ as the digits of the c.f.e of x . When x is understood we usually write a_i for the i 'th digit of the c.f.e of x .

Given a number x , it is natural to ask for information regarding the statistical properties of its c.f.e; that is, for any finite sequence of natural numbers $w = (w_1, \dots, w_k)$ (referred to hereafter as a *pattern*) one is interested in the asymptotic *frequency of appearance* of the pattern w in the c.f.e of x , or in other words in the existence and the value of the limit

$$D(x, w) = \lim_N \frac{1}{N} \# \{1 \leq n \leq N : w = (a_{n+1}, \dots, a_{n+k})\}. \quad (1.1)$$

We claim that for Lebesgue almost any x the limit in (1.1) exists and equals some explicit integral (depending only on the pattern w).

¹We shall completely ignore the rational numbers, which correspond to finite sequences as well as real numbers outside the unit interval, for which an additional integer digit a_0 is needed.

²This correspondence is in fact a homeomorphism when $\mathbb{N}^{\mathbb{N}}$ is considered with the product topology.

To see this, note that the c.f.e correspondence $x \leftrightarrow \{a_n(x)\}$ fits in the commutative diagram

$$\begin{array}{ccc} \mathbb{N}^{\mathbb{N}} & \xrightarrow{\sigma} & \mathbb{N}^{\mathbb{N}} \\ \downarrow & & \downarrow \\ [0, 1] \setminus \mathbb{Q} & \xrightarrow{S} & [0, 1] \setminus \mathbb{Q}, \end{array} \quad (1.2)$$

where $S(x) = \{\frac{1}{x}\} = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ is the so-called Gauss map and σ is the shift map $\sigma(a_1, a_2, \dots) = (a_2, a_3, \dots)$. It is well known that S preserves the Gauss-Kuzmin measure on the unit interval which is given by

$$\nu_{\text{Gauss}}(A) \stackrel{\text{def}}{=} \frac{1}{\log(2)} \int_A \frac{1}{1+x} dx. \quad (1.3)$$

The map S is ergodic with respect to ν_{Gauss} which implies by the pointwise ergodic theorem (see for example [EW11, §2.6, §9.6]) that for ν_{Gauss} (or equivalently Lebesgue) almost any x and any pattern $w = (w_1, \dots, w_k)$, the frequency $D(x, w)$ defined in (1.1) exists. More precisely, if we let

$$I_w = \{x \in [0, 1] \setminus \mathbb{Q} : w = (a_1(x), \dots, a_k(x))\} \quad (1.4)$$

denote the interval consisting of those points for which the c.f.e starts with the pattern w , then the pointwise ergodic theorem tells us that the ergodic averages of the characteristic function of I_w converge almost surely to $\nu_{\text{Gauss}}(I_w)$; that is

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} \chi_{I_w}(S^i(x)) = \nu_{\text{Gauss}}(I_w), \quad (1.5)$$

for ν_{Gauss} -almost any x . As the set of possible patterns is countable we conclude that for Lebesgue almost any x (1.5) holds for any pattern w . It is straightforward to check using the commutation in (1.2) that the limit in (1.5) is equal to the limit in (1.1).

1.2. Quadratic irrationals. Let

$$\text{QI} \stackrel{\text{def}}{=} \{\alpha \in \mathbb{R} : [\mathbb{Q}(\alpha) : \mathbb{Q}] = 2\}$$

be the set of real quadratic irrationals. By Lagrange's Theorem (see for example [EW11, §3.3]) QI is characterized as the set of $x \in \mathbb{R}$ for which the c.f.e is eventually periodic. For quadratic irrationals (which clearly form a Lebesgue-null set) it is clear that the limit in (1.1) always exists and is different from the almost sure value of the frequency.

In this paper we investigate the behavior of $D(x, w)$ where x varies in some fixed quadratic field. We make the convention to consider $x \bmod 1$ instead of x . This influences only the 0'th digit in the classical discussion on continued fractions and does not effect any statistical property of the c.f.e. As will become clear shortly, our approach manages to deal with sequences $\{x_n\}$ whose elements are arithmetically related in a way that “involves only finitely many primes”.

Before preparing the grounds for more general statements we state Theorem 1.2 which demonstrates the flavor of our results regarding continued fractions. To the best of our

knowledge, all the results in the literature regarding the evolution of the period of the c.f.e of quadratics involve averaging. In this respect, the results we present are of a new kind.

Notation 1.1. Throughout this paper we use the notation \ll in the following manner: Given two quantities A, B depending on some set of parameters P , we denote $A \ll B$ if there exists some absolute constant $c > 0$ (independent of any varying parameter) such that $A \leq cB$. Given a subset P' of the parameter set P , we denote $A \ll_{P'} B$ if there exists a constant $c_{P'} > 0$, depending possibly on the parameters in P' , such that $A \leq c_{P'}B$.

We denote by $|I_w|$ the length of the interval I_w defined in (1.4). For $\alpha \in \text{QI}$ we denote by $|P_\alpha|$ the length of the period of the c.f.e of α . The following Theorem follows from Corollary 2.10 and Theorem 2.12 as explained in Remark 2.14.

Theorem 1.2. *Let $\alpha \in \text{QI}$, $k \in \mathbb{N}$ be given. Then, for any finite pattern of natural numbers $w = (w_1, \dots, w_\ell)$ we have that $D(k^n \alpha, w) \rightarrow \nu_{\text{Gauss}}(I_w)$ as $n \rightarrow \infty$. Moreover, there exists a constant $c_{\alpha, k}$ such that for any $n \in \mathbb{N}$ the following holds*

$$|D(k^n \alpha, w) - \nu_{\text{Gauss}}(I_w)| \ll_{\alpha, k} |I_w|^{-1} k^{-\frac{n}{32}}; \quad (1.6)$$

$$|P_{k^n \alpha}| = c_{\alpha, k} k^n + o_{\alpha, k}(1) k^{(1 - \frac{1}{16})n}. \quad (1.7)$$

1.3. Structure of the paper. The results presented in this paper are split into two; results regarding the distribution of closed geodesics and results regarding continued fractions. The gist of the paper is concerned with the distribution of certain closed geodesics in (the unit tangent bundle of) the modular surface and the results regarding continued fractions are translations of our results about geodesics utilizing the connection between the two. Although this connection is considered well understood, we believe that some of the results we present that allow this translation are new and may find further applications (e.g. Theorem 8.8).

Although the statements of our main results (Theorems 4.8, 8.9, 8.10) require quite a bit of preparation, some of their consequences are fairly easy to state (e.g. Theorem 1.2), and will hopefully motivate the reader traversing through the necessary preparations needed for the statements and proofs of the more general results.

In §2 we begin fixing the notation, state Theorems 2.8, 2.12 which deal with continued fractions, and present some examples and open problems. In §3 we fix further notation. In §4 we discuss the notions of *S-Hecke graphs* and *generalized branches* which play a key role in the statement of the main Theorem 4.8. In §5 we discuss the relationship of Theorem 4.8 to existing results and state Lemma 5.1. This Lemma explains to some extent the phenomenon behind our results but is only used in the proof of growth statements such as (1.7) and is not needed for the proof of statements such as (1.6). In §6 we prove our main result, Theorem 4.8, where the main tool in the argument is the decay of matrix coefficients. In §7 we give an elementary proof of Lemma 5.1. In §8 we prove our main results regarding continued fractions, Theorems 8.9, 8.10, and deduce Theorems 2.8, 2.12 which are stated in §2. Theorems 8.9, 8.10 are the translation to the language of continued fractions of Theorem 4.8 and Lemma 5.1. The technical tool we develop in order for this

translation to carry through is Theorem 8.8 which allows us to translate statements with an error term from the world of closed geodesics to the continued fractions world. Finally, in sections §9,10 we prove Theorem 8.8 elaborating on the classical connection between the geodesic flow and continued fractions.

2. SOME RESULTS AND OPEN PROBLEMS

2.1. Early preliminaries.

Definition 2.1. Given a commutative unital ring \mathcal{R} we let

$$\mathrm{GL}_2(\mathcal{R}) \stackrel{\mathrm{def}}{=} \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_{2 \times 2}(\mathcal{R}) : ad - bc \in \mathcal{R}^\times \right\}$$

and $\mathrm{PGL}_2(\mathcal{R}) = \mathrm{GL}_2(\mathcal{R})/Z$, where $Z \stackrel{\mathrm{def}}{=} \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \in \mathrm{GL}_2(\mathcal{R}) \right\}$ is the center of $\mathrm{GL}_2(\mathcal{R})$. When $\mathcal{R}_1 \hookrightarrow \mathcal{R}_2$, we have a natural embedding of $\mathrm{PGL}_2(\mathcal{R}_1) \hookrightarrow \mathrm{PGL}_2(\mathcal{R}_2)$. We usually abuse notation and treat the elements of $\mathrm{PGL}_2(\mathcal{R})$ as matrices rather than equivalence classes of matrices.

Recall that $\mathrm{PGL}_2(\mathbb{R})$ acts on the real line by Möbius transformations; for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PGL}_2(\mathbb{R})$ and $x \in \mathbb{R}$, $gx \stackrel{\mathrm{def}}{=} \frac{ax+b}{cx+d}$. Recall the following basic result [RS92, Theorem 2]

Theorem 2.2. *For any $x \in \mathbb{R}$ the orbit of x under $\mathrm{PGL}_2(\mathbb{Z})$ is exactly the set of numbers having c.f.e with the same tail as the c.f.e of x . Equivalently, $\mathrm{PGL}_2(\mathbb{Z})x = \{y \in \mathbb{R} : \exists m, n > 0 \forall i \geq 0 a_{n+i}(y) = a_{m+i}(x)\}$.*

As the c.f.e of $\alpha \in \mathrm{QI}$ is eventually periodic, it follows from (1.2) that the orbit $\{S^n \alpha\}_{n \in \mathbb{N}}$ of α under the Gauss map is eventually periodic.

Definition 2.3. Let $\alpha \in \mathrm{QI}$. We denote by P_α the period of $\alpha \bmod 1$ under the Gauss map; that is, $P_\alpha \stackrel{\mathrm{def}}{=} \{x_1, \dots, x_\ell\} \subset [0, 1]$ where for some $n \geq 0$, $S^n(\alpha \bmod 1) = x_1$ and, $S(x_i) = x_{i+1}$ for all $i < \ell$ and $S(x_\ell) = x_1$. We denote by ν_α the normalized counting measure on $[0, 1]$ supported on the period P_α .

Let $\iota : \mathbb{R} \rightarrow \mathrm{PGL}_2(\mathbb{Z}) \backslash \mathbb{R}$ be the quotient map to the ‘set of orbits’. By Theorem 2.2, for any $\alpha \in \mathrm{QI}$ we have that

$$\text{For any } \beta \in \iota(\alpha), P_\alpha = P_\beta, \nu_\alpha = \nu_\beta. \tag{2.1}$$

We sometimes write $\nu_{\iota(\alpha)}, P_{\iota(\alpha)}$ when we wish to stress this fact.

Lemma 2.4. *Let $\alpha \in \mathrm{QI}$, for any pattern w , the frequency of appearance $D(\alpha, w)$ of the pattern w in the c.f.e of α equals $\nu_\alpha(I_w) = \frac{|P_\alpha \cap I_w|}{|P_\alpha|}$.*

Proof. Let (a_1, \dots, a_k) be the period of the c.f.e of α . By (1.2), $P_\alpha = \{x_1, \dots, x_k\}$ where $x_i \in [0, 1]$ is the number whose c.f.e is given by the infinite concatenation of the pattern $(a_i, \dots, a_k, a_1, \dots, a_{i-1})$. The statement of the Lemma now follows easily from (1.1) and (1.4). \square

Definition 2.5. Given a finite set of primes S we denote by $\mathcal{O}_S \stackrel{\text{def}}{=} \mathbb{Z}[p^{-1} : p \in S]$ the ring of S -integers. We denote by $\mathcal{O}_S^\times \stackrel{\text{def}}{=} \left\{ \pm \prod_{p \in S} p^{n_p} : n_p \in \mathbb{Z} \right\}$ the group of S -units; that is, the group of invertible elements in \mathcal{O}_S .

There is a natural embedding $\mathcal{O}_S^\times \hookrightarrow \text{PGL}_2(\mathcal{O}_S)$ given by $q \mapsto \text{diag}(q, 1)$. We denote by $\gamma_q = \text{diag}(q, 1)$ the image of q under this embedding. Note that $q\alpha = \gamma_q\alpha$.

Definition 2.6. For $\gamma \in \text{PGL}_2(\mathbb{Q})$ let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_{2 \times 2}(\mathbb{Z})$ be the unique representative of γ with co-prime entries. We define the *height* of γ to be $\text{ht}(\gamma) \stackrel{\text{def}}{=} |\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}|$. Given a rational number $q = \pm \prod_1^\ell p_i^{e_i}$, where the p_i 's are distinct primes and $e_i \in \mathbb{Z}$, we define the *height* of q to be $\text{ht}(q) \stackrel{\text{def}}{=} \text{ht}(\gamma_q) = \prod_1^\ell p_i^{|e_i|}$.

For $\gamma \in \text{PGL}_2(\mathbb{Q})$, $\text{ht}(\gamma)$ measures how far γ is from $\text{PGL}_2(\mathbb{Z})$. As the $\text{PGL}_2(\mathbb{Z})$ action does not change the period, it is natural to expect that a statement regarding the evolution of $\nu_{\gamma\alpha}$ will depend on $\text{ht}(\gamma)$. This is indeed the case as will be seen shortly.

2.2. Results. Our results are concerned with the convergence $\nu_{\gamma\alpha} \rightarrow \nu_{\text{Gauss}}$ and the growth of the length of the period $|P_{\gamma\alpha}|$ as $\text{ht}(\gamma) \rightarrow \infty$ and $\gamma \in \text{PGL}_2(\mathcal{O}_S)$ for a fixed finite set of primes S and $\alpha \in \text{QI}$. We give estimates on error terms and so refer to our results as *effective*. In these estimates there appears an exponent $\frac{25}{64} \leq \delta_0 \leq \frac{1}{2}$ whose exact value is not known (although according to the Ramanujan conjecture $\delta_0 = \frac{1}{2}$). The bigger it is the stronger the statements are and the best known lower bound for it to this date is $\delta_0 \geq \frac{25}{64}$; a bound given by Kim and Sarnak in the appendix of [Kim03]³.

Naturally, our results involve comparison of integrals with respect to measures which are mutually singular, and in order to make sense of an error term we need to restrict our attention to integrals of functions with some controlled behavior. This is usually done by looking at smooth functions and considering Sobolev norms. We choose to work with the more primitive notion of Lipschitz functions.

Definition 2.7. Let (X, d) be a metric space. For any $\kappa > 0$ we denote by $\text{Lip}_\kappa(X)$ the space of Lipschitz functions $f : X \rightarrow \mathbb{C}$ with κ as a Lipschitz constant. We sometimes refer to such functions as κ -Lipschitz.

The following Theorem is deduced from Theorem 8.9 in §8.

Theorem 2.8. *Let S be a finite set of primes, $\alpha \in \text{QI}$.*

- (1) *If $\{q_n\} \subset \mathcal{O}_S^\times$ is a sequence such that $\text{ht}(q_n) \rightarrow \infty$ then $\nu_{q_n\alpha} \rightarrow \nu_{\text{Gauss}}$. More precisely, given $\epsilon > 0$, $q \in \mathcal{O}_S^\times$, and $f \in \text{Lip}_\kappa([0, 1])$ the following estimate holds*

$$\left| \int_0^1 f d\nu_{\text{Gauss}} - \int_0^1 f d\nu_{q\alpha} \right| \ll_{\alpha, S, \epsilon} \max \{ \|f\|_\infty, \kappa \} \text{ht}(q)^{-\frac{\delta_0}{6} + \epsilon}. \quad (2.2)$$

- (2) *Assume that all the primes in S do not split in the extension $\mathbb{Q}(\alpha)$ of \mathbb{Q} . Then, if $\{\gamma_n\} \subset \text{PGL}_2(\mathcal{O}_S)$ is a sequence such that $\text{ht}(\gamma_n) \rightarrow \infty$ then $\nu_{\gamma_n\alpha} \rightarrow \nu_{\text{Gauss}}$. More*

³This parameter relates to the representation theory of GL_2 (see §6.3).

precisely, given $\epsilon > 0$, $\gamma \in \mathrm{PGL}_2(\mathcal{O}_S)$, and $f \in \mathrm{Lip}_\kappa([0, 1])$ the following estimate holds

$$\left| \int_0^1 f d\nu_{\mathrm{Gauss}} - \int_0^1 f d\nu_{\gamma\alpha} \right| \ll_{\iota(\alpha), S, \epsilon} \max \{ \|f\|_\infty, \kappa \} \mathrm{ht}(\gamma)^{-\frac{\delta_0}{6} + \epsilon}, \quad (2.3)$$

- (3) If one of the primes in S splits in the extension $\mathbb{Q}(\alpha)$ of \mathbb{Q} , then there exists sequences $q_n \in \mathcal{O}_S^\times, \gamma_n \in \mathrm{PGL}_2(\mathbb{Z})$ with $\mathrm{ht}(q_n) \rightarrow \infty$ such that $\nu_{q_n \gamma_n \alpha}$ does not converge to ν_{Gauss} and in particular, the implicit constant in (2.2) cannot be taken to be uniform on the orbit $\iota(\alpha)$ in contrast with (2.3)⁴.

Remark 2.9. (1) Under the assumption that no prime in S splits in $\mathbb{Q}(\alpha)$, (2.2) follows from (2.3) by choosing $\gamma = \gamma_q$.

- (2) It is an exercise to show that any $\gamma \in \mathrm{PGL}_2(\mathcal{O}_S)$ can be written as a product $\gamma = \gamma_1 \gamma_q \gamma_2$, where $\gamma_i \in \mathrm{PGL}_2(\mathbb{Z})$ and $q \in \mathcal{O}_S^\times$ (see the proof of Corollary 4.2). It now follows from (2.1) that $\nu_{\gamma\alpha} = \nu_{q\beta}$ where $\beta = \gamma_2\alpha$, and so although it seems more restrictive at first glance, instead of studying the evolution of $\nu_{\gamma\alpha}$ as $\gamma \in \mathrm{PGL}_2(\mathcal{O}_S)$, it is enough to consider the evolution of $\nu_{q\beta}$ as $q \in \mathcal{O}_S^\times, \beta \in \iota(\alpha)$.

When we use Theorem 2.8 to try and estimate the frequency of a pattern in the period of the c.f.e of $\gamma\alpha$ we obtain the following

Corollary 2.10. *Let S be a finite set of primes and $\alpha \in \mathrm{QI}$. For any finite pattern $w = (w_1 \dots w_k)$ of digits, and any $q \in \mathcal{O}_S^\times$*

$$|D(q\alpha, w) - \nu_{\mathrm{Gauss}}(I_w)| \ll_{\alpha, S, \epsilon} |I_w|^{-1} \mathrm{ht}(q)^{-\frac{\delta_0}{12} + \epsilon}. \quad (2.4)$$

Moreover, if all the primes in S do not split in the extension $\mathbb{Q}(\alpha)$ of \mathbb{Q} , then for any $\gamma \in \mathrm{PGL}_2(\mathcal{O}_S)$

$$|D(\gamma\alpha, w) - \nu_{\mathrm{Gauss}}(I_w)| \ll_{\iota(\alpha), S, \epsilon} |I_w|^{-1} \mathrm{ht}(\gamma)^{-\frac{\delta_0}{12} + \epsilon}. \quad (2.5)$$

Proof. By Lemma 2.4, $D(\beta, w) = \nu_{\iota(\beta)}(I_w)$ for any $\beta \in \mathrm{QI}$. The exponent in (2.2) (resp. (2.3)) is cut in half in (2.4) (resp. (2.5)) as a result of the fact that χ_{I_w} is not Lipschitz and one needs to use an approximation of it in order to apply Theorem 2.8. We leave the details to the reader. \square

Theorem 2.8 raises a natural question: Is it true that $\nu_{q_n \alpha} \rightarrow \nu_{\mathrm{Gauss}}$ for any sequence of rationals q_n with $\mathrm{ht}(q_n) \rightarrow \infty$? The following example which was essentially communicated to us by A. Ubis shows that the answer is negative and so the assumption that $q_n \in \mathcal{O}_S$ for a fixed finite set of primes S is crucial.

Example 2.11. Let D be a fundamental discriminant such that the negative Pell equation $x^2 - Dy^2 = -1$ has an integer solution (see [Lag80], [FK10] for example). A solution $x = k_1, y = n_1$ to the equation corresponds to a unit $\epsilon_1 \stackrel{\mathrm{def}}{=} k_1 + n_1 \sqrt{D}$ in the ring $\mathbb{Z}(\sqrt{D})$ of norm -1 and in turn, the odd powers $\epsilon_1^j = k_j + n_j \sqrt{D}$ give rise to infinitely many further solutions of the negative Pell equation. For odd j let α_j solve the equation $x = 2k_j + \frac{1}{x}$.

⁴In fact, it is possible to choose q_n so that $\nu_{q_n \gamma_n \alpha}$ is a constant sequence.

That is, the c.f.e of α_j is purely periodic with period of length 1 of digit $2k_j$ (note that here we abuse the notation introduced above and we do record the 0 digit). Solving for x in the above equation we see that α_j could be chosen to be $k_j + \sqrt{k_j^2 + 1}$. As (k_j, n_j) solve the negative Pell equation for D we get $\alpha_j = k_j + n_j\sqrt{D}$ which shows that the measures $\nu_{n_j\sqrt{D}}$ are not converging to the Gauss-Kuzmin measure and in fact are atomic measures supported on single points.

In §8 we prove Theorem 8.10 which discusses the length of the period of the c.f.e of $\gamma\alpha$, where $\alpha \in \text{QI}$ is fixed and $\gamma \in \text{PGL}_2(\mathcal{O}_S)$ varies. We state below Theorem 2.12 which is an adaptation of Theorem 8.10 that uses only the terminology presented so far. It is deduced from Theorem 8.10 in §8. Theorem 2.12 solves a conjecture of Hickerson [Hic73, p. 429] and strengthens [Coh77, Theorem 3]. Note that by the argument in Lemma 2.4, for any $\alpha \in \text{QI}$, $|P_\alpha|$ is the *length of the period of the c.f.e of α* .

In the Appendix of [Lag80], using the methods of Dirichlet [Dir56], Lagarias shows that, under some restrictive assumptions on α , one has that for any integer k there exists a constant C for which, $C\frac{k^n}{n} < |P_{k^n\alpha}|$. Under some restrictive assumptions on α , Grisel [Gri98] proved a stronger estimate of the form $C_1k^n \leq |P_{k^n\alpha}| \leq C_2k^n$. The following Theorem strengthens these results in several respects.

Theorem 2.12. *Let S be a finite set of primes. There exists a positive function $c(\alpha, \gamma)$ on the set $\text{QI} \times \text{PGL}_2(\mathcal{O}_S)$ satisfying the following: For any $\alpha \in \text{QI}$,*

- (1) (a) *For any $\epsilon > 0$, and $q \in \mathcal{O}_S^\times$*

$$|P_{q\alpha}| = c(\alpha, \gamma_q) \text{ht}(\gamma_q) + O_{\alpha, S, \epsilon}(1) \text{ht}(\gamma_q)^{1 - \frac{\delta_0}{6} + \epsilon}. \quad (2.6)$$

Moreover, if all the primes in S do not split in the quadratic field $\mathbb{Q}(\alpha)$, then for any $\epsilon > 0$, $\gamma \in \text{PGL}_2(\mathcal{O}_S)$

$$|P_{\gamma\alpha}| = c(\alpha, \gamma) \text{ht}(\gamma) + O_{\iota(\alpha), S, \epsilon}(1) \text{ht}(\gamma)^{1 - \frac{\delta_0}{6} + \epsilon}. \quad (2.7)$$

- (b) *The function c attains only finitely many values on \mathcal{O}_S^\times ; that is,*

$$|\{c(\alpha, \gamma_q) : q \in \mathcal{O}_S^\times\}| < \infty.$$

- (c) *If $q_n = \ell_1^{(n)}/\ell_2^{(n)}$, where $\ell_i^{(n)} \in \mathcal{O}_S^\times \cap \mathbb{N}$ satisfies $\ell_i^{(n)} | \ell_i^{(n+1)}$ for $i = 1, 2$, then $c(\alpha, \gamma_{q_n})$ stabilizes.*

- (2) $\sup \{c(\alpha, \gamma) : \gamma \in \text{PGL}_2(\mathcal{O}_S)\} \ll_{\iota(\alpha), S} 1$.

- (3) *All the primes in S do not split in the quadratic field $\mathbb{Q}(\alpha)$ if and only if*

$$\inf \{c(\alpha, \gamma) : \gamma \in \text{PGL}_2(\mathcal{O}_S)\} > 0. \quad (2.8)$$

As an immediate corollary we have for example the following

Corollary 2.13. *For any $\alpha \in \text{QI}$ and any positive integer k , $\lim_n \frac{|P_{k^n\alpha}|}{k^n}$ exists and is a positive real number.*

Remark 2.14. The first part of Theorem 1.2 is obtained from Corollary 2.10 by taking $q = k^n$, the Kim-Sarnak exponent $\delta_0 = \frac{25}{64}$, and choosing $\epsilon = \frac{1}{768}$ so that $-\frac{\delta_0}{12} + \epsilon = -\frac{1}{32}$. The second part of Theorem 1.2 is obtained similarly from Theorem 2.12(1a).

2.3. References to existing results. Although the question of the evolution of the c.f.e along arithmetically defined sequences in a fixed quadratic field is extremely natural, we did not find too many relevant papers to cite. Some earlier works studying the statistics of the period ‘in average’ (and also not in a fixed field), were initiated by Arnold (see [Arn08],[Arn07],[Ler10] and the references therein). See also [Pol86]. Other works, mostly related to the length of the period, which the reader might find related, may be found for example in many of the papers of Golubeva (such as [Gol02]) and in [Gri98],[BL05],[MF93],[CZ04],[Coh77],[Hic73],[Kei]. Standing out in this context is the recent paper of McMullen which provides examples of sequences of quadratic irrationals in a fixed quadratic field with uniformly bounded c.f.e digits [McM09]. We suspect that it should be very interesting to compare in detail how McMullen’s results fit together with the results of the present paper.

As for results regarding periodic geodesics the situation is completely different and we will not attempt to summarize the relevant results that appeared in the literature. We comment though, that as will be explained in §5, our main Theorem 4.8 is closely related to the work of Benoist and Oh [BO07].

2.4. Some open problems. We list below a few questions which emerge from our discussion and remain unsolved. Each of the problems below have a corresponding problem stated in terms of periodic geodesics on the modular surface.

- (1) Give satisfactory sufficient conditions on a sequence of rationals q_n to ensure that for a quadratic irrational α , the sequence of measures $\nu_{q_n\alpha}$ equidistribute to the Gauss-Kuzmin measure ν . It might be interesting to replace the quantifiers and allow the conditions to depend on α .
- (2) Is it true that for a quadratic irrational α which is not a unit in the ring of integers of $\mathbb{Q}(\alpha)$, the sequence of measures ν_{α^n} always equidistribute to the Gauss-Kuzmin measure along the subsequence of n ’s for which α^n is irrational (see [CZ04]). Note that our results deal with the case $\alpha = \sqrt{d}$.
- (3) Let p_n be an enumeration of the primes. Are there any quadratic irrationals α for which $\nu_{p_n\alpha}$ equidistribute to the Gauss-Kuzmin measure.
- (4) Is it true that for any quadratic irrational α there exist a sequence of distinct primes p_n so that $\nu_{p_n\alpha}$ equidistribute to the Gauss-Kuzmin measure.

2.5. Acknowledgments. We benefited from numerous stimulating conversations with Manfred Einsiedler, Elon Lindenstrauss and various other people which we try to name below. We would like to express our gratitude and appreciation to them for the generosity in which they exchange valuable ideas and create an ‘open source’ healthy environment in the community doing Homogeneous Dynamics. Thanks are also due to Yann Bugeaud, Étienne Fouvry, Tsachik Gelander, Alex Gorodnik, Chen Meiri, Philippe Michel, Shahar Mozes, Hee Oh, Peter Sarnak, Adrian Ubis, Akshay Venkatesh, and Barak Weiss.

Both Authors enjoyed the warm hospitality of the Centre Interfacultaire Bernoulli and the GANT semester held there.

M.A acknowledges the support of the Advanced research Grant 226135 from the European Research Council, the ISEF foundation, the Ilan and Asaf Ramon memorial foundation, and the Hoffman Leadership and Responsibility fellowship program at the Hebrew University of Jerusalem.

U.S acknowledges the support of the Advanced research Grant 228304 from the European Research Council.

3. PRELIMINARIES

In this section we fix the notation that we will use in §4–§8. In §9,10 our notation will slightly vary as will be explained at the beginning of §9.

For a prime p we let \mathbb{Q}_p denote the field of p -adic numbers and let \mathbb{Z}_p be the ring of p -adic integers. We sometimes denote $\mathbb{Q}_\infty = \mathbb{R}$. Let $\mathbf{P} \stackrel{\text{def}}{=} \{p \in \mathbb{N} : p \text{ is a prime}\}$. Given $S \subset \mathbf{P}$ we denote $S^* = S \cup \{\infty\}$. The set \mathbf{P}^* will be referred to as the set of *places* of \mathbb{Q} – the primes being the *finite places*.

Let $S \subset \mathbf{P}^*$ be given. Throughout $S_f = S \setminus \{\infty\}$. We denote by $\mathbb{Q}_S, \mathbb{Z}_S$ the product rings $\prod_{v \in S} \mathbb{Q}_v, \prod_{v \in S} \mathbb{Z}_v$ respectively (the latter makes sense only when $\infty \notin S$). Let \mathbb{G} denote the algebraic group PGL_2 . We denote $G_S = \mathbb{G}(\mathbb{Q}_S)$. We denote an element $g \in G_S$ by a sequence $g = (g_v)_{v \in S}$ where g_v is a 2×2 matrix over \mathbb{Q}_v . Keep in mind the slight abuse of notation arising from the fact that g_v is in fact an equivalence class of matrices. If $\infty \in S$ we usually abbreviate and write $g = (g_\infty, g_f)$ where g_f denotes the tuple of the components corresponding to the finite places in S . The identity elements in the various groups are denoted by e with the corresponding subscript. Thus for example $e_S = (e_v)_{v \in S}$ and if $\infty \in S$, $e_S = (e_\infty, e_f)$.

Hereafter $S \subset \mathbf{P}$. We may view the group $\Gamma_S \stackrel{\text{def}}{=} \mathbb{G}(\mathcal{O}_S)$ as a subgroup of $G_{S^*} = G_\infty \times \prod_{v \in S} G_v$ (embedded diagonally). When $S = \emptyset$, $\mathcal{O}_S = \mathbb{Z}$ and we denote $\Gamma_S = \mathbb{G}(\mathbb{Z})$ by Γ_∞ . It is well known that Γ_S is a lattice in G_{S^*} . We set

$$X_S \stackrel{\text{def}}{=} \Gamma_S \backslash G_{S^*}, \quad X_\infty \stackrel{\text{def}}{=} \Gamma_\infty \backslash G_\infty.$$

The gist of our discussion will be concerned with these homogeneous spaces. We denote by m_S (resp. m_∞) the G_{S^*} -invariant (resp. G_∞ -invariant) probability measure on X_S (resp. X_∞). The *real quotient* X_∞ is a *factor* of X_S in a natural way: Let $K_\infty \stackrel{\text{def}}{=} \text{PO}_2(\mathbb{R})$ denote the maximal compact subgroup of G_∞ . For a finite place $p \in \mathbf{P}$ we let $K_p = \mathbb{G}(\mathbb{Z}_p)$. We then let $K_S \stackrel{\text{def}}{=} \prod_{v \in S} K_v$. As we will explain shortly, the double coset space $X_S/K_S = \Gamma_S \backslash G_{S^*}/K_S$ is naturally identified with X_∞ . We denote by $\pi : X_S \rightarrow X_\infty$ the natural projection. This identification relies on two facts (i) $G_S = \Gamma_S K_S$, and (ii) $\Gamma_\infty = \Gamma_S \cap K_S$. Relying on these facts the identification is as follows: Given a double coset $\Gamma_S(g_\infty, g_f)K_S$, by (i) we may assume without loss of generality that $g_f = e_f$ and identify this double coset with $\Gamma_\infty g_\infty \in X_\infty$. The reader will easily check that (ii) implies that this map is indeed well defined and bijective. We leave the verification of conditions (i),(ii) to the reader ((ii) is straightforward and an argument similar to that giving (i) may be extracted from the proof of Lemma 6.9 for example).

Remark 3.1. In practice, given $x = \Gamma_S(g_\infty, g_f) \in X_S$ with representative (g_∞, g_f) such that $g_f \in K_S$, the projection $\pi(x)$ is $\Gamma_\infty g_\infty$. In other words, $\pi^{-1}(\Gamma_\infty g_\infty) = \{\Gamma_S(g_\infty, g_f) : g_f \in K_S\}$. Another useful observation to keep in mind here is that two points $x_1 = \Gamma_S(g_\infty, g_f), x_2 = \Gamma_S(g_\infty, h_f)$ are in the same fiber (that is $\pi(x_1) = \pi(x_2)$) if and only if the quotient $g_f^{-1}h_f$ belongs to K_S .

The group G_{S^*} (and all its subgroups) act on X_S by right translation. In particular, if $T \subset S^*$, we may view G_T (and its subgroups) as a subgroup of G_{S^*} and thus it acts on X_S . Note that $\pi : X_S \rightarrow X_\infty$ intertwines the G_∞ -actions. Of particular interest to us will be the action of the real diagonal group $A_\infty \stackrel{\text{def}}{=} \{\text{diag}(e^t, 1) : t \in \mathbb{R}\}$, the elements of which we often write as $a_\infty(t) = \text{diag}(e^t, 1)$.

We say that an orbit xL of a closed subgroup $L < G_{S^*}$ through a point $x \in X_S$ is *periodic* if it supports an L -invariant probability measure. Such a measure is unique and we refer to it as the Haar measure on the periodic orbit. Compact orbits are always periodic. Given a measure μ on X_S and $g \in G_{S^*}$ we let $g_*\mu$ denote the pushed forward measure by right translation by g . This notation is a bit awkward as $(gh)_*\mu = h_*(g_*\mu)$. This will not bother us as we will only use commutative subgroups to push measures.

For $v \in \mathbf{P}^*$, the Lie algebra of G_v will be denoted by \mathfrak{g}_v and is naturally identified with the space of traceless 2×2 matrices over \mathbb{Q}_v . Similarly to the notation introduced above we will denote by $\mathfrak{g}_{S^*} = \mathfrak{g}_\infty \oplus_{v \in S} \mathfrak{g}_v$ the Lie algebra of G_{S^*} . A basic fact that we will use is that if S is finite and $L < G_S$ is a closed subgroup then L contains an open product subgroup $\prod_{v \in S} L_v$ which allows us to speak of the Lie algebra of L which will be denoted $\text{Lie}(L)$. The exponential map $\exp_v : \mathfrak{g}_v \rightarrow G_v$ is defined for any place v by the usual power series and in fact, is only well defined for finite places on a certain neighborhood of 0. We denote its inverse by \log_v (it is defined on a small enough neighborhood of e_v) and use the obvious notation $\exp_S, \exp_{S^*}, \log_S, \log_{S^*}$ to denote the corresponding product maps from the corresponding product domains in $\mathfrak{g}_S, \mathfrak{g}_{S^*}, G_S, G_{S^*}$ respectively.

Given an element $g \in G_{S^*}$ and an element u (either of G_{S^*} or of \mathfrak{g}_{S^*}), we denote

$$u^g \stackrel{\text{def}}{=} g^{-1}ug.$$

If g is semisimple we denote by $(\mathfrak{g}_{S^*})_g^{\text{ws}}$ the *weak stable* subalgebra of \mathfrak{g}_{S^*} . It is defined as the direct sum of the eigenspaces (of the operator $u \mapsto u^g$) of modulus ≤ 1 or equivalently

$$(\mathfrak{g}_{S^*})_g^{\text{ws}} = \left\{ u \in \mathfrak{g}_{S^*} : \{u^{(g^n)}\}_{n>0} \text{ is bounded in } \mathfrak{g}_{S^*} \right\}.$$

For each place v we equip G_v, \mathfrak{g}_v with metrics in the following way: For $v = \infty$ we start with an inner product on \mathfrak{g}_∞ which is right K_∞ -invariant and use left translation to make it into a left invariant Riemannian metric on G_∞ which is also right K_∞ -invariant. This Riemannian metric induces left G_∞ -invariant, bi- K_∞ -invariant metric on G_∞ . For a finite place v , we start with a bi- K_v -invariant metric d_{K_v} on K_v (such that K_v equals the closed unit ball around e_v) and make it into a left invariant metric on G_v (which is also right K_v -invariant) by setting $d_{G_v}(g_1, g_2) = 2$ if $g_1^{-1}g_2 \notin K_v$ and $d_{G_v}(g_1, g_2) = d_{K_v}(g_1^{-1}g_2, e_v)$ otherwise. On the Lie algebra \mathfrak{g}_v we take the metric given by $d_{\mathfrak{g}_v}(u, w) = \max\{|u_{ij} - w_{ij}|_v : 1 \leq i, j \leq 2\}$ where the indices i, j stand for the

entries of the corresponding matrix. We usually denote the distance from $0 \in \mathfrak{g}_v$ by $\|u\| \stackrel{\text{def}}{=} d_{\mathfrak{g}_v}(u, 0)$ and refer to it as the *norm* of u . We define the metrics $d_{G_{S^*}}, d_{\mathfrak{g}_{S^*}}$ on $G_{S^*}, \mathfrak{g}_{S^*}$ respectively by taking the maximum of the metrics defined above over the places in S^* . The metric $d_{G_{S^*}}$ induces a right- $K_\infty \times K_S$ -invariant metric on X_S by setting $d_{X_S}(\Gamma_S g_1, \Gamma_S g_2) = \inf_{\gamma \in \Gamma_S} d_{G_{S^*}}(\gamma g_1, g_2)$.

In a metric space (X, d_X) we denote $B_r^X(x)$ the open ball of radius r around x . In case the space is a group, we denote by B_r^X the corresponding ball around the trivial element.

4. THE S -HECKE GRAPH AND THE MAIN THEOREM

Throughout this section we use the notation introduced in §3. We fix a finite set of finite places $S \subset \mathbf{P}$. The space X_∞ can be thought of as the moduli space of equivalence classes of 2-dimensional lattices in the plane \mathbb{R}^2 up to homothety. We will refer below to a point $x \in X_\infty$ as a class; here, the class $\Gamma_\infty g$ is composed of the lattice spanned by the rows of the matrix g (which is well defined up to scaling) and all its homotheties.

Our first aim is to state Theorem 4.8. Briefly, we will fix a class x with periodic A_∞ -orbit and consider a class x' on the S -Hecke graph (soon to be defined) through x and prove an effective equidistribution statement regarding the periodic orbit $x' A_\infty$ as x' drifts away from x in the graph.

4.1. Hecke friends. Given a class $x \in X_\infty$, we say that a class x' is a *Hecke friend* of x if one can choose lattices $\Lambda_x \in x, \Lambda_{x'} \in x'$ such that $\Lambda_{x'} < \Lambda_x$. After fixing the lattice Λ_x there is a unique choice of $\Lambda_{x'} \in x'$ such that $\Lambda_{x'} < \Lambda_x$ is *primitive*; that is, such that the index $[\Lambda_x : \Lambda_{x'}]$ is minimal. We denote this minimal index by $\text{ind}(x, x')$. We say that x' is an *S -Hecke friend* of x if $\text{ind}(x, x') \in \mathcal{O}_S^\times$. It is elementary to check that the S -Hecke friendship relation is an equivalence relation and that furthermore, if x, x' are Hecke friends then $\text{ind}(x, x') = \text{ind}(x', x)$.

4.2. The graph. For a class $x \in X_\infty$ we define

$$\mathcal{G}_S(x) = \{x' \in X_\infty : x, x' \text{ are } S\text{-Hecke friends}\} \quad (4.1)$$

The set $\mathcal{G}_S(x)$ has the structure of a graph⁵: We join $x_1, x_2 \in \mathcal{G}_S(x)$ with an edge if there exists $\Lambda_i \in x_i$ such that Λ_1 is a sublattice of Λ_2 of index p for some $p \in S$ (note that as p is prime this forces Λ_1 to be a primitive sublattice of Λ_2). In this case we declare the length of this edge to be $\log(p)$. This induces a distance function on the graph which we denote $d_{\mathcal{G}}(\cdot, \cdot)$ for which $d_{\mathcal{G}}(x_1, x_2) = \log(\text{ind}(x_1, x_2))$. We will refer to x as the *root* of $\mathcal{G}_S(x)$ and call $\mathcal{G}_S(x)$ the *S -Hecke graph through x* . Note that $\{\text{ind}(x, x') : x' \in \mathcal{G}_S(x)\} = \mathcal{O}_S^\times \cap \mathbb{N}$. We refer to numbers in $\mathcal{O}_S^\times \cap \mathbb{N}$ as *admissible radii* and denote for $h \in \mathcal{O}_S^\times \cap \mathbb{N}$ by $\mathcal{S}_h(x) \stackrel{\text{def}}{=} \{x' \in \mathcal{G}_S(x) : \text{ind}(x, x') = h\}$ the *sphere* of radius h around the root x .

⁵When S contains only one prime, this is the well known *p -Hecke tree* through x . In general, this graph is the product of the various p -Hecke trees for $p \in S$.

4.3. **The sphere.** For $q \in \mathcal{O}_S^\times$ let us define

$$a_f(q) \stackrel{\text{def}}{=} \left(e_\infty, \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}, \dots, \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix} \right) \in G_S. \quad (4.2)$$

Given $x \in X_\infty$ we wish to have a convenient algebraic description of the classes on the sphere $\mathcal{S}_h(x)$ for admissible radii h . We obtain this description using the extension $\pi : X_S \rightarrow X_\infty$ in the following way: Lemma 4.1 below shows that the various points on $\mathcal{S}_h(x)$ are obtained by choosing a lift $y \in \pi^{-1}(x)$ of x , and projecting $ya_f(h)$ via π back to X_∞ .

Lemma 4.1. *For $x = \Gamma_\infty g \in X_\infty$ and h an admissible radius we have*

$$\begin{aligned} \mathcal{S}_h(x) &= \pi(\{\Gamma_S(g, \gamma) : \gamma \in \Gamma_\infty\} a_f(h)) \\ &= \pi(\pi^{-1}(x) a_f(h)). \end{aligned} \quad (4.3)$$

Proof. Recall that the elementary divisors theorem attaches to any pair of lattices $\Lambda_1 < \Lambda_2$ in the plane, a pair of integers d_1, d_2 which are characterized by the following two properties: (1) the divisibility $d_2 | d_1$ holds, (2) there exists a basis v_1, v_2 of Λ_2 such that $d_1 v_1, d_2 v_2$ forms a basis of Λ_1 . Note that Λ_1 is a primitive sublattice of Λ_2 if and only if the second divisor satisfies $d_2 = 1$. We conclude from here that given a class $x = \Gamma_\infty g$, then a class x' lies on the sphere $\mathcal{S}_h(x)$ if and only if there exists a lattice $\Lambda_{x'} \in x'$ which is a sublattice of the lattice $\Lambda_x \in x$ spanned by the rows of g such that the elementary divisors are $d_1 = h, d_2 = 1$. In other words we have the equality

$$\mathcal{S}_h(x) = \{\Gamma_\infty \text{diag}(h, 1) \gamma g : \gamma \in \Gamma_\infty\}. \quad (4.4)$$

The following identity is crucial for us. It shows how the lattice Γ_S causes the desired interaction between the real and p -adic components in the extension X_S of X_∞ :

$$\begin{aligned} \Gamma_\infty \text{diag}(h, 1) \gamma g &= \pi(\Gamma_S(\text{diag}(h, 1) \gamma g, e_f)) \\ &= \pi\left(\Gamma_S \underbrace{\gamma^{-1} \text{diag}(1, h)}_{\in \Gamma_S} (\text{diag}(h, 1) \gamma g, e_f)\right) = \pi(\Gamma_S(g, \gamma^{-1}) a_f(h)). \end{aligned} \quad (4.5)$$

From equations (4.4), (4.5) we immediately conclude that

$$\mathcal{S}_h = \pi(\{\Gamma_S(g, \gamma) : \gamma \in \Gamma_\infty\} a_f(h)),$$

which is the first equality in (4.3). Using the first equality, the second equality follows once we show that for any given $\omega \in K_S$ there exist $\gamma \in \Gamma_\infty$ such that $\pi(\Gamma_S(g, \gamma) a_f(h)) = \pi(\Gamma_S(g, \omega) a_f(h))$. A short calculation using Remark 3.1 shows that this happens precisely when

$$\gamma^{-1} \omega \in a_f(h) K_S a_f(h)^{-1}. \quad (4.6)$$

Thus, let $\omega = (\omega_p)_{p \in S_f} \in K_S$ be given and write $\omega_p = \theta_p \cdot \text{diag}(1, \det(\omega_p))$, with $\theta_p \in \text{SL}_2(\mathbb{Z}_p)$. Let $U_n^p < \text{SL}_2(\mathbb{Z}_p)$ be the subgroup consisting of elements congruent to the identity modulo p^n . By the strong approximation Theorem for SL_2 (see [PR94, §7.4]), for any $n \in \mathbb{N}$ there exist $\gamma_n \in \Gamma_\infty$ such that for all $p \in S_f$

$$\gamma_n^{-1} \theta_p \in U_n^p.$$

Note that there exist $N = N(\mathfrak{h}) \in \mathbb{N}$ such that for all $n > N$ we have that the image of $\prod_{p \in S} U_n^p$ in G_S lies in $a_f(\mathfrak{h})K_S a_f(\mathfrak{h})^{-1}$. As $a_f(\mathfrak{h})K_S a_f(\mathfrak{h})^{-1}$ is a group that contains $(\text{diag}(1, \det(\omega_p)))_{p \in S_f}$ we conclude that $\prod_{p \in S_f} U_n^p \cdot \text{diag}(1, \det(\omega_p)) \subset a_f(\mathfrak{h})K_S a_f(\mathfrak{h})^{-1}$. Therefore any γ_n with $n > N$ will satisfy equation (4.6). This concludes the proof of the Lemma. \square

Corollary 4.2. *For $x = \Gamma_\infty g \in X_\infty$ and \mathfrak{h} an admissible radius another description of the sphere is given by*

$$\mathcal{S}_\mathfrak{h} = \{\Gamma_\infty \gamma g : \gamma \in \Gamma_S, \text{ht}(\gamma) = \mathfrak{h}\}.$$

Proof. Similarly to the proof of Lemma 4.1 one can show that any element $\gamma \in \Gamma_S$ can be written as a product $\gamma = \gamma_1 \text{diag}(\mathfrak{h}, 1) \gamma_2$, where $\gamma_i \in \Gamma_\infty$ and $\mathfrak{h} \in \mathcal{O}_S^\times \cap \mathbb{N}$. This implies first that $\text{ht}(\gamma) = \mathfrak{h}$ and moreover, together with (4.4) we obtain that $\mathcal{S}_\mathfrak{h} = \{\Gamma_\infty \gamma g : \gamma \in \Gamma_S, \text{ht}(\gamma) = \mathfrak{h}\}$ as desired. \square

Definition 4.3. Let $x \in X_\infty$ be given. Let $g_x \in G_\infty$ be a choice of a representative for x so that $x = \Gamma_\infty g_x$. For any choice $\omega \in K_S$ we define the *generalized ray* $\mathcal{L}_{g_x, \omega} \subset \mathcal{G}_S(x)$ to be the set

$$\mathcal{L}_{g_x, \omega} = \pi(\{\Gamma_S(g_x, \omega) a_f(\mathfrak{h}) : \mathfrak{h} \text{ is an admissible radius}\}). \quad (4.7)$$

When ω is a rational element (i.e. for any $p \in S$ the p 'th component ω_p of ω satisfies $\omega_p \in K_p \cap \text{PGL}_2(\mathbb{Q})$) we call the generalized ray $\mathcal{L}_{g_x, \omega}$ a *rational generalized ray*.

The reader should think of the generalized rays as prescribed ways to go to infinity in the graph $\mathcal{G}_S(x)$. When S is composed of a single prime the generalized rays are exactly the rays on the Hecke tree that start from the root x .

Remark 4.4. We wish to point out a few things regarding the definition of generalized rays and fix some notation that will be used in the sequel. Let $x = \Gamma_\infty g_x \in X_\infty$ be given.

- (1) For any $\omega \in K_S$ and any admissible radius \mathfrak{h} we denote $y_{\omega, \mathfrak{h}} = \Gamma_S(g_x, \omega) a_f(\mathfrak{h}) \in X_S$, $x_{\omega, \mathfrak{h}} = \pi(y_{\omega, \mathfrak{h}}) \in X_\infty$. With this notation the generalized ray $\mathcal{L}_{g_x, \omega}$ intersects the sphere $\mathcal{S}_\mathfrak{h}(x)$ in a single point, namely

$$\{x_{\omega, \mathfrak{h}}\} = \mathcal{L}_{g_x, \omega} \cap \mathcal{S}_\mathfrak{h}(x). \quad (4.8)$$

When the generalized ray is fixed (that is when ω is fixed) we sometimes denote $x_\mathfrak{h} = x_{\omega, \mathfrak{h}}$. We stress here the dependency on the representative g_x of x . Note that we do not recall this dependency in the notation $x_{\omega, \mathfrak{h}}, y_{\omega, \mathfrak{h}}$.

- (2) Two generalized rays $\mathcal{L}_{g_x, \omega_1}, \mathcal{L}_{g_x, \omega_2}$ intersect the sphere $\mathcal{S}_\mathfrak{h}(x)$ at the same point, that is, $x_{\omega_1, \mathfrak{h}} = x_{\omega_2, \mathfrak{h}}$, if and only if the points $y_{\omega_i, \mathfrak{h}}$ lie in the same fiber of π . This is in turn equivalent to saying that the conjugation $(\omega_2^{-1} \omega_1)^{a_f(\mathfrak{h})}$ lies in K_S (see Remark 3.1). This happens if and only if the lower left coordinate of each of the components of $\omega_2^{-1} \omega_1$ is divisible by \mathfrak{h} in the corresponding ring \mathbb{Z}_p . In particular, it follows that it is divisible by any integer that divides \mathfrak{h} which means by the same reasoning, that the two rays intersect all the spheres $\mathcal{S}_{\mathfrak{h}'}$ at the same points, for any choice of admissible radius \mathfrak{h}' dividing \mathfrak{h} . Moreover, it follows from here

that given $\omega_1, \omega_2 \in K_S$, the two generalized rays $\mathcal{L}_{g_x, \omega_i}$ are identical if and only if the quotient $\omega_2^{-1}\omega_1$ is an upper triangular element of K_S .

- (3) From the above it follows that the collection of generalized rays may be identified with the quotient K_S/B , where $B < K_S$ denotes the group of upper triangular elements (this identification depends of course on the choice of the representative g_x).
- (4) If we replace g_x by another representative γg_x for $\gamma \in \Gamma_\infty$, then it readily follows that for any $\omega \in K_S$, $\mathcal{L}_{\gamma g_x, \omega} = \mathcal{L}_{g_x, \gamma^{-1}\omega}$. In particular, the notion of rationality of a generalized ray is well defined.

4.4. Periodic A_∞ -orbits.

Definition 4.5. Let $x \in X_\infty$ be a class with a periodic A_∞ -orbit and let $g_x \in G_\infty$ be a representative, so that $x = \Gamma_\infty g_x$. We denote by

- (1) t_x the *length of the period*, i.e. the minimal positive t for which $xa_\infty(t) = x$;
- (2) μ_x the unique A_∞ -invariant probability measure supported on xA_∞ ;
- (3) γ_x the unique element of Γ_∞ solving the equation $\gamma_x^{-1}g_x = g_x a(t_x)$;
- (4) \mathbb{F}_x the quadratic extension of \mathbb{Q} that is generated by the eigenvalues of γ_x .

Note that γ_x depends on the choice of the representative g_x and thus is only well defined up to conjugation in Γ_∞ . The quadratic field \mathbb{F}_x on the other hand, only depends on this conjugacy class and so is well defined.

Let $x \in X_\infty$ be a class with a periodic A_∞ -orbit. It is straightforward to argue that any $x' \in \mathcal{G}_S(x)$ has a periodic orbit as well. We are interested in understanding the way the orbit $x'A_\infty$ is distributed in X_∞ as $d_G(x, x')$ goes to ∞ .

Remark 4.6. It turns out that the answer to this question has to do with the question of whether or not the primes $p \in S$ split in the quadratic extension \mathbb{F}_x of \mathbb{Q} . Let $\gamma \in \Gamma_\infty$ be a matrix such that the roots of its characteristic polynomial generate \mathbb{F}_x . Recall that a prime p splits in \mathbb{F}_x if and only if γ is diagonalizable over \mathbb{Q}_p . A short exercise in linear algebra shows that $\gamma \in \Gamma_\infty$ is diagonalizable over \mathbb{Q}_p if and only if it can be triangulized over \mathbb{Z}_p .

Definition 4.7. Let x be a class with a periodic A_∞ -orbit, g_x a representative so that $x = \Gamma_\infty g_x$, and $\omega \in K_S$.

- (1) We say that the generalized ray $\mathcal{L}_{g_x, \omega}$ is *degenerate* (for S) if there exists $p \in S_f$ such that $\omega_p^{-1}\gamma_x^n\omega_p$ is upper triangular for some positive integer n (here ω_p is the p 'th component of $\omega \in K_S$).⁶
- (2) We say that the class x is *split* (for S) if there exists $p \in S$ which splits over \mathbb{F}_x . By Remark 4.6, this is equivalent to the existence of a degenerate generalized ray.

We are now ready to state our main theorem.

Theorem 4.8. *Let $x = \Gamma_\infty g_x \in X_\infty$ be such that xA_∞ is periodic.*

⁶This is equivalent to saying that the Lie algebra of the closure of the group generated by γ_x^ω in G_p is upper triangular.

- (1) Let $\mathcal{L} = \mathcal{L}_{g_x, \omega}$ be a non-degenerate generalized ray of the graph $\mathcal{G}_S(x)$ and h an admissible radius, then for any $\varphi_0 \in \text{Lip}_\kappa(X_\infty) \cap L^2(X_\infty, m_\infty)$ and any $\epsilon > 0$ the following holds

$$\left| \int_{X_\infty} \varphi_0 d\mu_{x_h} - \int_{X_\infty} \varphi_0 dm_\infty \right| \ll_{x, S, \mathcal{L}, \epsilon} \max \{ \|\varphi_0\|_2, \kappa \} h^{-\frac{\epsilon_0}{2} + \epsilon}. \quad (4.9)$$

- (2) If x is non-split (i.e. all generalized rays are non-degenerate), the implicit constant in (4.9) may be chosen to be independent of the generalized ray and we have uniform rate of equidistribution along the full graph.
- (3) If x is split and $\mathcal{L}_{g_x, \omega}$ is a degenerate generalized ray, then there is a sequence of admissible radii $h_n \rightarrow \infty$ such that for the sequence of classes x_{h_n} , the lengths $t_{x_{h_n}}$ of the orbits $x_{h_n} A_\infty$ are bounded and in particular, the orbits do not equidistribute.
- (4) Rational generalized rays are always non-degenerate and so (4.9) holds automatically.
- (5) Nonetheless, in case x is split, the implicit constants in (4.9) cannot be taken to be uniform for the rational generalized rays.

5. RELATIONS TO OTHER ARGUMENTS

Before turning to the proof of Theorem 4.8 we wish to make some comments that will clarify its relation to arguments giving equidistribution of collections of periodic orbits. The result of Benoist and Oh [BO07, Theorem 1.1] imply that given a class x with a periodic A_∞ -orbit, then the collection of orbits $\{x' A_\infty : x' \in \mathcal{S}_h(x)\}$ (counted without multiplicities) is becoming equidistributed as $h \rightarrow \infty$.

Ignoring the effectivity of Theorem 4.8 and just interpreting it as saying that $\mu_{x'} \rightarrow m_\infty$ as x' drifts away from the root x along a non-degenerate generalized ray, it seems tempting to think that it is considerably stronger than the result of Benoist and Oh, as it deals with the equidistribution of single orbits as opposed to the equidistribution of the full collection. We will show in §5.1 below that this (non-effective) equidistribution in fact follows quite elementarily from the work of Benoist and Oh. Nonetheless, the argument we give for Theorem 4.8 is independent of [BO07] and as far as we know the effective statements in Theorem 4.8 do not follow easily from known results.

5.1. Total vs. individual growth. Let $x \in X_\infty$ be a class with a periodic A_∞ -orbit and consider the union of the periodic orbits $x' A_\infty$ for $x' \in \mathcal{S}_h(x)$ (where h is an admissible radius). We denote the total length of this union by $\mathbf{t}_x(h)$; that is, $\mathbf{t}_x(h) \stackrel{\text{def}}{=} \sum t_{x'}$ where the sum is taken over a set of representatives of the classes on the sphere giving rise to different orbits. The following Lemma shows that the growth rate of the length of individual periodic orbits along a non-degenerate generalized ray is the same as the growth rate of the total length. Although we only use this Lemma in the course of the proofs regarding the growth rate of the periods (cf. Theorems 2.12 and 8.10) it explains a phenomenon that to some extent stands behind all of our results. Its proof is given in §7.

Lemma 5.1. *Let $x \in X_\infty$ be a class with a periodic A_∞ -orbit. For any generalized ray \mathcal{L} of $\mathcal{G}_S(x)$, let $c_{\mathcal{L}}(\mathfrak{h}) \stackrel{\text{def}}{=} t_{x_{\mathfrak{h}}}/\mathfrak{h}$, where $x_{\mathfrak{h}}$ is the class in $\mathcal{S}_{\mathfrak{h}} \cap \mathcal{L}$.*

- (1) *The total length $t_x(\mathfrak{h})$ satisfies $\mathfrak{h} \ll_{x,S} t_x(\mathfrak{h}) \ll_{x,S} \mathfrak{h}$.*
- (2) *If \mathcal{L} is a non-degenerate generalized ray then $c_{\mathcal{L}}(\mathfrak{h})$ attains only finitely many values and moreover, if \mathfrak{h}_n is a divisibility sequence of admissible radii (that is $\mathfrak{h}_n \mid \mathfrak{h}_{n+1}$), then $c_{\mathcal{L}}(\mathfrak{h}_n)$ stabilizes.*
- (3) *The class x is non-split for S if and only if*

$$\inf \{c_{\mathcal{L}}(\mathfrak{h}) : \mathcal{L} \text{ non-degenerate, } \mathfrak{h} \in \mathcal{O}_S^\times \cap \mathbb{N}\} > 0. \quad (5.1)$$

The first two parts of Lemma 5.1 show that if \mathcal{L} is a non-degenerate generalized ray, then a single orbit $x_{\mathfrak{h}}A_\infty$ through the class $x_{\mathfrak{h}} \in \mathcal{L} \cap \mathcal{S}_{\mathfrak{h}}(x)$ actually occupies a positive proportion (bounded below by a constant independent of \mathfrak{h}) of the full collection $\{x'A_\infty : x' \in \mathcal{S}_{\mathfrak{h}}(x)\}$. Relying on [BO07] we may argue the non-effective version of Theorem 4.8 (that is, that $\mu_{x_{\mathfrak{h}}} \rightarrow m_\infty$ as $\mathfrak{h} \rightarrow \infty$) in the following way: Let $\mathfrak{h}_i \rightarrow \infty$ be a sequence of admissible radii such that $\mu_{x_{\mathfrak{h}_i}}$ converges to say μ_∞ (which is an A_∞ -invariant measure). We need to argue that $\mu_\infty = m_\infty$. Let $\eta_{\mathfrak{h}}$ be the natural A_∞ -invariant probability measure supported on the collection of periodic orbits $\{x'A_\infty : x' \in \mathcal{S}_{\mathfrak{h}}(x)\}$. By [BO07] $\eta_{\mathfrak{h}} \rightarrow m_\infty$. By the first two parts of Lemma 5.1 we can write $\eta_{\mathfrak{h}_i}$ as a convex combination of A_∞ -invariant probability measures in the following way: $\eta_{\mathfrak{h}_i} = c'_{\mathfrak{h}_i} \mu_{x_{\mathfrak{h}_i}} + (1 - c'_{\mathfrak{h}_i}) \nu_{\mathfrak{h}_i}$, where the constants $c'_{\mathfrak{h}_i}$ are bounded below by some constant c' independent of \mathfrak{h}_i . Taking i to ∞ (along an appropriate subsequences if necessary) we deduce that in the limit $m_\infty = c'_\infty \mu_\infty + (1 - c'_\infty) \nu_\infty$ for some positive constant $c'_\infty \leq 1$. By the ergodicity of m_∞ with respect to the A_∞ -action we deduce that the limit μ_∞ that appears in the above convex combination with positive weight, must be equal to m_∞ . This establishes the desired convergence.

6. PROOF OF THEOREM 4.8.

Throughout this section we fix $x \in X_\infty$ to be a class with a periodic A_∞ -orbit and a representative $g_x \in G_\infty$ such that $x = \Gamma_\infty g_x$. Using the notation of Definition 4.5, it follows that there exists $\gamma_x \in \Gamma_\infty$ such that

$$\gamma_x g_x a_\infty(t_x) = g_x. \quad (6.1)$$

We briefly discuss the relations between the various parts of Theorem 4.8. As the eigenvalues of γ_x are irrational (and not roots of unity) it follows that γ_x (or any of its powers) is not triangulizable over \mathbb{Q} and so all the rational generalized rays are non-degenerate. This establishes part (4) of the theorem. Part (5) of the theorem follows from part (3) because of (4.3) which shows that any class on the S -Hecke graph $\mathcal{G}_S(x)$ lies on a rational generalized ray; the sequence $x_{\mathfrak{h}_n}$ produced by part (3) may be viewed as a sequence of classes lying on (varying) rational generalized rays, showing that a uniform implicit constant for all rational generalized rays in (4.9) is impossible.

We begin with the necessary preparations for the arguments yielding parts (1),(2), and (3). We will see below that part (3) is a simple observation once the stage is set correctly and so the main bulk of the theorem lies in establishing parts (1) and (2).

After fixing g_x we fix a generalized ray in $\mathcal{G}_S(x)$; that is, we fix an element $\omega \in K_S$ and set $\mathcal{L}_\omega = \mathcal{L}_{g_x, \omega}$. Although ω is fixed, the reader should bear in mind that at some point we will vary the choice of ω in order to change the generalized ray.

6.1. The lift of a closed loop. The following construction is fundamental to our argument. Let $y_\omega \in X_S$ be defined by $y_\omega = \Gamma_S(g_x, \omega)$. Consider the orbit $y_\omega A_\infty \subset X_S$ and note that

$$\pi(y_\omega) = x, \quad xA_\infty = \pi(y_\omega A_\infty) = \pi(\overline{y_\omega A_\infty}), \quad (6.2)$$

where the rightmost equality follows from the fact that xA_∞ is compact and the continuity of the projection π .

We now analyze the closure $\overline{y_\omega A_\infty}$. Each $t \in \mathbb{R}$ can be written in a unique way in the form $t = s + \ell t_x$ for some $s \in [0, t_x)$ and $\ell \in \mathbb{Z}$. It follows from (6.1) that

$$y_\omega a_\infty(t) = \Gamma_S(g_x a_\infty^\ell(t_x) a_\infty(s), \omega) = \Gamma_S(g_x a_\infty(s), \gamma_x^\ell \omega) = y_\omega(a_\infty(s), \omega^{-1} \gamma_x^\ell \omega). \quad (6.3)$$

If we denote for an element γ in a group H by $\langle \gamma \rangle_H$ the cyclic group generated by γ in H , then it follows from (6.3) that

$$y_\omega A_\infty = y_\omega(A_\infty \times \langle \omega^{-1} \gamma_x \omega \rangle_{G_S}). \quad (6.4)$$

Let

$$H_\omega = \omega^{-1} \overline{\langle \gamma_x \rangle_{G_S}} \omega = \overline{\langle \omega^{-1} \gamma_x \omega \rangle_{G_S}}. \quad (6.5)$$

Clearly, H_ω is a compact subgroup of K_S . We let

$$L_\omega = A_\infty \times H_\omega. \quad (6.6)$$

Lemma 6.1. *The orbit $y_\omega L_\omega$ is compact and*

$$\overline{y_\omega A_\infty} = y_\omega L_\omega. \quad (6.7)$$

Proof. We first establish (6.7). The inclusion \supset follows readily from (6.4). For the reverse inclusion, let $t_n \in \mathbb{R}$ be such that $y_\omega a_\infty(t_n) \rightarrow_{n \rightarrow \infty} y \in \overline{y_\omega A_\infty}$. Let $s_n \in [0, t_x)$, $\ell_n \in \mathbb{Z}$ be as defined before (6.3); that is $t_n = s_n + \ell_n t_x$. By compactness we may assume without loss of generality (after passing to a subsequence if necessary) that $s_n \rightarrow s$ and $\omega^{-1} \gamma_x^{\ell_n} \omega \rightarrow h$. We conclude from (6.3) that

$$y = \lim y_\omega a_\infty(t_n) = \lim y_\omega(a_\infty(s_n), \omega^{-1} \gamma_x^{\ell_n} \omega) = y_\omega(a_\infty(s), h) \in y_\omega L_\omega. \quad (6.8)$$

The fact that the orbit $y_\omega L_\omega$ is compact now follows from the fact that it is a closed set contained in $\pi^{-1}(xA_\infty)$ which is compact by the properness of π . \square

Remark 6.2. The above proof actually establishes a bit more: We have shown that in fact,

$$\overline{y_\omega A_\infty} = y_\omega L_\omega = \{y_\omega(a_\infty(t), h) : t \in [0, t_x), h \in H_\omega\}. \quad (6.9)$$

Definition 6.3. Let η_ω denote the L_ω -invariant probability measure supported on the compact (and hence periodic) orbit $y_\omega L_\omega$. For an admissible radius h let

$$y_{\omega,h} = y_\omega a_f(h), \quad L_{\omega,h} = L_\omega^{a_f(h)}, \quad H_{\omega,h} = H_\omega^{a_f(h)},$$

and note the identity $y_\omega L_\omega a_f(h) = y_{\omega,h} L_{\omega,h} = y_{\omega,h}(A_\infty \times H_{\omega,h})$. We denote the unique $L_{\omega,h}$ -invariant probability measure supported on the periodic orbit $y_{\omega,h} L_{\omega,h}$ by $\eta_{\omega,h}$. It follows that $(a_f(h))_* \eta_\omega = \eta_{\omega,h}$. Note that the notation $y_{\omega,h}$ is consistent with the one introduced in Remark 4.4(1).

Lemma 6.4. *Let h be an admissible radius and $x_{\omega,h} \in \mathcal{L}_\omega \cap \mathcal{S}_h(x)$. Then, the pushed orbit $y_\omega L_\omega a_f(h) = y_{\omega,h} L_{\omega,h}$ projects to the periodic orbit $x_{\omega,h} A_\infty$ and furthermore, the measure $\eta_{\omega,h}$ supported on it projects to $\mu_{x_{\omega,h}}$; i.e. $\pi_* \eta_{\omega,h} = \mu_{x_{\omega,h}}$.*

Lemma 6.4 puts us in a desirable situation from the dynamical point of view; instead of studying the orbits $x' A_\infty$ in the space X_∞ as x' drifts away from x on a generalized ray (the connection between which is not clear a priori), we will study the images of the fixed orbit $y_\omega L_\omega$ under the action of $a_f(h)$ for admissible radii h , which share a clear algebraic (and geometric) relation. This relation is the reason we needed to introduce the S -arithmetic extension X_S .

Proof. The fact that $x_{\omega,h} = \pi(y_{\omega,h})$ follows from Definition 6.3 and Remark 4.4(1). We have that

$$\begin{aligned} \pi(y_{\omega,h} L_{\omega,h}) &= \pi(y_\omega L_\omega a_f(h)) = \pi(\overline{y_\omega A_\infty a_f(h)}) \\ &= \overline{\pi(y_\omega a_f(h) A_\infty)} = \overline{\pi(y_{\omega,h}) A_\infty} = x_{\omega,h} A_\infty, \end{aligned} \tag{6.10}$$

where the first equality from the left follows from Definition 6.3, the second, from Lemma 6.1 and that fact that $a_f(h)$ acts on X_S by a homeomorphism, the third, from the commutation of A_∞ and $a_f(h)$ and from the continuity of π , the fourth, from the fact that π intertwines the A_∞ -actions on X_S, X_∞ , and finally the fifth equality follows from the fact that the orbit $x_{\omega,h} A_\infty$ is compact.

As $A_\infty < L_{\omega,h}$, $\eta_{\omega,h}$ is A_∞ -invariant. As a consequence, the projection $\pi_* \eta_{\omega,h}$ is an A_∞ -invariant probability measure supported on $x_{\omega,h} A_\infty$. As $\mu_{x_{\omega,h}}$ is the unique such measure, we conclude that $\pi_* \eta_{\omega,h} = \mu_{x_{\omega,h}}$ as desired. \square

Remark 6.5. It follows from (6.9) and the definition of $y_{\omega,h}, H_{\omega,h}$ that

$$y_{\omega,h} L_{\omega,h} = \{y_{\omega,h}(a_\infty(t), h) : t \in [0, t_x], h \in H_{\omega,h}\}.$$

By (6.10) the following equality follows:

$$x_{\omega,h} A_\infty = \pi(\{y_{\omega,h}(a_\infty(t), h) : t \in [0, t_x], h \in H_{\omega,h}\}). \tag{6.11}$$

The meaning of the above equation is that the only reason for the orbit $x_{\omega,h} A_\infty$ to become long is that the group $H_{\omega,h}$ stretches and ‘sticks out’ of K_S . This is illustrated in the following proof.

Proof of part (3) of Theorem 4.8. For an admissible radius h and $p \in S$ denote by $(H_{\omega,h})_p$ the projection of the group $H_{\omega,h}$ on its p -th component. Note that by definition, $(H_{\omega,h})_p = \text{diag}(1, h^{-1})(H_{\omega})_p \text{diag}(1, h)$.

Assume that the generalized ray \mathcal{L}_{ω} is degenerate. It follows that there exists $p \in S$ for which some power of the p -th component $(\omega^{-1}\gamma_x\omega)_p$ is upper triangular. Let d be the minimal positive integer for which $(\omega^{-1}\gamma_x^d\omega)_p$ is upper triangular. We conclude from (6.5) that $(H_{\omega})_p$ contains an index d subgroup that consists of upper triangular elements only. Choose $h_n = p^n$ and note that because of the above $(H_{\omega,h_n})_p \cap K_p$ is of index at most d in $(H_{\omega,h_n})_p$. Moreover, note that as p is a unit in $\mathbb{Z}_{p'}$ for any prime $p' \neq p$, we have that $(H_{\omega,h_n})_{p'} < K_{p'}$. It follows that along the chosen sequence h_n we have that $H_{\omega,h_n} \cap K_S$ has at most index d in H_{ω,h_n} . Let $h_i \in H_{\omega,h_n}, i = 1 \dots d', d' \leq d$, be representatives of the cosets of $H_{\omega,h_n} \cap K_S$ and denote $y_i = y_{\omega,h_n} h_i, i = 1 \dots d'$ and $x_i = \pi(y_i)$. We can rewrite (6.11) as

$$\begin{aligned} x_{\omega,h_n} A_{\infty} &= \pi \left(\cup_{i=1}^{d'} \{y_{\omega,h_n}(a_{\infty}(t), h_i h) : t \in [0, t_x], h \in H_{\omega,h_n} \cap K_S\} \right) \\ &= \pi \left(\cup_{i=1}^{d'} \{y_i(a_{\infty}(t), h) : t \in [0, t_x], h \in H_{\omega,h_n} \cap K_S\} \right) \\ &= \cup_{i=1}^{d'} \{x_i a_{\infty}(t) : t \in [0, t_x]\}, \end{aligned} \quad (6.12)$$

and so we conclude that $t_{x_{\omega,h_n}} \leq d' t_x$ which finishes the proof. \square

In order to finish the proof of Theorem 4.8 we are left to argue parts (1),(2). As said before, these are the main parts of the theorem.

6.2. Strategy of the proof of Theorem 4.8(1),(2). In the notation of Lemma 6.4, because $\pi_* \eta_{\omega,h} = \mu_{x_{\omega,h}}$, the validity of (4.9) is equivalent to saying that given $\varphi \in \text{Lip}_{\kappa}(X_S) \cap L^2(X_S, m_S)$ which is K_S -invariant (i.e. is of the form $\varphi_0 \circ \pi$ for $\varphi_0 \in \text{Lip}_{\kappa}(X_{\infty}) \cap L^2(X_{\infty}, m_{\infty})$)

$$\left| \int \varphi d\eta_{\omega,h} - \int \varphi dm_S \right| \ll_{x,S_f,\mathcal{L}_{\omega},\epsilon} \max\{\kappa, \|\varphi\|_2\} h^{-\frac{\delta_0}{2} + \epsilon}. \quad (6.13)$$

The argument giving this ‘effective equidistribution’ is a combination of an argument which we will refer to as *the mixing trick* and spectral gap (or effective decay of matrix coefficients). As far as we know the mixing trick originates from Margulis’ thesis [Mar04]. We briefly describe its heuristics: One slightly thickens the initial orbit $y_{\omega} L_{\omega}$ to an open set $\mathcal{T} \subset X_S$ in directions which are (weakly) contracted by the action of $a_f(h)$. The set \mathcal{T} will be called below *a tube around the orbit* $y_{\omega} L_{\omega}$. Let $m_{\mathcal{T}}$ denote the normalized restriction of m_S to \mathcal{T} . The pushed measure $(a_f(h))_* m_{\mathcal{T}}$ is the normalized restriction of m_S to the pushed tube $\mathcal{T} a_f(h)$, which is a tube around the orbit $y_{\omega,h} L_{\omega,h}$. Because the thickening used to construct \mathcal{T} is taken in directions which are (weakly) contracted by $a_f(h)$, the size of the thickening giving the tube $\mathcal{T} a_f(h)$ is even smaller than the size of the initial thickening. Hence, there shouldn’t be much of a difference between integrating against the measure $\eta_{\omega,h}$ and integrating against $(a_f(h))_* m_{\mathcal{T}}$. The fact that the action of $a_f(h)$ is mixing on (X_S, m_S) means that the pushed measure $(a_f(h))_* m_{\mathcal{T}}$ is ‘close’ to m_S

(here, the effective mixing Theorem 6.6 will allow us to pin down the meaning of ‘close’ in a precise way). Combining these things together will give us the desired estimate given in (6.13).

In order to make this strategy into a rigorous proof we discuss in the next two subsections in detail the construction of tubes and decay of matrix coefficients.

6.3. Effective mixing. Let $\mathcal{H} = L^2(X_S, m_S)$. Our goal in this section is to prove the following:

Theorem 6.6. *Let h be an admissible radius and $w_1, w_2 \in \mathcal{H}$ be vectors with the following properties: w_1 is K_S -fixed and w_2 is stabilized by a product subgroup $K^* = \prod_{v \in S} K_v^* < K_S$ of index d in K_S . Then for any $\epsilon > 0$,*

$$|\langle w_1, a_f(h)w_2 \rangle - \langle w_1, 1 \rangle \langle 1, w_2 \rangle| \ll_\epsilon \|w_1\| \|w_2\| d^{\frac{1}{2}} h^{-\delta_0 + \epsilon}, \quad (6.14)$$

The meaning of the exponent δ_0 that appears in (6.14) will be explicated shortly. Before turning to the proof of the above theorem, we need to discuss three lemmas. For $v \in S$ let \mathcal{H}_v denote the orthocomplement of the G_v -invariant functions in \mathcal{H} . The following is [Ven10, Lemma 9.1]. It is the key input in the proof of Theorem 6.6.

Lemma 6.7. *Let $w_1, w_2 \in \mathcal{H}_v$ ($v \in S$) be two vectors which are stabilized respectively by finite index subgroups $K^{(1)}, K^{(2)}$ of K_v , let $d_i = [K_v, K^{(i)}]$, and $a_v(t) = \text{diag}(1, t)$, $t \in \mathbb{Q}_v^\times$. Then the following holds*

$$|\langle w_1, a_v(t)w_2 \rangle| \ll_\epsilon \|w_1\| \|w_2\| d_1^{\frac{1}{2}} d_2^{\frac{1}{2}} \max\{|t|_v, |t^{-1}|_v\}^{-\delta_0 + \epsilon}. \quad (6.15)$$

The exponent δ_0 comes from the following discussion. Let ρ_v be the unitary representation of G_v on \mathcal{H}_v . Let σ_0 be the smallest number so that no complementary series representations of parameter $\geq \sigma_0$ is weakly contained in ρ_v . Here we follow [Ven10] and parametrize the complementary series representations by the parameter $\sigma \in (0, \frac{1}{2})$; so $\sigma_0 = 0$ corresponds to ρ_v being tempered (the Ramanujan conjecture) and $\sigma_0 = \frac{1}{2}$ corresponds to ρ_v having no almost invariant vectors. The best bound known today towards Ramanujan is given by Kim and Sarnak in the appendix of [Kim03] and establishes the bound $\sigma_0 \leq \frac{7}{64}$. The exponent δ_0 that appears in Lemma 6.7 and in the other results is defined by

$$\delta_0 = \frac{1}{2} - \sigma_0, \quad (6.16)$$

so the Kim-Sarnak bound reads as $\delta_0 \geq \frac{25}{64}$.

Lemma 6.7 is stated for one place $v \in S$ but in Theorem 6.6 we wish to take advantage of the various places h is supported on. In order to do this, we will need to use Lemma 6.7 iteratively and the following abstract lemma in Hilbert space theory allows us to do so.

Lemma 6.8. *Let $G = G_1 \times G_2$ be a group acting unitarily on a Hilbert space \mathcal{H} . Let $K_i < G_i$ be subgroups, $g_i \in G_i$ be two given elements, and $F(g_i)$ two positive numbers satisfying the following statement: For each i , if $v, w \in \mathcal{H}$ are K_i -fixed vectors, then*

$$\langle g_i v, w \rangle \leq \|v\| \|w\| F(g_i).$$

Then for any $v, w \in \mathcal{H}$ which are $K_1 \times K_2$ -fixed we have that

$$\langle g_1 g_2 v, w \rangle \leq \|v\| \|w\| F(g_1) F(g_2).$$

Proof. Let us denote for $i = 1, 2$ $V_i = \{v \in \mathcal{H} : v \text{ is } K_i\text{-fixed}\}$ and $U = V_1 \cap V_2$. Let V'_i denote the orthocomplement of U in V_i and denote for a subspace W of \mathcal{H} by P_W the orthogonal projection on W . We first note that V_1, V_2 are K_2, K_1 -invariant respectively (because K_1, K_2 commute) and so the projections P_{V_1}, P_{V_2} commute with the actions of K_2, K_1 respectively. It follows from here that given $v_1 \in V_1$, say, the projection $P_{V_2}(v_1)$ is fixed by both K_1 and K_2 i.e. $P_{V_2}(v_1) \in U$. This proves that V'_1 is orthogonal to V_2 or in a more symmetric manner, V'_1 is orthogonal to V'_2 .

Let now v, w be two $K_1 \times K_2$ -fixed vectors. As $g_1 v$ is K_2 -fixed, i.e. $g_1 v \in V_2$, we may write $g_1 v = P_U(g_1 v) + P_{V'_2}(g_1 v)$ and similarly $g_2 w = P_U(g_2 w) + P_{V'_1}(g_2 w)$. It follows that

$$\begin{aligned} \langle g_1 v, g_2 w \rangle &= \langle P_U(g_1 v) + P_{V'_2}(g_1 v), P_U(g_2 w) + P_{V'_1}(g_2 w) \rangle \\ &= \langle P_U(g_1 v), P_U(g_2 w) \rangle \leq \|P_U(g_1 v)\| \|P_U(g_2 w)\|. \end{aligned} \quad (6.17)$$

Let $\tilde{v} = \frac{P_U(g_1 v)}{\|P_U(g_1 v)\|}$. Then \tilde{v} is K_2 -fixed and so by the assumption of the lemma we conclude that

$$\|P_U(g_1 v)\| = \langle g_1 v, \tilde{v} \rangle \leq \|v\| F(g_1).$$

Similarly, $\|P_U(g_2 w)\| \leq \|w\| F(g_2)$. Plugging this into (6.17) yields

$$\langle g_1 v, g_2 w \rangle \leq \|v\| \|w\| F_1(g_1) F_2(g_2),$$

which is equivalent to the desired statement up to replacing g_2 by its inverse (note that the assumption on g_i implies the corresponding assumption on g_i^{-1}). \square

The final ingredient needed for the proof of Theorem 6.6 is the following

Lemma 6.9. *For each place $v \in S$ the group generated by G_v and K_S acts ergodically on X_S , that is, $\{w \in \mathcal{H} : w \text{ is both } G_v, K_S\text{-fixed}\}$ is the one dimensional space of constant functions.*

Proof. Recall that $S^* = S \cup \{\infty\}$. Let $Y_S = \mathrm{SL}_2(\mathcal{O}_S) \backslash \prod_{v \in S^*} \mathrm{SL}(\mathbb{Q}_v)$. The strong approximation property for SL_2 implies that for any $v \in S$ the lattice $\mathrm{SL}_2(\mathcal{O}_S)$ embeds densely in $\prod_{v' \in S^* \setminus \{v\}} \mathrm{SL}_2(\mathbb{Q}_{v'})$. This is equivalent to saying that $\mathrm{SL}_2(\mathbb{Q}_v)$ acts minimally on Y_S (i.e. that any orbit is dense). In turn, this implies that $\mathrm{SL}_2(\mathbb{Q}_v)$ acts ergodically on Y_S (by the duality trick for example). Now, consider the natural map $\psi : \mathrm{SL}_2 \rightarrow \mathrm{PGL}_2$. This map induces a map from Y_S to X_S (which we also denote by ψ) which intertwines the actions of $\mathrm{SL}_2(\mathbb{Q}_v)$ and $\psi(\mathrm{SL}_2(\mathbb{Q}_v)) < G_v$ on these spaces respectively. It follows that the action of $\psi(\mathrm{SL}_2(\mathbb{Q}_v))$ on $\psi(Y_S)$ is ergodic.

Let $w \in \mathcal{H}$ be a function on X_S which is both G_v and K_S -invariant. Its restriction to $\psi(Y_S)$ is constant by the ergodicity proved above. It follows that in order to show that w is constant it is enough to show that the translates of $\psi(Y_S)$ by K_S cover X_S . We briefly sketch the argument: There is a natural ‘determinant map’ $\det : G_{S^*} \rightarrow \prod_{v \in S^*} \mathbb{Q}_v^\times / (\mathbb{Q}_v^\times)^2$. Let us denote $\Delta = \prod_{v \in S^*} \mathbb{Q}_v^\times / (\mathbb{Q}_v^\times)^2$ and $\Delta' = \det(\Gamma_S) < \Delta$. It follows that there is a well defined map $\det : \Gamma_S \backslash G_{S^*} = X_S \rightarrow \Delta' \backslash \Delta$. We leave it to the reader to show that the

space $\psi(Y_S)$ is characterized as the preimage of the identity coset Δ' under $\widetilde{\det}$. Since \det takes K_S onto $\Delta' \setminus \Delta$, we conclude that indeed, translates of $\psi(Y_S)$ under K_S cover X_S as desired. \square

Proof of Theorem 6.6. Let $\mathcal{H} = L^2(X_S, m_S)$ and for $v \in S$ let \mathcal{H}_v be the orthocomplement to the G_v -fixed vectors. Let $\mathcal{H}_0 = \bigcap_{v \in S} \mathcal{H}_v$ and let $w_1, w_2 \in \mathcal{H}$ be as in the statement of the theorem. Write

$$w_i = P_{\mathcal{H}_0}(w_i) + P_{\mathcal{H}_0^\perp}(w_i),$$

and note that the decomposition $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_0^\perp$ is G_S -invariant. It follows that

$$\begin{aligned} \langle a_f(\mathfrak{h})w_1, w_2 \rangle &= \langle a_f(\mathfrak{h}) \left(P_{\mathcal{H}_0}(w_1) + P_{\mathcal{H}_0^\perp}(w_1) \right), P_{\mathcal{H}_0}(w_2) + P_{\mathcal{H}_0^\perp}(w_2) \rangle \\ &= \underbrace{\langle a_f(\mathfrak{h})P_{\mathcal{H}_0}(w_1), P_{\mathcal{H}_0}(w_2) \rangle}_{(*)} + \underbrace{\langle a_f(\mathfrak{h})P_{\mathcal{H}_0^\perp}(w_1), P_{\mathcal{H}_0^\perp}(w_2) \rangle}_{(**)}. \end{aligned} \quad (6.18)$$

Let us first argue that $(**) = \langle w_1, 1 \rangle \langle 1, w_2 \rangle$. The space \mathcal{H}_0^\perp is the space generated by $\{\mathcal{H}_v^\perp\}_{v \in S}$. This implies that the vector $P_{\mathcal{H}_0^\perp}(w_1)$ is in the span of the vectors $P_{\mathcal{H}_v^\perp}(w_1)$ as v runs through S . For each $v \in S$ the vector $P_{\mathcal{H}_v^\perp}(w_1)$ is both G_v and K_S -fixed and so by Lemma 6.9 this implies that $P_{\mathcal{H}_v^\perp}(w_1) \in \mathcal{H}_c$, where \mathcal{H}_c denotes here the 1-dimensional space of constant functions. We conclude that $P_{\mathcal{H}_0^\perp}(w_1) \in \mathcal{H}_c$, or in other words, $P_{\mathcal{H}_0^\perp}(w_1) = P_{\mathcal{H}_c}(w_1) = \langle w_1, 1 \rangle$. Using this we see that

$$(**) = \langle \langle w_1, 1 \rangle, P_{\mathcal{H}_0^\perp}(w_2) \rangle = \langle w_1, 1 \rangle \langle 1, P_{\mathcal{H}_0^\perp}(w_2) \rangle. \quad (6.19)$$

In turn, $\langle P_{\mathcal{H}_0^\perp}(w_2), 1 \rangle$ is the orthogonal projection of $P_{\mathcal{H}_0^\perp}(w_2)$ on \mathcal{H}_c , but as $\mathcal{H}_c \subset \mathcal{H}_0^\perp$, this projection equals $\langle w_2, 1 \rangle$. We conclude from (6.19) that $(**) = \langle w_1, 1 \rangle \langle 1, w_2 \rangle$ as claimed.

We now analyze $(*)$ in (6.18). Because the decomposition $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_0^\perp$ is G_S -invariant the vectors $P_{\mathcal{H}_0}(w_1), P_{\mathcal{H}_0}(w_2)$ are fixed under K_S, K^* respectively (where K^* is as in the statement of the theorem). Order the primes in S in some way $p_1 \dots p_k$ and denote $d_{p_i} = [K_{p_i} : K_{p_i}^*]$, so $d = [K_S : K^*] = \prod_{i=1}^k d_{p_i}$. We leave it to the reader to prove by a simple induction, using Lemmas 6.7, 6.8 that for $j = 1, \dots, k$

$$\left\langle \prod_{i=1}^j a_{p_i}(\mathfrak{h}) P_{\mathcal{H}_0}(w_1), P_{\mathcal{H}_0}(w_2) \right\rangle \ll_\epsilon \|P_{\mathcal{H}_0}(w_1)\| \|P_{\mathcal{H}_0}(w_2)\| \prod_{i=1}^j d_{p_i}^{\frac{1}{2}} \prod_{i=1}^k |\mathfrak{h}|_{p_i}^{-\delta_0 + \epsilon}. \quad (6.20)$$

In particular, for $j = k$ we obtain

$$(*) = \langle a_f(\mathfrak{h})P_{\mathcal{H}_0}(w_1), P_{\mathcal{H}_0}(w_2) \rangle \ll_\epsilon \|w_1\| \|w_2\| d^{\frac{1}{2}} \mathfrak{h}^{-\delta_0 + \epsilon}. \quad (6.21)$$

Equations (6.21), (6.18) and the analysis carried above for $(**)$ now imply the validity of the theorem. \square

6.4. Tubes. As explained in §6.2, we start with a K_S -invariant ‘test function’ φ and we need to thicken the orbit $y_\omega L_\omega$ to a tube \mathcal{T} and then apply (6.14) to the vectors $w_1 = \varphi, w_2 = \chi_{\mathcal{T}}$. In order for the use of (6.14) to be meaningful we need to control d which is the index of the stabilizer of the tube in K_S . Also, the ‘width’ of the tube (i.e. the size of the thickening of the orbit) should be very small (at least in the real component) in order for the heuristics of §6.2 to take effect. This will hopefully motivate the constructions in this subsection.

Definition 6.10. Let $yL \subset X_S$ be a compact orbit of a closed subgroup $L < G_{S^*}$. Let $V = \bigoplus_{v \in S^*} V_v$ be a linear complement to $\text{Lie}(L)$ in \mathfrak{g}_{S^*} . Let $U \subset V$ be a small enough open neighborhood of 0 so that the map $yL \times U \rightarrow X_S$ defined by $(z, u) \mapsto z \exp_{S^*}(u)$ is a homeomorphism onto its image and its image is open in X_S . The set

$$\mathcal{T}_U(yL) = \{z \exp_{S^*}(u) : z \in yL, u \in U\}$$

is called a *tube around the orbit yL of width U* . We often denote the tube simply by \mathcal{T} . The width U and the tube \mathcal{T} are said to come from V .

A tube $\mathcal{T}_U(yL)$ gives us a coordinate system; a point of \mathcal{T} can be written uniquely as $z \exp_{S^*} u$. We refer to z as the *orbit coordinate* and to u as the *width coordinate*. We shall need a few lemmas about tubes which we now turn to describe.

6.4.1. Measures on tubes. Given a tube $\mathcal{T} = \mathcal{T}_U(yL)$ around the compact orbit yL coming from V , one could construct the following two natural probability measures supported on \mathcal{T} . The first is the normalized Haar measure $\frac{1}{m_S(\mathcal{T})} m_S|_{\mathcal{T}}$ which we will denote by $m_{\mathcal{T}}$. The second is the (pushforward of) the product measure $\eta \times m_U$ on $yL \times U \simeq \mathcal{T}$, where η is the unique L -invariant probability measure on the orbit yL and m_U is the normalized restriction of the Haar measure on V to U (that is $m_U = \frac{1}{m_V(U)} m_V|_U$). We shall need to understand to some extent the connection between these two measures.

Lemma 6.11. *The measure $m_{\mathcal{T}}$ is absolutely continuous with respect to $\eta \times m_U$. Moreover, if we denote by $F(z, u)$ the Radon-Nikodym derivative; that is $dm_{\mathcal{T}} = F(z, u) d\eta(z) dm_U(u)$, then for η -almost any $z \in yL$, $\int_U F(z, u) dm_U(u) = 1$.*

Proof. The absolute continuity is left to be verified by the reader. As for the claim about the density F , we argue as follows. Let $\varphi(z) = \int_U F(z, u) dm_U(u)$. We will show that φ is constant η -almost surely. As $\int_{yL} \varphi(z) d\eta(z) = m_{\mathcal{T}}(\mathcal{T}) = 1$ this constant must be equal to one.

Choose a fundamental domain \mathcal{E} in L for the orbit yL and identify it with the orbit. Note that with this identification η is just the restriction to \mathcal{E} of a Haar measure⁷ on L scaled so that $\eta(\mathcal{E}) = 1$. Assume to get a contradiction that φ is not constant η -almost surely. It follows that there are constants $c_2 < c_1$ so that the sets $E_1 = \{h \in \mathcal{E} : \varphi(yh) > c_1\}$, $E_2 = \{h \in \mathcal{E} : \varphi(yh) < c_2\}$ are of positive η -measure. There exists $h_0 \in L$ so that $\eta(E_1 \cap h_0^{-1} E_2) > 0$ and so if we let $\tilde{E}_1 = E_1 \cap h_0^{-1} E_2$ and $\tilde{E}_2 = h_0 \tilde{E}_1$, then $\tilde{E}_i \subset \mathcal{E}$ are both of

⁷Note that L must be unimodular, hence this measure is both left and right invariant.

(the same) positive η -measure and differ from one another by left translation by h_0 . The following calculation derives the desired contradiction:

$$\begin{aligned} c_1\eta(\tilde{E}_1) &\leq \int_{\tilde{E}_1} \varphi(z)d\eta(z) \\ &= m_S(\tilde{E}_1 \exp_{S^*}(U)) = m_S(h_0\tilde{E}_1 \exp_{S^*}(U)) = m_S(\tilde{E}_2 \exp_{S^*}(U)) \\ &= \int_{\tilde{E}_2} \varphi(z)d\eta(z) \leq c_2\eta(\tilde{E}_2) = c_2\eta(\tilde{E}_1). \end{aligned} \quad (6.22)$$

□

Our aim now is to define the relevant family of tubes around the orbit $y_\omega L_\omega$ that will be of use to us. The first stage is to choose the correct linear complement from which the tubes will come.

6.4.2. Choosing the linear complement. When we come to argue the validity of Theorem 4.8(1),(2) for a given admissible radius h , we may assume without loss of generality that S is the smallest set of primes for which $h \in \mathcal{O}_S^\times$. Hence, without loss of generality we may (and will) assume that h is divisible by all the primes in S . We refer to such a radius h as having *full support*. The assumption that an admissible radius has full support is equivalent to the fact that the weak stable algebra of $a_f(h)$ attains the form

$$(\mathfrak{g}_{S^*})_{a_f(h)}^{\text{ws}} = \mathfrak{g}_\infty \oplus_{p \in S} \left\{ \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \in \mathfrak{g}_p \right\}. \quad (6.23)$$

Definition 6.12. Let $V = \bigoplus_{v \in S^*} V_v$ be defined as follows

$$V_\infty = \left\{ \begin{pmatrix} 0 & * \\ * & 0 \end{pmatrix} \in \mathfrak{g}_\infty \right\}; \quad \text{For } p \in S, \quad V_p = \left\{ \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \in \mathfrak{g}_p \right\}.$$

Lemma 6.13. *If the generalized ray \mathcal{L}_ω is non-degenerate then the subspace $V \subset \mathfrak{g}_{S^*}$ from Definition 6.12 is indeed a linear complement of $\text{Lie}(L_\omega)$ which is contained in $(\mathfrak{g}_{S^*})_{a_f(h)}^{\text{ws}}$ for any admissible radius h of full support.*

Proof. The fact that $V \subset (\mathfrak{g}_{S^*})_{a_f(h)}^{\text{ws}}$ follows from the discussion preceding Definition 6.12.

Recall that $L_\omega = A_\infty \times H_\omega$ where $H_\omega = \omega^{-1} \overline{\langle \gamma_x \rangle}_{G_S} \omega$ (see (6.5),(6.6)). Writing $\text{Lie}(L_\omega) = \bigoplus_{S^*} \mathfrak{l}_v$ we see that V_∞ indeed complements \mathfrak{l}_∞ . Let \mathbb{T} be the algebraic subgroup of \mathbb{G} defined as the Zariski closure of the group generated by γ_x . It is a one dimensional torus and H_ω is a compact open subgroup of the conjugation $\omega^{-1} \mathbb{T} (\prod_{v \in S} \mathbb{Z}_v) \omega$. It follows that for any $v \in S$ the dimension of \mathfrak{l}_v is 1 and so in order to argue that it complements V_v we only need to argue that the inclusion $\mathfrak{l}_v \subset V_v$ does not hold. Such an inclusion would imply that there is a neighborhood of the identity in H_ω that consists of upper triangular matrices, which in turn would imply that a certain power of $\omega^{-1} \gamma_x \omega$ is upper triangular, contradicting the assumption that the generalized ray is non-degenerate. □

Henceforth, when speaking about a linear complement V to $\text{Lie}(L_\omega)$, we shall refer only to the subspace from Definition 6.12.

Remark 6.14. Because of the inclusion $V \subset (\mathfrak{g}_{S^*})_{a_f(h)}^{\text{ws}}$ (for any admissible radius h of full support), we conclude that if $U_0 \subset V$ is a small enough ball around zero, the conjugation $U_0^{a_f(h)}$ will be contained in the domain of \exp_{S^*} . This implies that for any $U \subset U_0$ the identity $(\exp_{S^*} U)^{a_f(h)} = \exp_{S^*}(U^{a_f(h)})$ holds. It follows that if \mathcal{T} is a tube of width U coming from V around $y_\omega L_\omega$, then if the width U is chosen within U_0 , the pushed tube $\mathcal{T}a_f(h)$ satisfies

$$\mathcal{T}a_f(h) = y_\omega L_\omega \exp_S(U)a_f(h) = y_{\omega,h} L_{\omega,h} \exp_S(U^{a_f(h)}). \quad (6.24)$$

That is, $\mathcal{T}a_f(h)$ is a tube of width $U^{a_f(h)}$ around the compact orbit $y_{\omega,h} L_{\omega,h}$. Below, we will make the implicit assumption that all the widths considered are contained in the ball U_0 .

6.4.3. *The tubes $\mathcal{T}_\omega^\delta$.* As explained above, we will need to construct tubes with shrinking real width component and with control on the subgroup of K_S that stabilizes them. After describing this family of tubes we state a few lemmas that describe their relevant properties. The proofs of these lemmas will be postponed till after concluding the proof of Theorem 4.8.

Let us denote by \tilde{B} a compact open subgroup of the group of upper triangular elements in K_S that lies in the domain of \log_S and for which Remark 6.14 applies (that is, all conjugations $\tilde{B}^{a_f(h)}$ are in the domain of \log_S for admissible radii h of full support). For $\delta > 0$ let $B_\delta^{V_\infty}$ be the ball of radius δ around 0 in the ∞ -component of the linear complement V from Definition 6.12.

Lemma 6.15. *There exists $\hat{\delta} > 0$ and an open compact subgroup $B = \prod_S B_p$ of \tilde{B} , such that for all $\delta < \hat{\delta}$, if we let $U^\delta = B_\delta^{V_\infty} \times \log_S(B)$, then for any $\omega \in K_S$ such that the generalized ray \mathcal{L}_ω is non-degenerate, the set $\mathcal{T}_\omega^\delta = y_\omega L_\omega \exp_S(U^\delta)$ is a tube around $y_\omega L_\omega$; that is, the map $y_\omega L_\omega \times U^\delta \rightarrow \mathcal{T}_\omega^\delta$ is a homeomorphism and the set $\mathcal{T}_\omega^\delta \subset X_S$ is open. Furthermore, the choice of $B, \hat{\delta}$ depends only on the original class x and the set of places S at hand. In particular, they are independent of ω .*

Lemma 6.16. *Let $B, \hat{\delta}$ be as in Lemma 6.15 and let $\omega \in K_S$ be such that the generalized ray \mathcal{L}_ω is non-degenerate.*

- (1) *There exists an open compact product subgroup $K^* = \prod_S K_v^* < K_S$ which stabilizes the tube $\mathcal{T}_\omega^\delta$ for any $\delta < \hat{\delta}$; that is $\mathcal{T}_\omega^\delta k = \mathcal{T}_\omega^\delta$ for any $k \in K^*, \delta < \hat{\delta}$. Moreover, if x is non-split, we may choose K^* to be independent of ω .*
- (2) *The measures $m_S(\mathcal{T}_\omega^\delta)$ satisfy $m_S(\mathcal{T}_\omega^\delta) \gg_{x,S,\mathcal{L}_\omega} \delta^2$. If x is non-split, the implicit constant may be chosen to be independent of the generalized ray.*

6.5. Concluding the main part of the proof.

Proof of parts(1),(2) of Theorem 4.8. We follow the strategy presented in §6.2 and use freely all the notation introduced so far. Let h be an admissible radius and assume without loss of generality that it is of full support. Let $\varphi_0 \in \text{Lip}_\kappa(X_\infty) \cap L^2(X_\infty, m_\infty)$.

We let $\varphi = \varphi_0 \circ \pi$ be the lift of φ_0 to X_S . As $\int_{X_\infty} \varphi_0 dm_\infty = \int_{X_S} \varphi dm_S$, we see by Lemma 6.4 that part (1) of the theorem will follow once we prove

$$\left| \int_{X_S} \varphi d\eta_{\omega,h} - \int_{X_S} \varphi dm_S \right| \ll_{x,S,\mathcal{L}_\omega,\epsilon} \max \{ \|\varphi\|_2, \kappa \} h^{-\frac{\delta_0}{2} + \epsilon}. \quad (6.25)$$

Part (2) will follow once we establish that in the non-split case, the implicit constant in (6.25) may be chosen independent of the generalized ray. Let $V < \mathfrak{g}_{S^*}$ be the linear complement from Definition 6.12. We apply Lemma 6.16 and use the notation introduced there to obtain a family of tubes $\mathcal{T}_\omega^\delta$ around $y_\omega L_\omega$ coming from V .

We denote $\mathcal{T}_{\omega,h}^\delta$ the (pushed) tube $\mathcal{T}_\omega^\delta a_f(h)$ around the orbit $y_{\omega,h} L_{\omega,h}$, and $m_{\mathcal{T}_{\omega,h}^\delta}$ the normalized restriction of m_S to $\mathcal{T}_{\omega,h}^\delta$. The width of⁸ $\mathcal{T}_{\omega,h}^\delta$ is $U^{\delta,h} = (U^\delta)^{a_f(h)}$, where U^δ is as in Lemma 6.15 (see Remark 6.14). We have

$$\left| \int_{X_S} \varphi d\eta_{\omega,h} - \int_{X_S} \varphi dm_S \right| \leq \underbrace{\left| \int_{X_S} \varphi d\eta_{\omega,h} - \int_{X_S} \varphi dm_{\mathcal{T}_{\omega,h}^\delta} \right|}_{(*)} + \underbrace{\left| \int_{X_S} \varphi dm_{\mathcal{T}_{\omega,h}^\delta} - \int_{X_S} \varphi dm_S \right|}_{(**)}. \quad (6.26)$$

To estimate (*) we define $\tilde{\varphi}_{\delta,h} : \mathcal{T}_{\omega,h}^\delta \rightarrow \mathbb{C}$ by $\tilde{\varphi}_{\delta,h}(z \exp_{S^*} u) = \varphi(z)$ for $z \in y_{\omega,h} L_{\omega,h}$, $u \in U^{\delta,h}$ and extend it to be zero outside the tube $\mathcal{T}_{\omega,h}^\delta$ to obtain a function on X_S . By Lemma 6.11 it follows that

$$\int_{X_S} \tilde{\varphi}_{\delta,h} dm_{\mathcal{T}_{\omega,h}^\delta} = \int_{y_{\omega,h} L_{\omega,h}} \int_{U^{\delta,h}} \varphi(z) F(z, u) dm_{U^{\delta,h}}(u) d\eta_{\omega,h}(z) = \int_{X_S} \varphi d\eta_{\omega,h} \quad (6.27)$$

We therefore have the following estimate for (*)

$$\begin{aligned} (*) &= \left| \int_{X_S} \tilde{\varphi}_{\delta,h} dm_{\mathcal{T}_{\omega,h}^\delta} - \int_{X_S} \varphi dm_{\mathcal{T}_{\omega,h}^\delta} \right| \\ &\leq \max \{ |\varphi(y \exp_{S^*}(w)) - \varphi(y)| : y \in y_{\omega,h} L_{\omega,h}, w \in U^{\delta,h} \}. \end{aligned} \quad (6.28)$$

Note that if we write $w \in U^{\delta,h}$ as (w_∞, w_f) , then for any $y \in y_{\omega,h} L_{\omega,h}$, $\varphi(y \exp_{S^*}(w)) = \varphi_0(\pi(y) \exp_\infty(w_\infty))$ by the K_S -invariance of φ . As the maps induced by the actions of elements of the form $\exp_\infty(w_\infty)$, $\|w_\infty\| < 1$ are all Lipschitz with some uniform Lipschitz constant c_1 , the distance between $\pi(y)$ and $\pi(y) \exp_\infty(w_\infty)$ is $\leq c_1 \delta$ and so by the Lipschitz assumption of φ_0 we obtain

$$(*) \leq c_1 \kappa \delta. \quad (6.29)$$

We now estimate (**). Let $\mathcal{H} = L^2(X_S, m_S)$ and denote

$$w_1 = \varphi, \quad w_2 = \frac{1}{m_S(\mathcal{T}_\omega^\delta)} \chi_{\mathcal{T}_\omega^\delta}.$$

⁸The reader should not confuse the superscript δ with our notation for conjugation.

In order to appeal to Theorem 6.6 we observe that w_1 is K_S -fixed and w_2 is K^* -fixed, where K^* is as in Lemma 6.16. By Lemma 6.16 the index $d = [K_S : K^*]$ depends only on x, S , and \mathcal{L}_ω and in the non-split case could be bounded by a number independent of the generalized ray. As for the norms, $\|w_1\| = \|\varphi\|$, and for w_2 we have $\|w_2\| = m_S(\mathcal{T}_\omega^\delta)^{-\frac{1}{2}}$. By Lemma 6.16 we have that $m_S(\mathcal{T}_\omega^\delta) \gg_{x,S,\mathcal{L}_\omega} \delta^2$ and so $\|w_2\| \ll_{x,S,\mathcal{L}_\omega} \delta^{-1}$. Furthermore, in the non-split case, the implicit constant can be taken to be independent of the generalized ray.

It now follows from Theorem 6.6 that

$$\begin{aligned}
 (**) &= \left| \langle \varphi, a_f(\mathfrak{h})^{-1} \left(\frac{1}{m_S(\mathcal{T}_\omega^\delta)} \chi_{\mathcal{T}_\omega^\delta} \right) \rangle - \int_{X_S} \varphi dm_S \right| & (6.30) \\
 &= |\langle a_f(\mathfrak{h})w_1, w_2 \rangle - \langle w_1, 1 \rangle \langle 1, w_2 \rangle| \\
 &\ll_{x,S,\mathcal{L}_\omega,\epsilon} \|\varphi\|_2 \delta^{-1} \mathfrak{h}^{-\delta_0+\epsilon},
 \end{aligned}$$

and that in the non-split case the implicit constant can be taken independent of the generalized ray. Combining (6.26), (6.29), (6.30), and choosing $\delta = c \mathfrak{h}^{\frac{1}{2}(-\delta_0+\epsilon)}$ (the meaning of c will become clear in a moment) we obtain (6.25) as desired (with ϵ replaced by $\frac{\epsilon}{2}$). Here the constant c is chosen to protect us from the possible finitely many \mathfrak{h} 's for which the inequality $\mathfrak{h}^{\frac{1}{2}(-\delta_0+\epsilon)} < \hat{\delta}$ does not hold ($\hat{\delta}$ as in Lemma 6.15). Note that indeed, the constant c depends only on $\hat{\delta}, S$, and ϵ . By Lemma 6.15 we see that it actually depends on x, S , and ϵ . This concludes the proof of Theorem 4.8. \square

6.6. Proofs of Lemmas 6.15, 6.16. We shall need the following auxiliary lemma which we leave without proof

Lemma 6.17. *There exists a neighborhood of the identity $W \subset G_{S^*}$, depending only on the class x , such that for any $\omega \in K_S$ and any $g \in W$, if $y_\omega L_\omega g \cap y_\omega L_\omega \neq \emptyset$ then $g \in L_\omega$.*

Proof of Lemma 6.15. The first restriction we impose on $\hat{\delta}$ is that it will be small enough so that in the real component, the map $(s, u) \mapsto \exp_\infty(s) \cdot \exp_\infty(u)$ from $B_\delta^{\text{Lie}(A_\infty)} \times B_\delta^{V_\infty} \rightarrow G_\infty$ is a homeomorphism onto its open image. Choose $B = \prod_S B_p$ to be any product compact open subgroup of \tilde{B} and define U^δ as in the statement of the Lemma. At this stage we observe that for any $\delta < \hat{\delta}$ the map $L_\omega \times U^\delta \rightarrow G_{S^*}$ given by $(g, u) \mapsto g \exp_{S^*}(u)$ has an open image. To see this, note that the image is a product of open sets in each component: In the real component the image equals $A_\infty \cdot \exp_\infty B_\delta^{V_\infty}$ which is open by the choice of $\hat{\delta}$, while for any finite place $p \in S$, the p 'th component of the image is $(H_\omega)_p \cdot B_p$ which is seen to be open in the following way: Because of the fact that $V_p = \text{Lie}(B_p)$ is a linear complement to $\text{Lie}((H_\omega)_p)$, the product $(H_\omega)_p \cdot B_p$ clearly contains an open neighborhood of the identity in G_p . It now follows from the fact that both $(H_\omega)_p, B_p$ are groups, that their product is actually an open set.

The above establishes in particular, that the set $\mathcal{T}_\omega^\delta = y_\omega L_\omega \exp_{S^*}(U^\delta) \subset X_S$ is open. It follows that in order to conclude that $\mathcal{T}_\omega^\delta$ is indeed a tube around $y_\omega L_\omega$, we only need to argue the injectivity of the map $(z, u) \mapsto z \exp_{S^*}(u)$ from $y_\omega L_\omega \times U^\delta$ to X_S . We denote this map by ψ_ω .

The second condition which we impose on $\hat{\delta}$ and on the choice of B is that the product $\left(\exp_\infty(B_\delta^{V_\infty})\right)^2 \cdot B^2 \subset W$, where W is as in Lemma 6.17. Assuming the injectivity of ψ_ω fails, we obtain elements $u_\infty^{(i)} \in B_\delta^{V_\infty}, b_i \in B, i = 1, 2$ and a non-trivial intersection of the form

$$y_\omega L_\omega \exp_\infty(u_\infty^{(1)})b_1 \cap y_\omega L_\omega \exp_\infty(u_\infty^{(2)})b_2.$$

This shows that $y_\omega L_\omega \cap y_\omega L_\omega \exp_\infty(u_\infty^{(2)}) \exp_\infty(-u_\infty^{(1)})b_2b_1^{-1} \neq \emptyset$. It now follows from our choice of $\hat{\delta}$ and B (by Lemma 6.17) that $\exp_\infty(u_\infty^{(2)}) \exp_\infty(-u_\infty^{(1)}) \in A_\infty$ and that $b_2b_1^{-1} \in H_\omega$. As B is a group which intersects H_ω trivially (this is our assumption that the generalized ray \mathcal{L}_ω is nondegenerate), we conclude that $b_1 = b_2$. Furthermore, from the fact that $\text{Lie}(A_\infty) \oplus V_\infty = \mathfrak{g}_\infty$, it is straightforward to deduce that if $\hat{\delta}$ is chosen small enough, then the inclusion $\exp_\infty(u_\infty^{(2)}) \exp_\infty(-u_\infty^{(1)}) \in A_\infty$ implies that $u_\infty^{(1)} = u_\infty^{(2)}$. This establishes the injectivity of ψ_ω as desired. \square

Proof of Lemma 6.16. We first argue the validity of part (1). As pointed out in the proof of Lemma 6.15 above, if ω is such that \mathcal{L}_ω is non-degenerate, then the set $H_\omega \cdot B \subset G_S$ is open (here B is as in Lemma 6.15). Moreover, as B is compact, there exist a neighborhood of the identity in G_S , and in particular, a compact open subgroup $K^* = \prod_S K_p^*$, with the property that for any $k \in K^*$ we have $Bk \subset H_\omega B$. We now claim that for any tube $\mathcal{T}_\omega^\delta$ as in Lemma 6.15 we have $\mathcal{T}_\omega^\delta k = \mathcal{T}_\omega^\delta$. To argue the inclusion \subset we note the following

$$\mathcal{T}_\omega^\delta k = y_\omega L_\omega \exp_\infty(B_\delta^{V_\infty})Bk \subset y_\omega L_\omega \exp_\infty(B_\delta^{V_\infty})H_\omega B = y_\omega L_\omega \exp_\infty(B_\delta^{V_\infty})B = \mathcal{T}_\omega^\delta.$$

The opposite inclusion follows by switching k with k^{-1} .

If x is non-spilt, it is not hard to see that the intersection $\cap_{\omega \in K_S} (H_\omega \cdot B)$ contains an open neighborhood around e_S . It then readily follows that this intersection contains an open neighborhood of B . We conclude similarly to the argument presented above that the group K^* may be chosen to work for all the ω 's simultaneously.

We briefly argue part (2) of the lemma. For each relevant ω , it is not hard to see that the volume of the tube $m_S(\mathcal{T}_\omega^\delta)$ satisfies $c_1 m_{V_\infty}(B_\delta^{V_\infty}) \leq m_S(\mathcal{T}_\omega^\delta) \leq c_2 m_{V_\infty}(B_\delta^{V_\infty})$, where the constants c_1, c_2 are determined by the volume of the orbit $y_\omega L_\omega$ and the position of the linear space V , from which the width is coming, with respect to $\text{Lie}(L_\omega)$. As the 2-dimensional volume $m_{V_\infty}(B_\delta^{V_\infty})$ is proportional to δ^2 , the claim regarding a single ω follows. In the non-split case, as the Lie algebras $\text{Lie}(L_\omega)$ are uniformly transverse to V , the constant c_1 above can be taken to be uniform for all ω which finishes the proof. \square

7. PROOF OF LEMMA 5.1

Let S be a finite set of primes. For an element $\delta \in G_S$ we denote $\Sigma_\delta \stackrel{\text{def}}{=} \overline{\langle \delta \rangle}_{G_S}$ and we say that δ is of *compact type* if Σ_δ is a compact group.

Definition 7.1. Let $\delta \in G_S$ be an element of compact type. Let us denote for any admissible radius h by $k_h(\delta)$ the minimal positive integer k for which δ^k belongs to the compact open subgroup $a_f(h)K_S a_f(h^{-1})$. Equivalently, $k_h(\delta) = [\Sigma_\delta : \Sigma_\delta \cap a_f(h)K_S a_f(h^{-1})]$.

Lemma 7.2. *Let $\delta \in G_S$ be an element of compact type such that no power of δ has an upper triangular component. Then, the ratio $e_h(\delta) \stackrel{\text{def}}{=} \frac{k_h(\delta)}{h}$ attains only finitely many values and furthermore, if h_n is a divisibility sequence (i.e. $h_n \mid h_{n+1}$), then the sequence $e_{h_n}(\delta)$ stabilizes.*

Proof. The strategy is to reduce to the case where S consists of a single prime p and $\delta \in K_p$ satisfies a certain congruence assumption. In this case the description of $k_h(\delta)$ becomes explicit and simple.

Step 1 - Reduction to one prime. Let us denote for an admissible radius h by $n_p(h)$ the integers satisfying $h = \prod_{p \in S} p^{n_p(h)}$. We have the equality

$$a_f(h)K_S a_f(h^{-1}) = \prod_{p \in S} a_p(p^{n_p(h)})K_p a_p(p^{-n_p(h)}),$$

and so, if we denote $\delta = (\delta_p)_{p \in S}$, then $k_h(\delta) = \text{lcm} \{k_{p^{n_p(h)}}(\delta_p) : p \in S\}$. From here it follows that the statement of the Lemma for a general finite set of primes S follows from the corresponding statement for a single prime. Therefore, henceforth we assume that S consists of a single prime p and our objective is to show that the sequence $k_{p^n}(\delta)/p^n$ stabilizes. For simplicity we denote $K_{p,n} \stackrel{\text{def}}{=} a_f(p^n)K_p a_f(p^{-n})$.

Step 2 - Replacing δ by a power. We wish to prove that for any $\ell > 0$ the statement of the Lemma for δ is equivalent to the statement of the Lemma for δ^ℓ . We first prove that $\cap_n(\Sigma_\delta \cap K_{p,n}) = \{e\}$. To see this, note that as

$$K_{p,n} = \left\{ \begin{pmatrix} a & p^{-n}b \\ p^n c & d \end{pmatrix} : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K_p \right\}, \quad (7.1)$$

we see that $\cap_n K_{p,n} \subset \left\{ \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \right\}$, and so if the intersection $\cap_n(\Sigma_\delta \cap K_{p,n})$ was non-trivial, then it would imply that Σ_δ contains a non-trivial upper triangular element, which in turn would imply that its (one-dimensional) Lie algebra is upper triangular. This is equivalent to saying that a power of δ is upper triangular, which contradicts our assumption on δ .

We conclude that for any $\ell > 0$, as $\Sigma_{\delta^\ell} < \Sigma_\delta$ is an open subgroup, for n large enough we have that $\Sigma_\delta \cap K_{p,n} = \Sigma_{\delta^\ell} \cap K_{p,n}$. This implies that for n large enough

$$k_{p^n}(\delta) = [\Sigma_\delta : \Sigma_{\delta^\ell} \cap K_{p,n}] = [\Sigma_\delta : \Sigma_{\delta^\ell}] \cdot [\Sigma_{\delta^\ell} : \Sigma_{\delta^\ell} \cap K_{p,n}] = [\Sigma_\delta : \Sigma_{\delta^\ell}] k_{p^n}(\delta^\ell), \quad (7.2)$$

and so, in particular, the sequence $k_{p^n}(\delta)/p^n$ stabilizes if and only if the sequence $k_{p^n}(\delta^\ell)/p^n$ does, as desired.

Step 3 - Concluding the proof. By Step 2 we may assume (by replacing δ by a suitable power of δ if necessary), that δ belongs to the (open) subgroup of K_p consisting of elements congruent to the identity modulo p^2 (or said differently, to the kernel of the natural homomorphism from K_p onto $\text{PGL}_2(\mathbb{Z}/p^2\mathbb{Z})$). Denote $B_n \stackrel{\text{def}}{=} K_p \cap K_{p,n}$. A direct calculation shows $B_n = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K_p : \frac{c}{p^n} \in \mathbb{Z}_p \right\}$. Under the above assumption we have that $k_{p^n}(\delta)$ is the order of (the image of) δ in the finite cyclic quotient $\Sigma_\delta / (\Sigma_\delta \cap B_n)$. Furthermore, as $B_{n+1} < B_n$, the divisibility relation $k_{p^n}(\delta) \mid k_{p^{n+1}}(\delta)$ holds. Let $n_0 \stackrel{\text{def}}{=} \max \{n > 0 : \delta \in B_n\}$.

Our assumption that δ is not upper triangular implies that n_0 is well defined. The proof of the Lemma will be concluded once we establish the following

Claim: For any $n \geq n_0$ we have $k_{p^n}(\delta) = p^{n-n_0}$ and moreover, $\delta^{p^{n-n_0}} \in B_n \setminus B_{n+1}$.

We prove this claim by induction on n . For $n = n_0$ the validity of the claim follows from the choice of n_0 . Let us assume it holds for n . As mentioned above, the divisibility relation $k_{p^n}(\delta) | k_{p^{n+1}}(\delta)$ holds. Moreover, from our inductive hypothesis saying that $\delta^{k_{p^n}(\delta)} \notin B_{n+1}$ we know that this divisibility relation is strict. It follows that $k_{p^{n+1}}(\delta) = j_0 p^{n-n_0}$ where j_0 is the minimal positive integer j so that $\delta^{j p^{n-n_0}} \in B_{n+1}$, or said differently, such that the bottom left coordinate of $\delta^{j p^{n-n_0}}$ is divisible by p^{n+1} in \mathbb{Z}_p . We will be finished once we show two things:

- (1) First, that $j_0 = p$ and so $k_{p^{n+1}}(\delta) = p^{n+1-n_0}$,
- (2) and second, that the bottom left coordinate of $\delta^{p^{n+1-n_0}}$ is not divisible by p^{n+2} and so $\delta^{k_{p^{n+1}}(\delta)} \in B_{n+1} \setminus B_{n+2}$ which completes the inductive step.

Consider the sequence $\delta^{j p^{n-n_0}} \stackrel{\text{def}}{=} \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}$, $j = 1, 2, \dots$ and note the recursive relation

$$c_{j+1} = c_1 a_j + c_j d_1. \quad (7.3)$$

We expand c_1 to a power series in \mathbb{Z}_p and use the inductive assumption that $\delta^{p^{n-n_0}} \in B_n \setminus B_{n+1}$ and write

$$c_1 = m_1 p^n + m_2 p^{n+1} + u p^{n+2}, \quad (7.4)$$

where $m_1 \in \{1, 2, \dots, p-1\}$, $m_2 \in \{0, 1, \dots, p-1\}$, $u \in \mathbb{Z}_p$. We claim that for any $1 \leq j$ we have

$$c_j = j m_1 p^n + j m_2 p^{n+1} + u_j p^{n+2} \text{ where } u_j \in \mathbb{Z}_p. \quad (7.5)$$

The validity of (1),(2) follows at once from (7.5) and the fact that $m_1 \in \{1, 2, \dots, p-1\}$. We prove the validity of (7.5) by induction on j . For $j = 1$, this is exactly (7.4). Now assume it holds for j and write (using the congruence assumption on δ)

$$a_j = 1 + p^2 A, \quad d_j = 1 + p^2 D, \quad A, D \in \mathbb{Z}_p.$$

Plugging this and (7.4),(7.5) into the recursive relation (7.3) we see that indeed

$$c_{j+1} = (j+1)m_1 p^n + (j+1)m_2 p^{n+1} + p^{n+2}(\dots)$$

as desired. This completes the proof of the Claim and by that concludes the proof of the Lemma. \square

Remark 7.3. It will be useful later on to note the following: A careful look at the argument giving Lemma 7.2 shows that for a fixed $\delta \in G_S$ of compact type, we have that there exists a positive constant c such that $c \leq e_h(\delta)$ for any admissible radius h , where c depends only on two things:

- (1) The power k_0 which we need to raise δ to so that each component $(\delta)_p^{k_0}$ will be in K_p and congruent to the identity mod p^2 .
- (2) The maximal admissible radius $h = \prod_{p \in S} p^{n_p}$ for which for any $p \in S$, $(\delta)_p^{k_0} \in B_{n_p}$ (this h measures how close δ^{k_0} is to being upper triangular).

Before turning to the proof of Lemma 5.1 we make yet another remark which will be used in the course of its proof.

Remark 7.4. Given a class $x \in X_\infty$ with a periodic A_∞ -orbit and a representative $\Lambda_x \in x$, then the matrix $a_\infty(t_x) = \text{diag}\left(e^{\frac{t_x}{2}}, e^{-\frac{t_x}{2}}\right)$ stabilizes the lattice Λ_x (that is, as a subset of \mathbb{R}^2 , $\Lambda_x = \Lambda_x a_\infty(t_x)$). As Λ_x is a lattice, it follows that $a_\infty(t_x)$ is conjugate to an integer matrix and so its eigenvalues $e^{\pm \frac{t_x}{2}}$, are algebraic integers of degree 2. The quadratic extension \mathbb{F}_x from Definition 4.5 is the one generated by them. As these eigenvalues are positive Galois conjugates whose product is equal to 1, we conclude furthermore that they belong to the group of totally positive units in the ring of integers of \mathbb{F}_x (i.e. units all of whose embeddings into the reals are positive). As such, by Dirichlet's unit theorem, they are integer powers of the fundamental unit of this field. In fact, we shall slightly abuse the classical terminology and use the term *fundamental unit* to refer to the unit in the ring of integers which is of absolute value > 1 and which generates the group of totally positive units. If the fundamental unit is $\epsilon = e^{\frac{t_0}{2}}$, then the reader will easily verify that the image $\Lambda_x a_\infty(t_0)$ is contained in the \mathbb{Q} -span of Λ_x . This shows that if we write $x = \Gamma_\infty g_x$, then there is a rational matrix δ_x which solves $\delta_x g_x = g_x a_\infty(t_0)$ and in fact, $t_x = k t_0$ where k is the minimal positive integer such that δ_x^k is an integer matrix.

Proof of Lemma 5.1. (1). A short counting argument shows that the cardinality of the sphere $\mathcal{S}_h(x)$ is proportional to h (were the proportionality constant depends on S). For each x' on the sphere, let $s_{x'}$ be the minimal positive number such that $x' a_\infty(s)$ returns to the sphere. The total length is then $\mathbf{t}_x(h) = \sum_{x' \in \mathcal{S}_h(x)} s_{x'}$. We will show below that for any $x' \in \mathcal{S}_h(x)$ $s_{x'} \leq t_x$. This will establish the inequality $\mathbf{t}_h(x) \ll_{x,S} h$ which is half of the statement in part (1) of the Lemma. The other half, namely the inequality $h \ll_{x,S} \mathbf{t}_x(h)$, actually follows from part (2) of the Lemma.

Let $x' \in \mathcal{S}_h(x)$ be given. By Lemma 4.1 we see that there exists $y \in \pi^{-1}(x)$ such that $x' = \pi(y a_f(h))$. As π intertwines the A_∞ -actions on X_S, X_∞ we see that $x = x a_\infty(t_x) = \pi(y a_\infty(t_x))$ and so if we let $\tilde{y} = y a_\infty(t_x)$ then $\tilde{y} \in \pi^{-1}(x)$ and again by Lemma 4.1 we have that $x'' = \pi(\tilde{y} a_f(h)) \in \mathcal{S}_h(x)$. The following calculation then shows that indeed $s_{x'} \leq t_x$ as was claimed:

$$x' a_\infty(t_x) = \pi(y a_f(h) a_\infty(t_x)) = \pi(y a_\infty(t_x) a_f(h)) = \pi(\tilde{y} a_f(h)) = x'' \in \mathcal{S}_h(x).$$

(2). Let $\mathcal{L} = \mathcal{L}_{g_x, \omega}$ be a non-degenerate generalized ray of $\mathcal{G}_S(x)$ (here $x = \Gamma_\infty g_x$ and $\omega \in K_S$). Let $t_0 > 0$ be such that $e^{\frac{t_0}{2}}$ is the fundamental unit of \mathbb{F}_x as in Remark 7.4. In the notation of the same remark, let δ_x be the rational matrix satisfying $\delta_x g_x = g_x a_\infty(t_0)$. We replace the set of places S by a bigger set if necessary \tilde{S} , so that $\delta_x \in \Gamma_{\tilde{S}}$. We then consider the bigger graph $\mathcal{G}_{\tilde{S}}(x)$ which contains the original graph and we further consider its following generalized ray: Write $\tilde{S} = S \cup T$ and define $\tilde{\omega} \in K_{\tilde{S}}$ to be identical to ω in the components corresponding to the primes in S and equal the identity in the components corresponding to primes in T . We then define $\tilde{\mathcal{L}}$ to be the generalized ray $\mathcal{L}_{g_x, \tilde{\omega}}$ of $\mathcal{G}_{\tilde{S}}(x)$. Note that because of the way we defined $\tilde{\omega}$, the generalized ray $\tilde{\mathcal{L}}$ is non-degenerate as well.

Denote as before by x_h the class in $\tilde{\mathcal{L}} \cap \mathcal{S}_h(x)$. We are interested in analyzing the length t_{x_h} of the orbit $x_h A_\infty$. By Remark 7.4, there exists a positive integer \widehat{k}_h satisfying

$$t_{x_h} = \widehat{k}_h t_0. \quad (7.6)$$

In fact, for later purposes, note that in our discussion x and the representative g_x are fixed but we will play with the ray later on, i.e. with the choice of $\tilde{\omega}$ (which in our setting is defined by ω), and so we should actually record the dependency in $\tilde{\omega}$ in our notation and denote $\widehat{k}_h(\tilde{\omega})$. The function $k_h(\cdot)$ from Definition 7.1 and $\widehat{k}_h(\cdot)$ are closely related as will be seen below.

The number $\widehat{k}_h(\tilde{\omega})$ is by definition the minimal positive integer such that $x_h a_\infty(kt_0) = x_h$ or, if we prefer working in the extension $X_{\tilde{S}}$, it is the minimal positive integer so that $\Gamma_{\tilde{S}}(g_x, \tilde{\omega}) a_f(h) a_\infty(kt_0)$ returns to the fiber $\pi^{-1}(x_h)$. Because of the identity $\delta_x g_x = g_x a_\infty(t_0)$ and the fact that $\delta_x \in \Gamma_{\tilde{S}}$ we see that $\Gamma_{\tilde{S}}(g_x, \tilde{\omega}) a_f(h) a_\infty(kt_0) = \Gamma_{\tilde{S}}(g_x, \delta_x^{-k} \tilde{\omega} a_f(h))$, and so this point lies in the same fiber as $\Gamma_{\tilde{S}}(g_x, \tilde{\omega} a_f(h))$ (i.e. above x_h) if and only if the quotient $a_f(h^{-1}) \tilde{\omega}^{-1} \delta_x^k \tilde{\omega} a_f(h)$ belongs to $K_{\tilde{S}}$ (see Remark 3.1). That is, $\widehat{k}_h(\tilde{\omega})$ is the minimal positive integer k for which the $(\tilde{\omega}^{-1} \delta_x \tilde{\omega})^k \in a_f(h) K_{\tilde{S}} a_f(h^{-1})$. This establishes the equality

$$k_h(\delta_x^{\tilde{\omega}}) = \widehat{k}_h(\tilde{\omega}).$$

The validity of part (2) of the Lemma now follows immediately from Lemma 7.2 and (7.6) which together imply $c_{\mathcal{L}_{g_x, \omega}}(h) = t_0 e_h(\delta_x^{\tilde{\omega}})$.

(3). Assume first that x is non-split with respect to S . As noted in Remark 7.3 the lower bound for the function $h \mapsto e_h(\delta_x^{\tilde{\omega}})$, which gives us the lower bounds for the functions $c_{\mathcal{L}_{g_x, \omega}}(h)$, depends only on two things:

- (1) The smallest power k_0 for which δ_x belongs to the subgroup of $K_{\tilde{S}}$ consisting of elements congruent to the identity modulo p^2 in each component (note that we may ignore the conjugation by $\tilde{\omega}$ as this is a normal subgroup of $K_{\tilde{S}}$).
- (2) The p -adic norms $|c_p|_p$, where c_p is the left bottom coordinate of the p -component of $(\tilde{\omega}^{-1} \delta_x \tilde{\omega})^{k_0}$ where $p \in \tilde{S}$.

It is clear that k_0 depends only on x and the original set of primes S and does not vary with ω (i.e. with the generalized ray). Also, for primes $p \in S$, the p -adic norm $|c_p|_p$ is bounded from below as ω ranges over K_S because x is non split. Finally, for the primes $p \in \tilde{S} \setminus S$, as the p 'th component of $\tilde{\omega}$ equals the identity, the p 'th component of $(\tilde{\omega}^{-1} \delta_x \tilde{\omega})^{k_0}$ is independent of ω . We conclude that

$$\inf \{c_{\mathcal{L}_{g_x, \omega}}(h) : \omega \in K_S, h \text{ is an admissible radius}\} > 0$$

as desired. We leave it as an exercise to the reader to show that in the split case this infimum equals zero. □

8. APPLICATIONS TO CONTINUED FRACTIONS

In this section we present the necessary terminology and results that will allow us to state and prove our main theorems regarding continued fractions and deduce Theorems 2.8, 2.12.

 8.1. The A_∞ -orbit attached to $\alpha \in \text{QI}$.

Definition 8.1. For $\alpha \in \text{QI}$, let α' denote its Galois conjugate and let

$$g_\alpha \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & \alpha' \\ 1 & 1 \end{pmatrix} \text{ if } \alpha > \alpha', \text{ and } g_\alpha \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha' \\ 1 & -1 \end{pmatrix} \text{ otherwise.} \quad (8.1)$$

Furthermore, let $x_\alpha \stackrel{\text{def}}{=} \Gamma_\infty g_\alpha \in X_\infty$.

Lemma 8.2. *Let $\alpha \in \text{QI}$. Then, the orbit $x_\alpha A_\infty \subset X_\infty$ is periodic.*

Proof. Consider the \mathbb{Z} -module $\Lambda_\alpha = \text{span}_{\mathbb{Z}} \{1, \alpha\}$ in the field $\mathbb{Q}(\alpha)$. There exists a unit ω in the ring of integers which stabilizes Λ_α and furthermore by replacing ω by ω^2 if necessary we may assume that both ω and its Galois conjugate ω' are positive. Note that the diagonal matrix $\text{diag}(\omega, \omega')$ is an element of A_∞ . Let $\gamma = \begin{pmatrix} n & m \\ k & \ell \end{pmatrix} \in \text{GL}_2(\mathbb{Z})$ be the matrix describing the passage from the basis $\{1, \alpha\}$ to the basis $\{\omega, \omega\alpha\}$ of Λ_α . That is

$$\begin{pmatrix} n & m \\ k & \ell \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix} = \begin{pmatrix} \omega\alpha \\ \omega \end{pmatrix}. \quad (8.2)$$

The reader will easily verify now that (8.2) implies that $\gamma g_\alpha = g_\alpha \text{diag}(\omega, \omega')$ or in other words that in the space X_∞ the orbit $x_\alpha A_\infty$ is periodic as desired. \square

Definition 8.3. Given $\alpha \in \text{QI}$, in the spirit of Definition 4.5, we denote $t_\alpha = t_{x_\alpha}$, $\mu_\alpha = \mu_{x_\alpha}$, and $\gamma_\alpha = \gamma_{x_\alpha}$, where γ_{x_α} is defined using the representative g_α of x_α .

Fix a finite set of primes S . For $\alpha \in \text{QI}$ consider the S -Hecke graph $\mathcal{G}_S(x_\alpha)$ and recall that by Corollary 4.2, for $\gamma \in \Gamma_S$ with $\text{ht}(\gamma) = h$ we have that the class $\Gamma_\infty \gamma g_\alpha$ lies on the sphere $\mathcal{S}_h(x_\alpha)$.

Definition 8.4. Let $\omega \in K_S$

- (1) We say that $\gamma \in \Gamma_S$ lies on the generalized ray $\mathcal{L}_{g_\alpha, \omega}$ if $\Gamma_\infty \gamma g_\alpha \in \mathcal{L}_{g_\alpha, \omega}$ and denote this by $\gamma \in [\omega]_{\text{br}}$. As will be explained shortly in Remark 8.5, the question of whether or not $\gamma \in [\omega]_{\text{br}}$ is indeed independent of α as suggested by the notation.
- (2) Similarly to the notation introduced in Remark (4.4)(1), we denote by $x_{\alpha, \omega, h}$ the class in $\mathcal{L}_{g_\alpha, \omega} \cap \mathcal{S}_h$.

Remark 8.5. With the above notation, for $\gamma \in \Gamma_S$ with $\text{ht}(\gamma) = h$ we have that $\gamma \in [\omega]_{\text{br}}$ if and only if $x_{\alpha, \omega, h} = \Gamma_\infty \gamma g_\alpha$. As mentioned in the proof of Corollary 4.2, an element $\gamma \in \Gamma_S$ can be written as $\gamma = \gamma_1 \text{diag}(h, 1) \gamma_2$ with $\gamma_i \in \Gamma_\infty$ and $h = \text{ht}(\gamma)$. It follows that for $\omega \in K_S$ we have that $\gamma \in [\omega]_{\text{br}}$ if and only if $\pi(\Gamma_S(g_\alpha, \gamma^{-1})) = \pi(\Gamma_S(g_\alpha, \omega) a_f(h))$. The latter happens, by Remark 3.1, exactly when $\gamma \omega a_f(h)$ belongs to K_S , or equivalently, when the lower left coordinate of γ_2 is divisible by h in each \mathbb{Z}_p for $p \in S$.

The following Lemma relates the orbit $\{\gamma\alpha : \gamma \in \Gamma_S\} \subset \mathbb{R}$ and the periodic A_∞ -orbits through points of $\mathcal{G}_S(x_\alpha)$.

Lemma 8.6. *Let S be a finite set of primes, $\alpha \in \text{QI}$, $\omega \in K_S$, and $\gamma \in [\omega]_{\text{br}}$, with $\text{ht}(\gamma) = h$. Denote by $\theta \in \Gamma_\infty$ the element $\theta = \text{diag}(1, -1)$. Then, one of the following two relations holds: Either $\mu_{x_\alpha, \omega, h} = \mu_{\gamma\alpha}$, or $\mu_{x_\alpha, \omega, h} = \theta_* \mu_{\gamma\alpha}$.*

Proof. Write $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and recall that the matrix g_α in Definition 8.1 has one of the forms $g_\alpha = \begin{pmatrix} \alpha & \alpha' \\ 1 & 1 \end{pmatrix}$ or $g_\alpha = \begin{pmatrix} \alpha & \alpha' \\ 1 & 1 \end{pmatrix} \theta$. The equation

$$\gamma \begin{pmatrix} \alpha & \alpha' \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \gamma\alpha & (\gamma\alpha)' \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{c\alpha+d} & 0 \\ 0 & \frac{1}{c\alpha'+d} \end{pmatrix}, \quad (8.3)$$

together with the fact that $g_{\gamma\alpha}$ has either the form $\begin{pmatrix} \gamma\alpha & (\gamma\alpha)' \\ 1 & 1 \end{pmatrix}$ or the form $\begin{pmatrix} \gamma\alpha & (\gamma\alpha)' \\ 1 & 1 \end{pmatrix} \theta$, imply that the two points $x_{\alpha, \omega, h} = \Gamma_\infty \gamma g_\alpha$, $x_{\gamma\alpha} = \Gamma_\infty g_{\gamma\alpha}$ are on the same orbit under the group generated by A_∞ and θ in G_∞ . This group contains A_∞ as a subgroup of index 2 and so either the two points are on the same A_∞ -orbit or otherwise, their A_∞ -orbits are related by the action of θ . The translation of the latter statement to the A_∞ -invariant probability measures that are supported on these orbits is exactly the statement sought. \square

The following Theorem relates the measures μ_α from Definition 8.3 to the measures ν_α from Definition 2.3 and will allow us to translate equidistribution results for geodesic loops to statements about periods of c.f.e while controlling error terms. We will use the following terminology

Definition 8.7. For $\alpha \in \text{QI}$, we denote

$$j_\alpha \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if the size } |P_\alpha| \text{ is even,} \\ 2 & \text{if the size } |P_\alpha| \text{ is odd.} \end{cases}$$

Theorem 8.8. *Let $\alpha \in \text{QI}$. There exists an absolute constant $T_0 > 1$ so that if we assume that for some $T > T_0$ the estimate $|\int f d\mu_\alpha - \int f dm_\infty| \leq \max\{\|f\|_2, \kappa\} T^{-1}$ holds for any $f \in \text{Lip}_\kappa(X_\infty) \cap L^2(X_\infty, m_\infty)$, then the following two statements hold*

(1) *For any $f \in \text{Lip}_\kappa([0, 1])$, and any $\epsilon > 0$*

$$\left| \int_0^1 f d\nu_\alpha - \int_0^1 f d\nu_{\text{Gauss}} \right| \ll_\epsilon \max\{\|f\|_\infty, \kappa\} T^{-\frac{1}{3} + \epsilon}.$$

(2) *There exists an absolute constant c_0 such that $\left| \frac{|P_\alpha|}{t_\alpha} - \frac{1}{j_\alpha c_0} \right| \ll_\epsilon T^{-\frac{1}{3} + \epsilon}$.*

Below we will use Theorem 8.8 while postponing its proof to §9.

8.2. Main Theorems regarding continued fractions. The following is the analogue of Theorem 4.8 in the language of continued fractions. We prove it and then deduce Theorem 2.8.

Theorem 8.9. *Let S be a finite set of primes and $\alpha \in \text{QI}$.*

- (1) Let $\omega \in K_S$ be such that the generalized ray $\mathcal{L}_{g_\alpha, \omega}$ is non-degenerate. Then for any $\gamma \in [\omega]_{\text{br}}$, any $\epsilon > 0$, and any $f \in \text{Lip}_\kappa([0, 1])$ the following estimate holds

$$\left| \int_0^1 f d\nu_{\text{Gauss}} - \int_0^1 f d\nu_{\gamma\alpha} \right| \ll_{\mathcal{L}_{g_\alpha, \omega}, S, \epsilon} \max \{ \|f\|_\infty, \kappa \} \text{ht}(\gamma)^{-\frac{\delta_0}{6} + \epsilon}, \quad (8.4)$$

- (2) If all the primes in S do not split in the quadratic extension $\mathbb{Q}(\alpha)$, then the implicit constant in (8.4) may be taken to be independent of the generalized ray.
 (3) On the other hand, if there exists $\omega \in K_S$ such that the generalized ray $\mathcal{L}_{g_\alpha, \omega}$ is degenerate, then there exists a sequence $\gamma_n \in [\omega]_{\text{br}}$ such that $\text{ht}(\gamma_n) \rightarrow \infty$ and the cardinality of the periods $P_{\gamma_n\alpha}$ is bounded. In particular, $\nu_{\gamma_n\alpha} \not\rightarrow \nu_{\text{Gauss}}$.
 (4) Rational generalized rays are always non-degenerate.
 (5) Nevertheless, in case x_α is split, then one cannot take the implicit constant in (8.4) to be uniform along the rational generalized rays.

Proof. (1). Let $\gamma \in [\omega]_{\text{br}}$ and $\epsilon > 0$ be given and denote $h = \text{ht}(\gamma)$. By the corresponding part of Theorem 4.8 we know that for the class $x_{\alpha, \omega, h} \in \mathcal{L}_{g_\alpha, \omega} \cap \mathcal{S}_h(x_\alpha)$, the measure $\mu_{x_{\alpha, \omega, h}}$ satisfies the estimate

$$\left| \int f d\mu_{\alpha, \omega, h} - \int f dm_\infty \right| \ll_{\mathcal{L}_{g_\alpha, \omega}, S, \epsilon} \max \{ \kappa, \|f\|_2 \} h^{-\frac{\delta_0}{2} + \epsilon}, \quad (8.5)$$

for any $f \in \text{Lip}_\kappa(X_\infty) \cap L^2(X_\infty, m_\infty)$. By Lemma 8.6 we know that either $\mu_{\gamma\alpha} = \mu_{x_{\alpha, \omega, h}}$ or $\theta_*\mu_{\gamma\alpha} = \mu_{x_{\alpha, \omega, h}}$. Assume that the first possibility holds. We now apply Theorem 8.8 with $T^{-1} = Ch^{-\frac{\delta_0}{2} + \epsilon}$ where C is the implicit constant in (8.5) and obtain the desired (8.4). One remark is in order here: For finitely many heights h it might happen that this choice of T is not valid as T needs to exceed the absolute constant T_0 from Theorem 8.8. We overcome this problem by choosing the implicit constant in (8.4) to be big enough so that this inequality will hold for these finitely many cases as well.

Assume now that when we apply Lemma 8.6 we obtain that $\theta_*\mu_{\gamma\alpha} = \mu_{x_{\alpha, \omega, h}}$. As θ acts as an isometry of X_∞ we have that (8.5) implies the same estimate for $\mu_{\gamma\alpha}$ replacing $\mu_{x_{\alpha, \omega, h}}$ and the argument concludes as before.

The argument giving part (2) of the Theorem is identical to the one giving part (1) of the Theorem but uses as an input the corresponding part of Theorem 4.8.

For part (3) of the Theorem follows from Theorem 4.8(3) because of the following general fact⁹: For $\beta \in \text{QI}$ $|P_\beta| \ll t_\beta$.

Part (4) of the Theorem is included in Theorem 4.8 and finally, part (5) of the Theorem follows from part (3) of the Theorem in the same way that the corresponding implication of Theorem 4.8 was proved in the beginning of §6. \square

Proof of Theorem 2.8. (1). Note that for $q \in \mathcal{O}_S$ if we define $\gamma_q = \text{diag}(q, 1)$ then $\gamma_q\alpha = q\alpha$. Define $\Omega \subset K_S$ as follows

$$\Omega \stackrel{\text{def}}{=} \{ \omega = (\omega_p)_{p \in S} \in K_S : \omega_p = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \omega_p = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \}. \quad (8.6)$$

⁹This fact will become clear in §9, in fact $|P_\beta| \leq \frac{t_\beta}{\epsilon_0}$, where ϵ_0 is as in Lemma 9.3.

Then, we leave it as an exercise to the reader to verify (using Remark 8.5), that in the notation of Definition 8.4, for any $q \in \mathcal{O}_S$, $\gamma_q \in [\omega]_{\text{br}}$ for some $\omega \in \Omega$. As by Theorem 8.9(4) the finitely many rational generalized rays $\mathcal{L}_{g_{\alpha,\omega}}$, $\omega \in \Omega$ are all non-degenerate we conclude from Theorem 8.9(1) that the estimate (2.2) indeed holds.

(2). Given $\gamma \in \Gamma_S$ choose $\omega \in K_S$ such that $\gamma \in [\omega]_{\text{br}}$. The estimate (2.3) holds with implicit constant depending only on α, S, ϵ but not on the the generalized ray. If we replace γ by $\gamma\gamma_0$ for some choice of $\gamma_0 \in \Gamma_\infty$ we change the generalized ray but this does not effect the right hand side of (8.4) by Theorem 8.9(2). We conclude that the implicit constant does not depend on α but only on the orbit $\iota(\alpha)$. This gives us the desired estimate (2.3).

(3). Let $\delta_n \in \Gamma_S$ be a sequence such that $\nu_{\delta_n\alpha}$ does not converge to ν_{Gauss} , as in part (3) of Theorem 8.9. Write $\delta_n = \gamma'_n \text{diag}(q_n, 1) \gamma_n$, where $\gamma_n, \gamma'_n \in \Gamma_\infty$. By (2.1), $\nu_{\delta_n\alpha} = \nu_{q_n\gamma_n\alpha}$ and so the sequences q_n, γ_n satisfy the statement. \square

The following Theorem is the most general statement we could extract from our analysis regarding the growth of the period. We prove it and then deduce Theorem 2.12.

Theorem 8.10. *Let S be a finite set of primes. There exists a positive function $c(\alpha, \omega, h)$ on the set $\text{QI} \times K_S \times (\mathcal{O}_S^\times \cap \mathbb{N})$ satisfying the following: For any $\alpha \in \text{QI}$*

- (1) *If the generalized brach $\mathcal{L}_{g_{\alpha,\omega}}$ is non-degenerate then*
 (a) *For any $\epsilon > 0$, for any $\gamma \in [\omega]_{\text{br}}$ with $\text{ht}(\gamma) = h$ we have*

$$|P_{\gamma\alpha}| = c(\alpha, \omega, h) h + O_{\alpha,\omega,S,\epsilon}(1) h^{1-\frac{\delta_0}{6}+\epsilon}. \quad (8.7)$$

Moreover, if all the primes in S do not split in $\mathbb{Q}(\alpha)$ then the the function $O_{\alpha,\omega,S,\epsilon}(1)$ in (8.7) is in fact $O_{\iota(\alpha),S,\epsilon}(1)$.

- (b) *The function c attains only finitely many values along the ray corresponding to ω ; that is, $|\{c(\alpha, \omega, h) : h \in \mathcal{O}_S^\times \cap \mathbb{N}\}| < \infty$.*
 (c) *If $h_n \in \mathcal{O}_S^\times \cap \mathbb{N}$ satisfies $h_n | h_{n+1}$, then $c(\alpha, \omega, h_n)$ stabilizes.*
 (2) $\sup \{c(\alpha, \omega, h) : \omega \in K_S, h \in \mathcal{O}_S^\times \cap \mathbb{N}\} \ll_{\iota(\alpha),S} 1$.
 (3) *All the primes in S do not split in the quadratic extension $\mathbb{Q}(\alpha)$ if and only if $\inf \{c(\alpha, \omega, h) : \omega \in K_S, h \in \mathcal{O}_S^\times \cap \mathbb{N}\} > 0$.*

Proof of Theorem 8.10. Fix $\alpha \in \text{QI}$ and let $x_\alpha = \Gamma_\infty g_\alpha$. Define $c(\alpha, \omega, h) \stackrel{\text{def}}{=} \frac{1}{j_\alpha c_0} c_{\mathcal{L}_{g_{\alpha,\omega}}}(h)$, where $c_{\mathcal{L}_{g_{\alpha,\omega}}}(\cdot)$ is defined in Lemma 5.1 by the equation $t_{x_\alpha,\omega,h} = c_{\mathcal{L}_{g_{\alpha,\omega}}}(h) h$ and c_0, j_α are as in Theorem 8.8.

Parts (1b),(1c),(2),(3) of the Theorem follow directly from Lemma 5.1. We now prove part (1a) in a similar manner to the argument for Theorem 8.9(1) given above. For any $\epsilon > 0$, $h \in \mathcal{O}_S^\times \cap \mathbb{N}$ we have by Theorem 4.8 that the measure $\mu_{x_\alpha,\omega,h}$ satisfies the estimate (8.5). For any $\gamma \in [\omega]_{\text{br}}$ with $\text{ht}(\gamma) = h$ we have by Lemma 8.6 that the measure $\mu_{\gamma\alpha}$ is equal either to $\mu_{x_\alpha,\omega,h}$ or to $\theta_*\mu_{x_\alpha,\omega,h}$. In any case, as θ is an isometry of X_∞ , the measure $\mu_{\gamma\alpha}$ satisfies (8.5) as well. Applying Theorem 8.8 with $T^{-1} = C h^{-\frac{\delta_0}{2}+\epsilon}$, where C

is the implicit constant in (8.5), we obtain

$$\left| \frac{|P_{\gamma\alpha}|}{t_{\gamma\alpha}} - \frac{1}{j_{\alpha}c_0} \right| \ll_{\alpha,\omega,S,\epsilon} h^{-\frac{\delta_0}{6}+\epsilon}. \quad (8.8)$$

Note though that we may apply Theorem 8.8 only when $T > T_0$ and so we choose the implicit constant in (8.8) to be big enough to handle the finitely many h 's for which $T \leq T_0$.

By Lemma 8.6 we have that $t_{x_{\alpha,\omega,h}} = t_{\gamma\alpha}$ and so by the definition of $c(\alpha, \omega, h)$ we obtain that $t_{\gamma\alpha} = j_{\alpha}c_0c(\alpha, \omega, h)h$. Substituting this in (8.8) and recalling that by part (2) of the Theorem – that was already established above – $c(\alpha, \omega, h) \ll_{\alpha,S} 1$, we obtain $|P_{\gamma\alpha}| = c(\alpha, \omega, h)h + O_{\alpha,\omega,S,\epsilon}(1)h^{1-\frac{\delta_0}{6}+\epsilon}$ as desired. The last statement regarding the big O in the non-split case follows from the fact that in this case the implicit constant C from (8.5) may be chosen independent of ω . \square

Proof of Theorem 2.12. For any $\alpha \in \text{QI}$ and $\gamma \in \Gamma_S$, let $h = \text{ht}(\gamma)$ and choose $\omega \in K_S$ so that $\gamma \in [\omega]_{\text{br}}$. Let $\tilde{c}(\alpha, \gamma) \stackrel{\text{def}}{=} c(\alpha, \omega, h)$ where $c(\alpha, \omega, h)$ is the function appearing in Theorem 8.10. Note that although the choice of ω is not unique, the value $\tilde{c}(\alpha, \gamma)$ is well defined. We prove that $\tilde{c}(\cdot, \cdot)$ satisfies the conclusions of the theorem where the change of notation is in order to avoid confusion.

Parts (2),(3) of the theorem follow directly from the corresponding part of Theorem 8.10. Part (1a) of the theorem follows from the corresponding part of Theorem 8.10 with the additional remark that for any $q \in \mathcal{O}_S^\times$, $\gamma_q \in [\omega]_{\text{br}}$ for some $\omega \in \Omega$, where Ω is as in (8.6). The finiteness of Ω implies that the big O in (2.6) is independent of the generalized ray (as opposed to the big O in (8.7)). For the same reason, part (1b) of the theorem follows from the corresponding part of Theorem 8.10. Finally, part (1c) of the theorem also follows from the corresponding part of Theorem 8.10 where here we need to remark that if $q_n = \ell_n^{(1)}/\ell_n^{(2)}$ is a sequence as in part (1c) of the theorem, then $\gamma_{q_n} \in [\omega]_{\text{br}}$ for a fixed choice of $\omega \in \Omega$. \square

9. PROOF OF THEOREM 8.8

The proof of Theorem 8.8 utilizes and expands on the tight connection between the geodesic flow and the Gauss map. This connection was discovered by Artin [Art82], who used the flexibility of continued fractions to construct dense geodesics. As we will need to use technical aspects of this connection, we choose to give below a brief – essentially self contained – treatment which allows us to introduce the language and notation needed in the proof of Theorem 8.8. We follow closely the notation and exposition of [EW11, §9.6] (see also [Ser85]).

Our notation henceforth will differ slightly from the notation used in previous sections. We elaborate about these changes: Note that the natural map $\text{PSL}_2 \rightarrow \text{PGL}_2$ induces an isomorphism between the quotients $\text{PSL}_2(\mathbb{Z}) \backslash \text{PSL}_2(\mathbb{R})$ and $\text{PGL}_2(\mathbb{Z}) \backslash \text{PGL}_2(\mathbb{R})$. Due to the geometric nature of the arguments below, it would be easier for us to work with

the space $X \stackrel{\text{def}}{=} \Gamma \backslash G$, where $G \stackrel{\text{def}}{=} \text{PSL}_2(\mathbb{R})$ and $\Gamma \stackrel{\text{def}}{=} \text{PSL}_2(\mathbb{Z})$ rather than with the quotient of $\text{PGL}_2(\mathbb{R})$. As before, we shall abuse notation and treat elements of G as matrices rather than equivalence classes of such. The group G acts on the upper half plane $\mathbb{H} \stackrel{\text{def}}{=} \{z = x + iy : y > 0\}$ by Möbius transformations and this action preserves the hyperbolic metric $ds^2 = \frac{dx^2 + dy^2}{y^2}$ and so induces an action of G on the unit tangent bundle $T^1\mathbb{H}$. The action of G on $T^1\mathbb{H}$ is free and transitive hence allows us to identify G with $T^1\mathbb{H}$ once we choose a base point. We make the usual choice of the base point to be the tangent vector pointing upwards through $i \in \mathbb{H}$. With this identification the geodesic flow on $G = T^1\mathbb{H}$ corresponds to the action from the right of the positive diagonal subgroup

$$A = \{a(t)\} = \{\text{diag}(e^{t/2}, e^{-t/2}) : t \in \mathbb{R}\} < G.$$

Remark 9.1. The reason we chose to work with PGL_2 rather than PSL_2 to begin with is as follows: We were trying to analyze the c.f.e of numbers of the form $q\alpha$ and therefore used the fundamental conjugation relation $\begin{pmatrix} q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & q\alpha \\ 0 & 1 \end{pmatrix}$. Working with PSL_2 would have forced us to use conjugation by say $\text{diag}(q, q^{-1})$ which would have produced results regarding $q^2\alpha$.

9.1. Cross-sections. We now wish to introduce the notion of a cross-section. We are being rather restrictive below as we only want to discuss a specific example hence we see no use in greater generality. Given a Borel measurable set $C \subset X$, we let $r_C : C \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$ be defined by $r_C(x) = \inf\{t > 0 : xa(t) \in C\}$. The function r_C is called *the return time function to C*. The set C is called a *cross-section* for $a(t)$ if the return time functions for positive and negative times are bounded from below by some fixed positive number and the map $(x, t) \mapsto xa(t)$ from $\{(x, t) : x \in C, 0 \leq t < r_C(x)\} \rightarrow X$ is a measurable isomorphism onto its image in X . The *first return map* T_C is defined to be $T_C(x) = xa(r_C(x))$, where this makes sense; i.e. for x belonging to $\{x \in C : r_C(x) < \infty\}$. In fact, we will be interested only in points which return infinitely often in the future and past to C , thus we define the domain of the first return map to be

$$\text{Dom}_{T_C} = \{x \in C : \text{there are infinitely many} \tag{9.1} \\ \text{positive and negative } t\text{'s with } xa(t) \in C\}.$$

Note that $T_C : \text{Dom}_{T_C} \rightarrow \text{Dom}_{T_C}$ is invertible.

We now wish to define the relevant cross-section for the geodesic flow in X . An element $g \in G$ represented by a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ corresponds to a tangent vector of unit length to the upper half plane. It then defines a geodesic in \mathbb{H} which hits the boundary of \mathbb{H} in two points. We denote the *endpoint* and *startpoint* of the geodesic it defines by $e_+(g), e_-(g)$ respectively. Clearly we have $e_+(g) = \frac{a}{c}, e_-(g) = \frac{b}{d}$, where we allow ∞ as a possible value. Any element $g \in G$ has a unique decomposition (the Iwasawa decomposition) of the form

$$g = n(t)a(s)k_\theta = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{s/2} & 0 \\ 0 & e^{-s/2} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \tag{9.2}$$

where $t, s \in \mathbb{R}$, and $\theta \in [0, \pi)$. The notation $n(t), a(s), k_\theta$ should be understood from (9.2). An element g having the above decomposition corresponds to the tangent vector to the

point $t + ie^s \in \mathbb{H}$ of angle 2θ in the *clockwise* direction from the vector pointing upwards. Consider the following sets:

$$\begin{aligned}\mathcal{C}^+ &= \{g = a(s)k_\theta \in G : e_+(g) \in (0, 1), e_-(g) < -1\}; \\ \mathcal{C}^- &= \{g = a(s)k_\theta \in G : e_+(g) \in (-1, 0), e_-(g) > 1\}; \\ \mathcal{C} &= \mathcal{C}^+ \cup \mathcal{C}^-.\end{aligned}\tag{9.3}$$

The set \mathcal{C} consists of those tangent vectors whose base-point lies on the imaginary axis with some restriction on the angle θ related to the height e^s of the base point. It should be clear from the geometric picture described above that the range of ‘allowed angles’ for such a tangent vector, say in \mathcal{C}^+ , is a subinterval of $(\frac{\pi}{4}, \frac{\pi}{2})$ with $\frac{\pi}{2}$ being its right-end-point. In §10 we will work out these intervals exactly. Let $\pi : G \rightarrow X$ be the quotient map. We denote the sets $\pi(\mathcal{C}), \pi(\mathcal{C}^+), \pi(\mathcal{C}^-)$ by C, C^+, C^- respectively. The following lemma is proved in [EW11, §9.6].

Lemma 9.2. *The following hold*

- (1) *The set \mathcal{C} injects into X under π ; that is, for each $x \in C$ corresponds a unique $g \in \mathcal{C}$ with $\pi(g) = x$.*
- (2) *The set C is a cross-section for the geodesic flow on X .*
- (3) *The domain of T_C corresponds to those $g \in \mathcal{C}$ for which both $e_+(g), e_-(g)$ are irrational.*
- (4) *For $g \in \mathcal{C}^+$, if $T_C(\pi(g))$ is defined, then $T_C(\pi(g)) \in C^-$. An analogue statement with $+$ replaced by $-$ holds.*

It will be convenient for us to introduce a ‘thickening’ of the cross-section C which will be denoted by B . The following lemma is left to be verified by the reader.

Lemma 9.3. *There exists a constant $\epsilon_0 > 0$ (which will be fixed throughout) such that the following statements hold*

- (1) *The map $(g, t) \mapsto ga(t)$ from $\mathcal{C} \times (0, \epsilon_0)$ to the set*

$$\mathcal{B} \stackrel{\text{def}}{=} \{ga(t) : g \in \mathcal{C}, t \in (0, \epsilon_0)\}\tag{9.4}$$

is one to one and onto, and the set \mathcal{B} is open in G .

- (2) *Let $B \stackrel{\text{def}}{=} \pi(\mathcal{B})$. The restriction $\pi : \mathcal{B} \rightarrow B$ is one to one and onto and the set $B \subset X$ is open.*

The constant ϵ_0 introduced in the above lemma is a lower bound for the return time function, r_C , to the cross-section C . The importance of part (2) of the above lemma is that it gives us a well defined way of lifting points in X near the cross-section to the group G in which it is more convenient to work. The combination of parts (1) and (2) gives us natural coordinates on B ; any point $x \in B$ can be written uniquely as $x_C a(t)$ where $x_C \in C$ and $t \in (0, \epsilon_0)$.

In our discussion we will encounter certain measures on the cross-section C which are invariant under the first return map and we will need a procedure to construct from them

measures on the ambient space X which are invariant under the geodesic flow; that is, under the action of the group A .

Definition 9.4. Let $\tilde{\mu}$ be a probability measure on C . We define the *suspension* of $\tilde{\mu}$ to be the measure $\sigma_{\tilde{\mu}}$ on X which is given by the following rule of integration: For $f \in C_c(X)$

$$\int_X f(x) d\sigma_{\tilde{\mu}}(x) = \int_C \int_0^{r_C(x)} f(xa(t)) dt d\tilde{\mu}(x). \quad (9.5)$$

Lemma 9.5. *If $\tilde{\mu}(\text{Dom}_{T_C}) = 1$ and $\tilde{\mu}$ is T_C -invariant, then the suspension $\sigma_{\tilde{\mu}}$ is A -invariant. Furthermore, $\sigma_{\tilde{\mu}}(X) = \int_C r_C d\tilde{\mu}$.*

Proof. This follows from [EW11, Lemma 9.23] taking into account that T_C is invertible on Dom_{T_C} . \square

Definition 9.6. Given a function $f : C \rightarrow \mathbb{C}$, we denote by $\hat{f} : X \rightarrow \mathbb{C}$ the following function

$$\hat{f}(x) = \begin{cases} f(x_C) & \text{if } x \in B \text{ has coordinates } (x_C, t), \\ 0 & \text{if } x \notin B \end{cases}$$

Note that with the above definition, given a measure $\tilde{\mu}$ on C and a function $f : C \rightarrow \mathbb{C}$, equation (9.5) translates to the following useful formula which will be used frequently below

$$\int_X \hat{f} d\sigma_{\tilde{\mu}} = \epsilon_0 \int_C f d\tilde{\mu}. \quad (9.6)$$

9.2. The Gauss map. Let $I = (0, 1)$ and $S : I \rightarrow I$ be the Gauss map; i.e. the map defined by the formula $S(y) = \frac{1}{y} - \lfloor \frac{1}{y} \rfloor$. Note that strictly speaking $S(y)$ is not in I for points of the form $y = \frac{1}{m}$. The reader will easily verify that $S^n(y)$ is well defined for all positive n if and only if y is irrational. This slight inconvenience will not bother us as we will only apply the Gauss map to irrational points. Let $I_{\text{irr}} = I \setminus \mathbb{Q}$. Consider the following subsets of \mathbb{R}^2 :

$$D = \left\{ (y, z) : y \in I, 0 < z < \frac{1}{1+y} \right\}, \quad D_{\text{irr}} = \{(y, z) \in D : y \in I_{\text{irr}}\}. \quad (9.7)$$

Let $\bar{S} : D \rightarrow D$ be the map given by $\bar{S}(y, z) = (S(y), y(1 - yz))$ and note similarly that strictly speaking, in order to iterate \bar{S} as many times as we wish we need to restrict to points in D_{irr} . Recall (see for example [EW11, §3.4]) that the normalized restriction of the Lebesgue measure on \mathbb{R}^2 to D , which we denote here by λ , is an \bar{S} -invariant probability measure. This is the so called *invertible¹⁰ extension of the Gauss map* as when one projects on the first coordinates, one recovers the Gauss map and the Gauss-Kuzmin measure ν_{Gauss} introduced in the introduction. That is if $p : D \rightarrow I$ denotes the projection on the first coordinate, then

$$p_* \lambda = \nu_{\text{Gauss}}. \quad (9.8)$$

¹⁰The term ‘invertible’ refers to the fact that when restricted to a subset of D , \bar{S} is indeed invertible. This subset is obtained by neglecting a certain set of Lebesgue measure zero (see [EW11, Prop. 3.15]).

9.3. Relation to the Gauss map. Consider the maps $\tau_+ : C^+ \rightarrow D, \tau_- : C^- \rightarrow D$ defined by the following formulas: For $x = \pi(g) \in C$, where $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{C}$:

$$\text{For } g \in \mathcal{C}^+, \tau_+(x) = (e_+(g), \frac{1}{e_+(g) - e_-(g)}) = (\frac{a}{c}, cd), \quad (9.9)$$

$$\text{For } g \in \mathcal{C}^-, \tau_-(x) = (-e_+(g), \frac{1}{-e_+(g) + e_-(g)}) = (-\frac{a}{c}, -cd).$$

We let $\tau : C \rightarrow D$ be the union of τ_+ and τ_- . The formulas in (9.9) can be stated geometrically as follows: For a tangent vector $g \in \mathcal{C}$ and $x = \pi(g), \tau(x) = (y, z) \in D$, where y is the absolute value of the end point of the semicircle corresponding to g and z^{-1} is the diameter of it. For any endpoint $y \in (0, 1)$ (resp. $y \in (-1, 0)$) and any diameter $z^{-1} > 1$, we can attach a well defined semicircle in \mathbb{H} which corresponds to a unique point in \mathcal{C}^+ (resp. \mathcal{C}^-). This shows that τ is one to one and onto (and in fact, a homeomorphism) from C^+ (resp. C^-) to D which is the area below the graph of the function $y \mapsto (1+y)^{-1}$. The following basic lemma is proved in [EW11, §9.6]. It establishes the link between the geodesic flow and the Gauss map.

Lemma 9.7. *The following diagram commutes (for points $x \in C$ for which $T_C(x)$ is defined)*

$$\begin{array}{ccc} C & \xrightarrow{T_C} & C \\ \tau \downarrow & & \downarrow \tau \\ D & \xrightarrow{\bar{S}} & D. \end{array}$$

Note that $\tau : C \rightarrow D$ is ‘almost’ an isomorphism (it is two to one), and so the above lemma basically says that any dynamical question about the system $\bar{S} : D \rightarrow D$ can be pulled to a corresponding question on $T_C : C \rightarrow C$. In our case the dynamical question is that of equidistribution of certain \bar{S} -invariant measures. Using the suspension construction we will see that the equidistribution questions for the dynamical system $T_C : C \rightarrow C$ translate to equidistribution questions of certain A -invariant measures on X .

We will be interested in two types of measures on the cross-section C defined above. The first is the following version of the Lebesgue measure: We use τ_+ (resp. τ_-) to pull the (normalized restriction of) Lebesgue measure λ from D to C^+ (resp. C^-) and denote the resulting measure by $\tilde{\lambda}^+$ (resp. $\tilde{\lambda}^-$). Further denote $\tilde{\lambda} = \frac{1}{2}\tilde{\lambda}^+ + \frac{1}{2}\tilde{\lambda}^-$. Clearly $\tilde{\lambda}$ is T_C -invariant and $\tau_*(\tilde{\lambda}) = \lambda$.

The second type of measures on C are those coming from quadratic irrationals. We recall Definitions 8.1, 8.3. Let $\alpha \in \text{QI}$ and let g_α be as in (8.1). We chose to define g_α as we did so as to ensure that its determinant is positive and hence it corresponds naturally to an element of G with endpoint α . Let $x_\alpha \in X$ be the corresponding point (that is $x_\alpha = \pi(\frac{1}{\sqrt{\det g_\alpha}}g_\alpha)$) and μ_α the A -invariant probability measure supported on the periodic orbit $x_\alpha A = \{x_\alpha a(t) : t \in [0, t_\alpha)\}$, where t_α is the length of the orbit. We claim that the intersection $C \cap x_\alpha A$ is a non-empty finite set contained in Dom_{T_C} . In fact, any geodesic in the upper half plane that corresponds to a semi-circle, projects to a set in X

that intersects C non-trivially. By Lemma 9.2(3), if the end points of the geodesic are irrational, the intersection is in Dom_{T_C} . Finally, the finiteness follows from the fact that C is a cross-section together with the fact that the orbit $x_\alpha A$ is of finite length.

Let us denote by $\tilde{\mu}_\alpha$ the normalized counting measure on $C \cap x_\alpha A$. Clearly $\tilde{\mu}_\alpha$ is invariant under the first return map T_C . Let us denote the G -invariant probability measure on X by m_X . The following lemma links between the measures m_X, μ_α and the suspensions $\sigma_{\tilde{\lambda}}, \sigma_{\tilde{\mu}_\alpha}$ given in Definition 9.4.

Lemma 9.8. *Let $\alpha \in \text{QI}$. The suspensions $\sigma_{\tilde{\lambda}}, \sigma_{\tilde{\mu}_\alpha}$ of the probability measures $\tilde{\lambda}, \tilde{\mu}_\alpha$ are proportional to m_X, μ_α respectively.*

Proof. The fact that $\sigma_{\tilde{\lambda}}$ is proportional to the Haar measure m_X is proved in [EW11, p. 325-326]. The outline of the proof is as follows: By Lemma 9.5, $\sigma_{\tilde{\lambda}}$ is A -invariant. One shows that it is absolutely continuous with respect to m_X and deduces the result from the ergodicity of m_X with respect to the A -action. Regarding $\sigma_{\tilde{\mu}_\alpha}$, note that it is clearly a measure that is supported on the orbit $x_\alpha A$ and it is A -invariant by Lemma 9.5. The assertion now follows from the uniqueness (up to proportionality) of an A -invariant measure on the periodic orbit $x_\alpha A$. \square

Definition 9.9. Let c_0 be the absolute constant satisfying $\sigma_{\tilde{\lambda}} = c_0 m_X$. Similarly, for any $\alpha \in \text{QI}$ let c_α be the constant satisfying $\sigma_{\tilde{\mu}_\alpha} = c_\alpha \mu_\alpha$.

The following lemma is the last bit of information we need in order to translate the statement of Theorem 8.8 to the cross-section.

Lemma 9.10. *Let $p : D \rightarrow I$ be the projection on the first coordinate. Then*

$$\begin{aligned} (p \circ \tau)_*(\tilde{\lambda}) &= \nu_{\text{Gauss}}, \\ (p \circ \tau)_*(\tilde{\mu}_\alpha) &= \nu_\alpha. \end{aligned} \tag{9.10}$$

Proof. The first equality in (9.10) follows from the fact that $\tau_*(\tilde{\lambda}) = \lambda$ (which is basically the definition of $\tilde{\lambda}$) and the observation $p_*(\lambda) = \nu$ which was pointed out in (9.8). We argue the second equality: By Lemma 9.7 the measure $\tau_*(\tilde{\mu})$ is \tilde{S} -invariant. By the above discussion it is finitely supported. Since $x_\alpha A$ is a loop, the first return map T_C acts transitively on the support of $\tilde{\mu}$ and so the support of $\tau_*(\tilde{\mu})$ consists of a single \tilde{S} orbit. This implies that $(p \circ \tau)_*(\tilde{\mu})$ is supported on a single periodic orbit of the Gauss map S . Denote this period by P'_α . We need to argue why $P_\alpha = P'_\alpha$, which is equivalent to $P_\alpha \cap P'_\alpha \neq \emptyset$.

Consider the matrix g_α defined in (8.1). The tangent vector corresponding to g_α defines a geodesic in $T^1\mathbb{H}$ which is a semicircle with endpoint $e_+(g_\alpha) = \alpha$. At some point along this geodesic we find a point g which projects to C^+ under π . Let $x = \pi(g) \in C^+$ and $g' \in C^+$ the corresponding point in C^+ . Clearly x is in the support of $\tilde{\mu}_\alpha$ and hence the endpoint $e_+(g') = p \circ \tau(x)$ is a point of P'_α . As $\pi(g) = x = \pi(g')$ we deduce that there exists $\gamma \in \Gamma$ such that $\gamma g = g'$. Therefore the semicircle corresponding to g and to g' are related by the action of γ as a Möbius transformation. It follows that the endpoints $\alpha, e_+(g')$ are related by the action of γ as well. By Theorem 2.2 this action can effect only

finitely many digits of the c.f.e of α and we conclude that the periods of the c.f.e of α and of $e_+(g')$ must be the same (up to a possible cyclic rotation) which finishes the proof. \square

Finally, in light of (9.10), Theorem 8.8 will follow if we prove the following

Theorem 9.11. *Let $\alpha \in \text{QI}$. There exists an absolute constant $T_0 > 1$ so that if we assume that for some $T > T_0$ the estimate $|\int f d\mu_\alpha - \int f dm_X| \leq \max\{\|f\|_2, \kappa\} T^{-1}$ holds for any $f \in \text{Lip}_\kappa(X) \cap L^2(X, m_X)$, then the following two statements hold*

(1) For any $f \in \text{Lip}_\kappa(D)$, and any $\epsilon > 0$

$$\left| \int_C f \circ \tau d\tilde{\mu}_\alpha - \int_C f \circ \tau d\tilde{\lambda} \right| \ll_\epsilon \max\{\|f\|_\infty, \kappa\} T^{-\frac{1}{3} + \epsilon}. \quad (9.11)$$

(2) The constant c_0 from Definition 9.9 satisfies $\left| \frac{|P_\alpha|}{t_\alpha} - \frac{1}{j_\alpha c_0} \right| \ll_\epsilon T^{-\frac{1}{3} + \epsilon}$.

The argument yielding Theorem 9.11 is slightly technical because of the following issue: We start with a κ -Lipschitz function $f : D \rightarrow \mathbb{C}$ and construct from it the function $\widehat{f \circ \tau} : X \rightarrow \mathbb{C}$ as in Definition 9.6. As we wish to appeal to Theorem 4.8 we need to remedy $\widehat{f \circ \tau}$ to be Lipschitz in a way that will allow us to control its Lipschitz constant. In order to achieve this we shall need the following technical lemma which is proved in §10.

Lemma 9.12. *For any $M > 1$ and $0 < \rho < 1$ there exist a function $\varphi = \varphi_{\rho, M} : X \rightarrow [0, 1]$ with the following properties*

- (1) The function φ is ρ^{-1} -Lipschitz.
- (2) We have $\int_X 1 - \varphi dm_X \ll M^{-1} + \rho \log M$.
- (3) Given $f : D \rightarrow \mathbb{C}$ a κ -Lipschitz function, the product $\widehat{f \circ \tau} \cdot \varphi : X \rightarrow \mathbb{C}$ is Lipschitz with Lipschitz constant $\ll \max\{\|f\|_\infty, \kappa\} \rho^{-1} M$.

Proof of Theorem 9.11. (1). Let $T > 1$ and $\epsilon > 0$ be fixed. Under the assumption in the statement of the Theorem we need to argue the validity of (9.11). Let $f \in \text{Lip}_\kappa(D)$ be given. Let c_0, c_α be as in Definition 9.9. Using (9.6) we have the following estimate:

$$\begin{aligned} \left| \int_C f \circ \tau d\tilde{\mu}_\alpha - \int_C f \circ \tau d\tilde{\lambda} \right| &= \left| \frac{c_\alpha}{\epsilon_0} \int_X \widehat{f \circ \tau} d\mu_\alpha - \frac{c_0}{\epsilon_0} \int_X \widehat{f \circ \tau} dm_X \right| \\ &\leq \underbrace{|c_\alpha - c_0|}_{(*)} \|f\|_\infty \epsilon_0^{-1} + c_0 \epsilon_0^{-1} \underbrace{\left| \int_X \widehat{f \circ \tau} d\mu_\alpha - \int_X \widehat{f \circ \tau} dm_X \right|}_{(**)}. \end{aligned} \quad (9.12)$$

We first estimate the expression (**) in (9.12). Given $M > 1, 0 < \rho < 1$ we let $\varphi = \varphi_{\rho, M}$ be as in Lemma 9.12 and denote $\psi = 1 - \varphi$.

$$\begin{aligned} (**) &= \left| \int_X \widehat{f \circ \tau} \cdot (\varphi + \psi) d\mu_\alpha - \int_X \widehat{f \circ \tau} \cdot (\varphi + \psi) dm_X \right| \\ &\leq \left| \int_X \widehat{f \circ \tau} \cdot \varphi d\mu_\alpha - \int_X \widehat{f \circ \tau} \cdot \varphi dm_X \right| + \left| \int_X \widehat{f \circ \tau} \cdot \psi d\mu_\alpha \right| + \left| \int_X \widehat{f \circ \tau} \cdot \psi dm_X \right|. \end{aligned} \quad (9.13)$$

We will estimate each of the three summands in the right hand side of the inequality (9.13). By Lemma 9.12(2) we have

$$\left| \int_X \widehat{f \circ \tau} \cdot \psi dm_X \right| \leq \|f\|_\infty \int_X \psi dm_X \ll \|f\|_\infty (M^{-1} + \rho \log M). \quad (9.14)$$

Next, note that by Lemma 9.12(1) ψ is ρ^{-1} -Lipschitz and so our assumption together with the estimate (9.14) yields

$$\begin{aligned} \left| \int_X \widehat{f \circ \tau} \cdot \psi d\mu_\alpha \right| &\leq \|f\|_\infty \int_X \psi d\mu_\alpha \\ &\leq \|f\|_\infty \left(\int_X \psi dm_X + \max\{1, \rho^{-1}\} T^{-1} \right) \\ &\ll \|f\|_\infty (M^{-1} + \rho \log M + \rho^{-1} T^{-1}). \end{aligned} \quad (9.15)$$

Finally, by Lemma 9.12(3) our assumption applies to the Lipschitz function $\widehat{f \circ \tau} \cdot \varphi$ and we conclude the following

$$\left| \int_X \widehat{f \circ \tau} \cdot \varphi d\mu_\alpha - \int_X \widehat{f \circ \tau} \cdot \varphi dm_X \right| \leq \max\{\|f\|_\infty, \kappa\} \rho^{-1} M T^{-1}. \quad (9.16)$$

We now make the choice $M = \rho^{-1} = T^{\frac{1}{3} - \frac{\epsilon}{2}}$ and combine estimates (9.14), (9.15), (9.16) into (9.13) to obtain

$$(**) \ll_\epsilon \max\{\|f\|_\infty, \kappa\} T^{-\frac{1}{3} + \epsilon}, \quad (9.17)$$

where in the above estimate we used $\rho \log M \ll_\epsilon T^{-\frac{1}{3} + \epsilon}$.

In order to finish we need to further estimate (*) in (9.12). To obtain this estimation from the above we take $f : D \rightarrow \mathbb{C}$ to be identically 1 and note that in this case $\widehat{f \circ \tau} = \chi_B$ and so using (9.6) we have

$$\left| \int_X \widehat{f \circ \tau} d\mu_\alpha - \int_X \widehat{f \circ \tau} dm_X \right| = |\mu_\alpha(B) - m_X(B)| = \left| \frac{\epsilon_0}{c_\alpha} - \frac{\epsilon_0}{c_0} \right|. \quad (9.18)$$

The left hand side of (9.18) is (**) for this choice of f and so by (9.17) we obtain

$$|c_\alpha^{-1} - c_0^{-1}| \ll_\epsilon T^{-\frac{1}{3} + \epsilon}. \quad (9.19)$$

We choose the absolute constant T_0 so that the inequality (9.19) (applied say with $T = T_0$ and $\epsilon = \frac{1}{6}$) implies that $c_\alpha^{-1} > c_0^{-1}/2$ and so is bounded away from 0 by an absolute constant. As the derivative of the function $x \mapsto x^{-1}$ is bounded for x 's bounded away from 0, we conclude from (9.19) that

$$(*) = |c_\alpha - c_0| \ll_\epsilon T^{-\frac{1}{3} + \epsilon}. \quad (9.20)$$

Plugging this estimation of (*) together with (9.17) to (9.12) we obtain the desired inequality (9.11).

(2). Let \tilde{P}_α denote the support of $\tilde{\mu}_\alpha$. It follows from (9.10) that $p \circ \tau(\tilde{P}_\alpha) = P_\alpha$. We will show below in Lemma 9.13 that the map $p \circ \tau : \tilde{P}_\alpha \rightarrow P_\alpha$ is j_α to 1 (that is, two to

one if $|P_\alpha|$ is odd or one to one if it is even), and so the inequality sought will follow once we show $\left| \frac{|\tilde{P}_\alpha|}{t_\alpha} - \frac{1}{c_0} \right| \ll_\epsilon T^{-\frac{1}{3}+\epsilon}$.

Recall that by (9.18) $\mu_\alpha(B) = \frac{\epsilon_0}{c_\alpha}$. On the other hand, the geodesic $x_\alpha A$ which is of length t_α penetrates B exactly $|\tilde{P}_\alpha|$ times and stays in B along a time interval of length ϵ_0 each time and so $\mu_\alpha(B) = \frac{|\tilde{P}_\alpha| \cdot \epsilon_0}{t_\alpha}$. It follows that $c_\alpha^{-1} = \frac{|\tilde{P}_\alpha|}{t_\alpha}$ and we conclude from (9.19) the desired inequality $\left| \frac{|\tilde{P}_\alpha|}{t_\alpha} - \frac{1}{c_0} \right| \ll_\epsilon T^{-\frac{1}{3}+\epsilon}$. \square

Lemma 9.13. *For any $\alpha \in \text{QI}$ the map $p \circ \tau : \tilde{P}_\alpha \rightarrow P_\alpha$ is j_α to 1.*

Proof. In what follows we do not always distinguish between the cross-section C and the subset $\mathcal{C} \subset G$ used to define it. We first observe that if a semicircle in the upper half plane that corresponds to the geodesic $\{ga(t)\}$ projects under π to a periodic geodesic then $e_-(g), e_+(g) \in \mathbb{R}$ are quadratic irrationals that are Galois conjugates of each other. This implies in particular, that if we denote the lift of \tilde{P}_α from C to \mathcal{C} by $\tilde{\mathcal{P}}_\alpha$ and by $\tilde{\mathcal{P}}^\pm = \mathcal{C}^\pm \cap \tilde{\mathcal{P}}_\alpha$, then $p \circ \tau$ is injective when viewed as a map from either $\tilde{\mathcal{P}}^+$ or from $\tilde{\mathcal{P}}^-$. This follows from the fact that $\tau : \mathcal{C}^\pm \rightarrow D$ is one to one and onto and that according to the observation made above, the first coordinate $p \circ \tau(g)$ determines the second one as it is the reciprocal of the diameter of the corresponding semicircle.

This shows that the pre-image of a point in P_α is of size 1 or 2. Choose $\beta \in P_\alpha$ and a pre-image of it $g \in \tilde{\mathcal{P}}_\alpha$. We Apply Lemma 9.7 and follow the orbits $S^i(\beta)$ $i = 0, 1, \dots$ and the orbit $T_C^i(\pi(g))$ above it. If $|P_\alpha|$ is odd, then Lemma 9.2(4) tells us that when the orbit in the unit interval closes up, the orbit in the cross-section cannot close up (as it switched from C^+ to C^- or vice versa), and therefore we see that each of $\tilde{\mathcal{P}}^\pm$ projects onto P_α and so the map is 2 to 1. Similarly, in case $|P_\alpha|$ is even, when the orbit in the unit interval closes up Lemma 9.2(4) tells us that the orbit in the cross-section must return to the the same set C^+ or C^- that $\pi(g)$ belongs to and therefore it must close up by the injectivity which was observed at the beginning. It follows that one of the sets $\tilde{\mathcal{P}}^\pm$ is empty while the other one projects onto P_α , and so the map is 1 to 1. \square

10. CONSTRUCTION OF φ - PROOF OF LEMMA 9.12

10.1. Motivation. We start with a function $f : D \rightarrow \mathbb{C}$ which is κ -Lipschitz and we consider the function $\tilde{f} : X \rightarrow \mathbb{C}$ given by $\tilde{f} = \widehat{f \circ \tau}$. The points of discontinuity of \tilde{f} are contained in ∂B . We wish to find an approximation of f which is not only continuous but for which we will have clear control on its Lipschitz constant. To achieve this, we construct an auxiliary function φ which vanishes in an ϵ -thickening of ∂B and is equal to 1 outside a 2ϵ -thickening of ∂B . This will clearly make $\tilde{f} \cdot \varphi$ continuous, but in order to control its Lipschitz constant we will have to make φ vanish ‘high in the cusp’ where the differential of τ explodes (see Lemma 10.7 below). Along the construction we need to pay attention to two more quantities which we should control: The Lipschitz constant of φ and $\int \psi$, where $\psi = 1 - \varphi$. These clearly fight one against the other; in order to make

$\int \psi$ small we wish to take ϵ (which control the above thickening) to be small which makes the Lipschitz constant of φ large.

Below, in §10.2-10.5, we discuss a somewhat eclectic collection of observations that we will use in order to carry out the arguments in §10.6 with little interruption.

10.2. General metric observations. Let (Y, d) be a metric space. For a subset $F \subset Y$ we denote

$$(F)_\epsilon = \{y \in Y : d(y, F) \leq \epsilon\};$$

that is, the set of all points of distance $\leq \epsilon$ from F . The following general construction allows us to build Lipschitz functions in abundance. The proof is left to the reader.

Lemma 10.1 (Fundamental construction). *Let (Y, d) be a metric space and $F \subset Y$ a subset. For $\epsilon > 0$ define $\varphi_{\epsilon, F} : Y \rightarrow [0, 1]$ by $\varphi_{\epsilon, F}(y) = \min\{1, \epsilon^{-1} d(y, F)\}$. Then $\varphi_{\epsilon, F}$ attains the constant values 0 on F and 1 on $Y \setminus (F)_\epsilon$. Furthermore, $\varphi_{\epsilon, F}$ is ϵ^{-1} -Lipschitz.*

We now make two remarks regarding Lipschitz constants:

Remark 10.2. Consider two functions, $f : Y \rightarrow \mathbb{C}$ and $\varphi : Y \rightarrow [0, 1]$, on a metric space (Y, d) and assume that they are κ_f, κ_φ -Lipschitz respectively with $\kappa_\varphi \geq 1$. Then, for any $x, y \in Y$ we have

$$\begin{aligned} |f \cdot \varphi(x) - f \cdot \varphi(y)| &\leq |f(x) - f(y)| \varphi(x) + |f(y)| |\varphi(x) - \varphi(y)| \\ &\leq 2 \max\{\kappa_f, \|f\|_\infty\} \kappa_\varphi d(x, y), \end{aligned}$$

that is $f \cdot \varphi$ has Lipschitz constant $\ll \max\{\kappa_f, \|f\|_\infty\} \kappa_\varphi$.

Remark 10.3. Let $f : Y \rightarrow \mathbb{C}$ be a continuous function on a metric space (Y, d) in which between any two points x, y there exists a path whose length equals $d(x, y)$. Suppose there is an open cover $\{U_i\}$ of $\text{supp}(f)$ such that for each i the restriction $f : U_i \rightarrow \mathbb{C}$ is κ -Lipschitz. Then we claim that f is κ -Lipschitz as a function on Y . To see this, take two points $x, y \in Y$ and connect them by a path γ whose length is $d(x, y)$. As f is assumed to be continuous we can turn the open cover $\{U_i\}$ of the support of f to an open cover of Y by joining in the open set $U_0 = Y \setminus \text{supp}(f)$. Clearly f is κ -Lipschitz on U_0 as well. Now let $\epsilon > 0$ be a Lebesgue number for the induced open cover of the path γ . Choose points $x = x_0, x_1 \dots x_n = y$ on γ in a monotone way (so that $d(x, y) = \sum_1^n d(x_i, x_{i-1})$) and such that the distance between x_i to x_{i-1} is less than ϵ . It follows that for each $1 \leq i \leq n$ there exists an open set from the cover U_{j_i} such that $x_{i-1}, x_i \in U_{j_i}$. As f is assumed to be κ -Lipschitz on U_{j_i} , we conclude that

$$|f(x) - f(y)| \leq \sum_1^n |f(x_i) - f(x_{i-1})| \leq \sum_1^n \kappa d(x_i, x_{i-1}) = \kappa d(x, y).$$

10.3. Coordinates. We wish to define a convenient coordinate system which will allow us to carry out the relevant computations. Recall the open subsets B, \mathcal{B} of X, G respectively that were defined in Lemma 9.3. We define similarly to (9.3)

$$\begin{aligned}\mathcal{B}^+ &= \{ga(t) : g \in \mathcal{C}^+, t \in (0, \epsilon_0)\} \\ \mathcal{B}^- &= \{ga(t) : g \in \mathcal{C}^+, t \in (0, \epsilon_0)\}.\end{aligned}\tag{10.1}$$

A point $g \in \mathcal{B}$ can be written uniquely in the form $a(s)k_\theta a(t)$ where $s \in \mathbb{R}$, $t \in (0, \epsilon_0)$ and the angle $\theta \in [0, \pi)$ has some restrictions on it, arising from the requirements about the endpoints of the semicircle corresponding to g . We shall refer to (s, θ, t) as the *coordinates* of the point $g \in \mathcal{B}$ or of the corresponding point $\pi(g) \in B$.

As the action of $a(t)$ from the right does not effect the endpoints, the restrictions on the θ -coordinate are a function of s alone. We work out these restrictions for, say, $g \in \mathcal{B}^+$: We already observed (after (9.3)) that $\theta \in (\frac{\pi}{4}, \frac{\pi}{2})$ (in order to ensure that $e_+(g) \in (0, 1)$). It is easy to see from the definition of the start and end points that for $s \in \mathbb{R}$, $a(s)k_\theta \in \mathcal{C}^+$, where $\theta \in [0, \pi)$, if and only if $e^s \cot \theta \in (0, 1)$ and $-e^s \tan \theta < -1$. This is equivalent to saying $\tan \theta \in (\min\{e^s, e^{-s}\}, \infty)$. We choose an inverse $\tan^{-1} : \mathbb{R} \rightarrow (0, \frac{\pi}{2})$ and conclude that for a given s , the range of allowed angles for points $g \in \mathcal{B}^+$ with coordinates (s, θ, t) , is an interval I_s^+ which is defined by

$$I_s^+ = (\theta_{\min}(s), \frac{\pi}{2}), \text{ where } \theta_{\min}(s) = \tan^{-1}(\min\{e^s, e^{-s}\}) > \frac{\pi}{4}.\tag{10.2}$$

Let us denote

$$\mathcal{E}^+ = \{(s, \theta, t) \in \mathbb{R}^3 : s \in \mathbb{R}, t \in (0, \epsilon_0), \theta \in I_s^+\},\tag{10.3}$$

and define similarly \mathcal{E}^- and $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$. Let $\xi : \mathbb{R}^3 \rightarrow G$ be the function

$$\xi(s, \theta, t) = a(s)k_\theta a(t).\tag{10.4}$$

Clearly, we have $\xi(\mathcal{E}) = \mathcal{B}$, $\xi(\mathcal{E}^+) = \mathcal{B}^+$, and $\xi(\mathcal{E}^-) = \mathcal{B}^-$.

Lemma 10.4. *There is an absolute constant c such that for any $\epsilon > 0$, an ϵ -ball in \mathcal{E} is mapped by ξ into a ball of radius $c\epsilon$ in \mathcal{B} .*

In the course of the proof of Lemma 10.4 we will use the following elementary observation

Lemma 10.5. *Let $h(t)$ be a one parameter subgroup of G . Then for any $g \in G$, $d_G(g, gh(t)) \leq \|\dot{h}(0)\|t$, where $\|\dot{h}(0)\|$ is the norm of the derivative of $h(t)$ at the identity.*

Proof. We link any two points $g_i = \xi(s_i, \theta_i, t_i) \in \mathcal{B}$, $i = 1, 2$ by the path which changes linearly the s -coordinate first, then the θ -coordinate, and finally the t -coordinate. Each such change corresponds to the action from the right by a one-parameter subgroup $h(t)$ as in Lemma 10.5. The change in the s -coordinate corresponds to $h(s) = a(-t_1)k_{-\theta_1}a(s)k_{\theta_1}a(t_1)$, the change in the θ -coordinate corresponds to $h(\theta) = a(-t_1)k_\theta a(t_1)$, and finally, the change in the t -coordinate corresponds to $h(t) = a(t)$. As the family of one-parameter subgroups that are involved in this process are conjugations of $a(t)$ and k_θ , where the conjugating

element is varying in a compact set, we conclude that the norm of the derivative at the identity $\dot{h}(0)$ is $\ll 1$ for some absolute implicit constant. Lemma 10.5 implies then that

$$d_G(g_1, g_2) \ll |s_1 - s_2| + |\theta_1 - \theta_2| + |t_1 - t_2|,$$

which establishes the claim. \square

10.4. Height. The map τ defined in (9.9) was considered so far as a map from the cross-section C . As we wish to use differentiation it will be more convenient to extend it to a map $\tau : B \rightarrow D$ in the following way: Given a point $x \in B$ it can be written uniquely as $x_C a(t)$ where $x_C \in C$ and $t \in (0, \epsilon_0)$. We define $\tau(x) = \tau(x_C)$; that is, we view τ as a function on B which is constant along the direction of the geodesic flow.

As will be seen shortly, the norm of the differential of $\tau : B \rightarrow D$ is not bounded and so, in order to be able to control the Lipschitz constant of the function appearing in Lemma 9.12(3) we need to force its support to be contained in a domain in which we have some control on $\|d\tau\|$.

Recall the Iwasawa decomposition (9.2). Let \mathcal{F} denote the usual fundamental domain of Γ in G , that is,

$$\begin{aligned} \mathcal{F} &= \left\{ n(t)a(s)k_\theta \in G : |t| < \frac{1}{2}, t^2 + e^{2s} > 1 \right\}, \\ \overline{\mathcal{F}} &= \left\{ n(t)a(s)k_\theta \in G : |t| \leq \frac{1}{2}, t^2 + e^{2s} \geq 1 \right\} \end{aligned} \quad (10.5)$$

We define the height function $\text{ht} : G \rightarrow \mathbb{R}$ to be $\text{ht}(g) = e^s$ if $g = n(t)a(s)k_\theta$. This is indeed the imaginary coordinate of the base-point of the tangent vector to \mathbb{H} corresponding to g . This function respects the identifications induced by Γ on the boundary of $\overline{\mathcal{F}}$ and so descends to a function (which we continue to denote $\text{ht}(\cdot)$) on X . For any $M > 1$ we let

$$\begin{aligned} \mathcal{H}_M &= \{g \in \overline{\mathcal{F}} : \text{ht}(g) \geq M\}, & \mathcal{K}_M &= \{g \in \overline{\mathcal{F}} : \text{ht}(g) < M\}; \\ H_M &= \{x \in X : \text{ht}(x) \geq M\}, & K_M &= \{x \in X : \text{ht}(x) < M\}. \end{aligned} \quad (10.6)$$

Remark 10.6. It is well known that $m_X(H_M) = m_G(\mathcal{H}_M) = M^{-1}$, which is an identity that will be needed later (need to add reference).

10.5. Estimating norms of differentials.

Lemma 10.7. *The differentials of $\tau : B \rightarrow D$ and $\text{ht} : X \rightarrow \mathbb{R}$ at a point y satisfy $\|d_y \tau\| \ll \text{ht}(y)$, $\|d_y(\text{ht})\| \ll \text{ht}(y)$.*

Proof. We calculate for example $\|d_y \tau\|$ for $y \in B^+$ (here $B^+ = \pi(\mathcal{B}^+)$). Let N, H , and W denote the respective derivatives at time $t = 0$ of the one parameter subgroups $n(s), a(t)$, and k_θ which appear in (9.2);

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Let $g \in \mathcal{B}^+$ be such that $y = \pi(g)$ and write $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ so that $\tau(y) = \begin{pmatrix} a \\ c, cd \end{pmatrix}$ as given in (9.9). The tangent space $T_y(X)$ is identified (as an inner product space) with $T_g(G)$

which is in turn identified with the Lie algebra $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$ via the map sending a matrix $V \in \mathfrak{g}$ to gV ; here we make a choice of an inner product on \mathfrak{g} which induces the left-invariant Riemannian metric on G and hence on the quotient X . Thus, we will obtain an upper bound for the norm of $d_y \tau$ if we calculate an upper bound for the norms in \mathbb{R}^2 of the vectors $d_y \tau(gV)$ for $V = N, H, W$ (where here we abuse notation and think of $d_y \tau$ as a map from $T_g(G)$ to \mathbb{R}^2).

We may think of the above 2×2 matrices as vectors in \mathbb{R}^4 (where the first row corresponds to the first two coordinates) and then we get that $d_y \tau$ is given by the matrix

$$d_y \tau = \begin{pmatrix} \frac{1}{c} & 0 & -\frac{a}{c^2} & 0 \\ 0 & 0 & d & c \end{pmatrix}.$$

A short calculation shows that

$$d_y \tau(gN) = \begin{pmatrix} 0 \\ c^2 \end{pmatrix}, \quad d_y \tau(gH) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad d_y \tau(gW) = \begin{pmatrix} c^{-2} \\ c^2 - d^2 \end{pmatrix}.$$

We conclude that $\|d_y \tau\| \ll \max\{c^2, c^{-2}, d^2\}$, where the implicit constant comes from the fact that we did not specify an inner product on \mathfrak{g} . Writing g in its (s, θ, t) -coordinates $g = a(s)k_\theta a(t)$ we calculate c, d and conclude that as $|t| \leq \epsilon_0$, $\|d_y \tau\| \ll e^{|s|}$. Remark 10.9 now gives $\|d_y \tau\| \ll \text{ht}(y)$ as desired.

We briefly describe the estimate for $d_y(\text{ht})$. Let $g \in \overline{\mathcal{F}}$ be such that $y = \pi(g)$. Assume for a start that the Iwasawa decomposition of g is given by $g = n(t)a(s)$. Then the derivative in the directions of W and N are trivial (because the actions from the right of the one parameter groups $k_\theta, u(t)$ do not change the height). The derivative in the direction of H is e^s which equals $\text{ht}(y)$. It follows that for such points $\|d_y(\text{ht})\| \ll \text{ht}(y)$. Now for the general case, let $g = n(t)a(s)k_\theta \in \overline{\mathcal{F}}$ be the Iwasawa decomposition and consider the composition $G \rightarrow G \rightarrow \mathbb{R}$ given by first acting on the right by $k_{-\theta}$ and then applying ht . As ht is invariant under the action from the right by $k_{-\theta}$, this composition equals ht . Its differential at y equals by the chain rule to the composition of the differential of right multiplication by $k_{-\theta}$ at the point y and the differential of ht at the point $y' = \pi(g')$, where $g' = n(t)a(s)$. As right multiplication by $k_{-\theta}$ is an isometry the first differential has norm 1 (here we use the fact that the left invariant Riemannian metric we chose on G is also right $\{k_\theta\}$ -invariant). We evaluated the norm of the second differential before and we conclude that the composition satisfies the desired estimate. \square

Remark 10.8. As the differential of $\text{ht} : X \rightarrow \mathbb{R}$ is $\ll M$ on K_M . It follows that it is Lipschitz there with a Lipschitz constant $\ll M$ (see Remark 10.10). We conclude that there exists some absolute constant ℓ (which is the implicit constant in the estimate $\|d_y(\text{ht})\| \ll \text{ht}(y)$), such that the following two statements hold

- (1) For any $0 < \epsilon < 1$, $(H_M)_\epsilon \subset H_{\frac{M}{\ell}}$.
- (2) For any $0 < \epsilon < 1$, $(K_M)_\epsilon \subset K_{\ell M}$.

To see (1) for example, note that if this was false, then we could find $x \in K_{\frac{M}{\ell}}$ the distance of which from H_M is ≤ 1 . We conclude that there must be a point x' such that $\text{ht}(x') = M$ and $d_X(x, x') \leq 1$. This of course contradicts the fact that ht is M -Lipschitz on K_M .

Remark 10.9. We wish to comment on the height of a point $y = \pi(g) \in B$, where $g \in \mathcal{B}$ has coordinates (s, θ, t) . By Lemma 10.5, if we let $g' \in \mathcal{C}$ be the point with coordinates $(s, \theta, 0)$, then $d_G(g, g') \ll \epsilon_0$ (here we take $h(t) = a(t)$ to ‘cancel’ the t -coordinate in at most ϵ_0 time). The height of g' is by definition $\text{ht}(g') = e^{|s|}$ (the reason for the absolute value is that g' might be in the lower fundamental domain $k_{\frac{\pi}{2}}\mathcal{F}$). We conclude from parts (1),(2) of Remark 10.8 that

$$|s| - \log \ell \leq \log(\text{ht}(g)) \leq |s| + \log \ell.$$

10.6. The argument.

Proof of Lemma 9.12. Fix $M > 1$ and $0 < \epsilon < 1$ (below ϵ replaces the number ρ in the statement of Lemma 9.12). Let $F \subset X$ be defined by

$$F = (\partial B)_\epsilon \cup H_M. \quad (10.7)$$

Define $\varphi_{\epsilon, F} : X \rightarrow [0, 1]$ as in Lemma 10.1. To ease the notation we simply denote it by φ bearing in mind the dependencies on ϵ, M . Lemma 10.1 implies the assertion in Lemma 9.12(1). Let $\psi = 1 - \varphi$. As φ attains the value 1 on $X \setminus (F)_\epsilon$ we have that $\psi \leq \chi_{(F)_\epsilon}$. Furthermore, by Remark 10.8(1) and from the definitions we see that

$$(F)_\epsilon \subset (\partial B)_{2\epsilon} \cup (H_M)_\epsilon \subset ((\partial B)_{2\epsilon} \cap K_M) \cup H_{\frac{M}{\epsilon}}.$$

It follows that

$$\int_X \psi dm_X \leq m_X(((\partial B)_{2\epsilon} \cap K_M)) + m_X(H_{\frac{M}{\epsilon}}).$$

Hence, by Remark 10.6, Lemma 9.12(2) will follow once we show that the following estimate holds for all $M > 1$

$$m_X(((\partial B)_{2\epsilon} \cap K_M)) \ll \epsilon \log M. \quad (10.8)$$

In order to establish (10.8) we argue as follows: We first want to pull the calculation to G and then to \mathbb{R}^3 . It is clear that $\pi(\partial \mathcal{B} \cap \mathcal{K}_M) = \partial B \cap K_M$ and as π can only decrease distances (that is π is 1-Lipschitz), we must have $\pi((\partial \mathcal{B})_{2\epsilon} \cap \mathcal{K}_M) \supset (\partial B)_{2\epsilon} \cap K_M$. By the definition of the measure m_X it follows that

$$m_X((\partial B)_{2\epsilon} \cap K_M) \leq m_G((\partial \mathcal{B})_{2\epsilon} \cap \mathcal{K}_M). \quad (10.9)$$

Hence, we are reduced to estimate $m_G((\partial \mathcal{B})_{2\epsilon} \cap \mathcal{K}_M)$. We will work out below the estimation for $m_G((\partial \mathcal{B}^+)_{2\epsilon} \cap \mathcal{K}_M)$ only. Let $N_\epsilon(L)$ denote the number of ϵ -balls needed to cover a set L . Clearly,

$$N_{3\epsilon}((\partial \mathcal{B}^+)_{2\epsilon} \cap \mathcal{K}_M) \leq N_\epsilon(\partial \mathcal{B} \cap \mathcal{K}_M).$$

We know that a ball of radius ϵ in G has volume $\ll \epsilon^3$ and so we deduce that

$$m_G((\partial \mathcal{B}^+)_{2\epsilon} \cap \mathcal{K}_M) \ll \epsilon^3 N_\epsilon(\partial \mathcal{B} \cap \mathcal{K}_M). \quad (10.10)$$

Consider the following four subsets of $\bar{\mathcal{E}}^+ \subset \mathbb{R}^3$ which are mapped by ξ onto the boundary $\partial\mathcal{B}$

$$\begin{aligned}\mathcal{Q}_1 &= \{(s, \theta, t) : s \in \mathbb{R}, t \in (0, \epsilon_0), \theta = \theta_{\min}(s)\}; \\ \mathcal{Q}_2 &= \left\{ (s, \theta, t) : s \in \mathbb{R}, t \in (0, \epsilon_0), \theta = \frac{\pi}{2} \right\}; \\ \mathcal{Q}_3 &= \{(s, \theta, t) : s \in \mathbb{R}, \theta \in I_s^+, t = 0\}; \\ \mathcal{Q}_4 &= \{(s, \theta, t) : s \in \mathbb{R}, \theta \in I_s^+, t = \epsilon_0\}.\end{aligned}$$

Let $\mathcal{Q} = \cup_{i=1}^4 \mathcal{Q}_i$. A point in $\mathcal{B} \cap \mathcal{K}_M$ with coordinates (s, θ, t) must satisfy $|s| \leq \log M + \log \ell$ as explained in Remark 10.9. Hence, we conclude by Lemma 10.4 that

$$N_\epsilon(\partial\mathcal{B}^+ \cap \mathcal{K}_M) \ll N_{c^{-1}\epsilon}(Q \cap \{(s, \theta, t) : |s| \leq \log M + \log \ell\}). \quad (10.11)$$

This reduces the problem to a Euclidean one: For each $1 \leq i \leq 4$ the surface

$$\mathcal{Q}_i \cap \{(s, \theta, t) : |s| \leq \log M + \log 2\}$$

is a graph of a function from a domain in \mathbb{R}^2 to \mathbb{R} . The variables vary in a range that is of bounded length in one direction and of length $2(\log M + \log \ell)$ in the other. As all these functions have derivatives which are uniformly bounded (in fact, all of them are constant apart from the function $(s, t) \mapsto \theta_{\min}(s)$ corresponding to \mathcal{Q}_1 , see (10.2)), we deduce that

$$N_{c^{-1}\epsilon}(Q \cap \{(s, \theta, t) : |s| \leq \log M + \log \ell\}) \ll \frac{\log M}{\epsilon^2}. \quad (10.12)$$

Combining (10.12), (10.11), (10.10), and (10.9) gives (10.8), which as explained above concludes the proof of Lemma 9.12(2). We turn now to the proof of Lemma 9.12(3).

Let $f : D \rightarrow \mathbb{C}$ be κ -Lipschitz and denote $\tilde{f} = \widehat{f \circ \tau}$. The support of the product $\tilde{f} \cdot \varphi$ is contained in the intersection of the supports of \tilde{f} and φ . By definition of the $\widehat{}$ operator, the support of \tilde{f} is contained in B . By definition of φ its support is contained in the intersection $\{x \in X : d_X(x, \partial B) \geq \epsilon\} \cap \bar{K}_M$. It follows that

$$\text{supp}(\tilde{f} \cdot \varphi) \subset \{x \in B : d_X(x, \partial B) \geq \epsilon\} \cap \bar{K}_M. \quad (10.13)$$

As the points of discontinuity of \tilde{f} are contained in ∂B we conclude that $\tilde{f} \cdot \varphi : X \rightarrow \mathbb{C}$ is continuous. In order to estimate its Lipschitz constant we wish to appeal to Remark 10.3. Cover the open set $\mathcal{B} \cap \mathcal{K}_{2M}$ by open balls $\mathcal{U}_i \subset \mathcal{B} \cap \mathcal{K}_{2M}$. Note that each \mathcal{U}_i is contained in either \mathcal{B}^+ or \mathcal{B}^- . Consider the open cover $\{\mathcal{U}_i\}$ of $\text{supp}(\tilde{f} \cdot \varphi)$, where $U_i = \pi(\mathcal{U}_i)$. By Remark 10.3, Lemma 9.12(3) will follow once we prove that $\tilde{f} \cdot \varphi : U_i \rightarrow \mathbb{C}$ is $\max\{\kappa, \|f\|_\infty\} \epsilon^{-1} M$ -Lipschitz. As φ is ϵ^{-1} -Lipschitz we see that by Remark 10.2 it is enough to argue that for each i , $\tilde{f} : U_i \rightarrow \mathbb{C}$ is Lipschitz with Lipschitz constant $\ll \kappa M$. As $U_i \subset K_{2M}$ we know by Lemma 10.7 that the norm of the differential of τ is $\ll M$ on U_i . It follows that the Lipschitz constant of the composition $\tilde{f} = f \circ \tau$ is $\ll \kappa M$ as desired. \square

Remark 10.10. We remark here about a slight inaccuracy in the arguments presented above and how to remedy it: Let M, N be two Riemannian manifolds and $f : U \rightarrow N$ a

smooth map from an open set $U \subset M$. Assume the differential of f has norm bounded by some constant κ on U . We used above (in two places) the conclusion that f must be κ -Lipschitz. Strictly speaking, this shows indeed that f is κ -Lipschitz, but with respect to the metric induced from the restriction of the Riemannian metric from M to U . This need not be the restricted metric on U in which we are interested. In order to remedy this, one needs to prove that the following property holds: *There exists some absolute constant c such that given any two points in $x, y \in U$ one is able to find a path connecting them inside U of length $\leq cd(x, y)$ (here d is the metric of the ambient space containing U).*

Once this property is established, the conclusion is that f has Lipschitz constant $\ll \kappa$. The above property clearly holds in any Euclidean ball. Using the fact that the exponential map from the Lie algebra to G is bi-Lipschitz when restricted to a small enough neighborhood of zero, we see that any image of a small enough Euclidean ball around zero is an open neighborhood of the identity in G which satisfies the desired property. Using left translations (which are isometries of G) we see that each point of G has a basis of neighborhoods satisfying the above properties. Regarding the argument in the very end of the proof of Lemma 9.12, we should simply define the sets \mathcal{U}_i to be such neighborhoods instead open balls. Regarding the use of this in Remark 10.8, we leave the details to the reader.

REFERENCES

- [Arn07] V. I. Arnol'd, *Continued fractions of square roots of rational numbers and their statistics*, Uspekhi Mat. Nauk **62** (2007), no. 5(377), 3–14. MR2373750 (2008m:11155)
- [Arn08] V. I. Arnol'd, *Statistics of the periods of continued fractions for quadratic irrationals*, Izv. Ross. Akad. Nauk Ser. Mat. **72** (2008), no. 1, 3–38. MR2394969 (2009e:11014)
- [Art82] E. Artin, *Collected papers*, Springer-Verlag, New York, 1982. Edited by Serge Lang and John T. Tate, Reprint of the 1965 original. MR671416 (83j:01083)
- [BL05] Y. Bugeaud and F. Luca, *On the period of the continued fraction expansion of $\sqrt{2^{2n+1} + 1}$* , Indag. Math. (N.S.) **16** (2005), no. 1, 21–35. MR2138048 (2006e:11015)
- [BO07] Y. Benoist and H. Oh, *Equidistribution of rational matrices in their conjugacy classes*, Geom. Funct. Anal. **17** (2007), no. 1, 1–32. MR2306651 (2008m:11068)
- [Coh77] J. H. E. Cohn, *The length of the period of the simple continued fraction of $d^{1/2}$* , Pacific J. Math. **71** (1977), no. 1, 21–32. MR0457335 (56 #15543)
- [CZ04] P. Corvaja and U. Zannier, *On the rational approximations to the powers of an algebraic number: solution of two problems of Mahler and Mendès France*, Acta Math. **193** (2004), no. 2, 175–191. MR2134865 (2005m:11131)
- [Dir56] P. G. L. Dirichlet, *Une propri et e des formes quadratiques a determinant positif*, J. Math. Pures Appl. (1856), 76–79.
- [EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London Ltd., London, 2011. MR2723325
- [FK10]  . Fouvry and J. Kl uners, *On the negative Pell equation*, Ann. of Math. (2) **172** (2010), no. 3, 2035–2104. MR2726105 (2011h:11122)
- [Gol02] E. P. Golubeva, *On the class numbers of indefinite binary quadratic forms of discriminant dp^2* , Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **286** (2002), no. Anal. Teor. Chisel i Teor. Funkts. 18, 40–47, 227–228. MR1937366 (2003j:11040)

- [Gri98] G. Grisel, *Length of continued fractions in principal quadratic fields*, Acta Arith. **85** (1998), no. 1, 35–49. MR1623357 (99g:11015)
- [Hic73] D. R. Hickerson, *Length of period simple continued fraction expansion of \sqrt{d}* , Pacific J. Math. **46** (1973), 429–432. MR0321881 (48 #246)
- [Kei] M. Keith, *On the continued fraction expansion of $\sqrt{2^{2n+1}}$* . Unpublished, available on <http://www.numbertheory.org/pdfs/period.pdf>.
- [Kim03] H. H. Kim, *Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2* , J. Amer. Math. Soc. **16** (2003), no. 1, 139–183 (electronic). With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak. MR1937203 (2003k:11083)
- [Lag80] J. C. Lagarias, *On the computational complexity of determining the solvability or unsolvability of the equation $X^2 - DY^2 = -1$* , Trans. Amer. Math. Soc. **260** (1980), no. 2, 485–508. MR574794 (81g:10029)
- [Ler10] E. Y. Lerner, *About statistics of periods of continued fractions of quadratic irrationalities*, Funct. Anal. Other Math. **3** (2010), no. 1, 75–83. MR2734562
- [Mar04] G. A. Margulis, *On some aspects of the theory of Anosov systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2004. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska. MR2035655 (2004m:37049)
- [McM09] C. T. McMullen, *Uniformly Diophantine numbers in a fixed real quadratic field*, Compos. Math. **145** (2009), no. 4, 827–844. MR2521246 (2010h:11114)
- [MF93] M. Mendès France, *Remarks and problems on finite and periodic continued fractions*, Enseign. Math. (2) **39** (1993), no. 3-4, 249–257. MR1252067 (94i:11045)
- [Pol86] M. Pollicott, *Distribution of closed geodesics on the modular surface and quadratic irrationals*, Bull. Soc. Math. France **114** (1986), no. 4, 431–446. MR882589 (88j:58102)
- [PR94] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., Boston, MA, 1994. Translated from the 1991 Russian original by Rachel Rowen. MR1278263 (95b:11039)
- [RS92] A. M. Rockett and P. Szűsz, *Continued fractions*, World Scientific Publishing Co. Inc., River Edge, NJ, 1992. MR1188878 (93m:11060)
- [Ser85] C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31** (1985), no. 1, 69–80. MR810563 (87c:58094)
- [Ven10] A. Venkatesh, *Sparse equidistribution problems, period bounds and subconvexity*, Ann. of Math. (2) **172** (2010), no. 2, 989–1094. MR2680486