LINNIK’S PROBLEMS AND MAXIMAL ENTROPY METHODS

ANDREAS WIESER

Abstract. We use maximal entropy methods to examine the distribution properties of primitive integer points on spheres and of CM points on the modular surface. The proofs we give are a modern and dynamical interpretation of Linnik’s original ideas and follow techniques presented by Einsiedler, Lindenstrauss, Michel and Venkatesh in 2011.

1. Introduction

1.1. Integer points on spheres. Consider the set of primitive integral solutions to the equation

\[ x^2 + y^2 + z^2 = d \]

for some positive integer \( d \). The question whether or not such a solution exists was raised by Legendre (amongst others), who claimed that the equation \( x^2 + y^2 + z^2 = d \) has a solution if and only if \( d \) is not of the form \( d = 4^a(8b + 7) \) for non-negative integers \( a, b \). A full proof of Legendre’s so-called three-squares theorem was given by Gauss [Gau86]. In fact, a primitive integral solution to \( x^2 + y^2 + z^2 = d \) exists if and only if \( d \) satisfies Legendre’s condition

\[ d \in \mathbb{D} := \{ d \in \mathbb{N} \mid d \not\equiv 0, 4, 7 \mod 8 \}. \]

A further question treated by Gauss concerns a refinement of the above: As \( d \) tends to infinity with \( d \) satisfying Legendre’s condition, how many primitive integral solutions to \( x^2 + y^2 + z^2 = d \) are there? The number of such solutions turns out to be closely related to the class number of the quadratic number field \( \mathbb{Q}(\sqrt{-d}) \) (see for instance Section 4 of [EMV13]) for which an asymptotic as \( d \to \infty \) is well-known thanks to Dirichlet’s class number formula and Siegel’s lower bound. These results imply that the number of primitive integral solutions to \( x^2 + y^2 + z^2 = d \) is \( d^{\frac{1}{2} + o(1)} \) for \( d \to \infty, d \in \mathbb{D} \). In this paper, we will be interested in the distribution properties of these solutions: Let

\[ \mathcal{I}_d := \left\{ (x, y, z) \in \mathbb{Z}^3 \mid \gcd(x, y, z) = 1, x^2 + y^2 + z^2 = d \right\} \]

be the set of primitive integral solutions to \( x^2 + y^2 + z^2 = d \) projected onto the unit sphere \( S^2 \). Using ergodic-theoretic methods, we will give a proof of the following non-effective result due to Linnik [Lin68]:

Linnik’s Theorem A (Equidistribution of primitive integer points on the sphere). Let \( p \) be an odd prime and let \( \mathbb{D}(p) = \{ d \in \mathbb{D} \mid -d \in (\mathbb{F}_p^\times)^2 \} \). As \( d \) tends to infinity with \( d \in \mathbb{D}(p) \), the normalized sums of Dirac measures \( \frac{1}{|\mathcal{I}_d|} \sum_{x \in \mathcal{I}_d} \delta_x \) equidistribute to the uniform probability measure on the unit sphere \( S^2 \).

Date: January 30, 2018.

The author was supported by SNF Grant 200021-152819.
Since the choice of the prime \( p \) is arbitrary, the splitting\(^1\) condition \(-d \in (\mathbb{F}_p^2)^2\), known as Linnik’s condition, seems to be superfluous. Indeed, Linnik was able to eliminate it assuming GRH. In 1988, Duke [Duk88] succeeded in proving Linnik’s theorem unconditionally using entirely different methods building on work of Iwaniec.

A modern exposition of Linnik’s theorem using expander graphs is given by Ellenberg, Michel and Venkatesh in [EMV13]. The present article aims to give an ergodic theoretic proof of Linnik’s theorem using maximal entropy methods and following Einsiedler, Lindenstrauss, Michel and Venkatesh [ELMV12]. The motivation for such a proof originates from a refinement of Linnik’s theorem by Akka, Einsiedler and Shapira in [AES16].

In [ELMV12], the authors prove a theorem due to Duke concerning equidistribution of collections of closed geodesics (associated to positive discriminants) on the modular surface. In analogy to [ELMV12], we will study certain collections of orbits \( \mathcal{O}_d \) in the \( p \)-adic extension \( SO_3(\mathbb{Z}_p^2)) \) which arise through the stabilizer subgroup in \( SO_3 \) of a primitive integer point of length \( \sqrt{d} \). We note that one dynamical reason for working in the \( p \)-adic extension instead of \( SO_3(\mathbb{Z}) \) is that the acting subgroup \( SO_3(\mathbb{Q}_p) \) is non-compact for all odd primes \( p \).

Furthermore, if \( v \in \mathbb{Z}^3 \) is a primitive integer point then the stabilizer subgroup \( \mathbb{H}_v = \{ g \in SO_3 \mid gv = v \} \) is the orthogonal group of the quadratic form \( x^2 + y^2 + z^2 \) restricted to the plane \( v^\perp \). Thus, \( \mathbb{H}_v(\mathbb{Q}_p) \) is split if and only if \( v^\perp \) contains an isotropic vector. One can show by elementary means that the latter is equivalent to Linnik’s condition for \( d = Q_0(v) \) (see also Proposition 2.5). Therefore, Linnik’s condition is an artefact of our dynamical proof (it is comparable to the positivity assumption on discriminants in [ELMV12]). Using the vector \( v \) the collection \( \mathcal{O}_d \) is constructed to be a finite union of certain orbits under the group \( \mathbb{H}_v(\mathbb{R} \times \mathbb{Q}_p) \).

We will see that as the integer \( d \in \mathbb{D}(p) \) increases the uniform measures on the collections of orbits \( \mathcal{O}_d \) mentioned above equidistribute to the Haar measure on the \( p \)-adic extension (Theorem 2.2). From this, Linnik’s Theorem A is readily obtained by projecting onto the real quotient. The main step in the proof of Theorem 2.2 is to show that any weak*-limit of the measures has maximal entropy with respect to the action of a fixed diagonalizable element in \( SO_3(\mathbb{Q}_p) \) (cf. Section 2.7). This maximal entropy statement can then be used to deduce Theorem 2.2 by means of a uniqueness result on measures of maximal entropy see e.g. [EL07, Theorem 7.9] or [MT94] (cf. Section 2.6). To show maximal entropy one verifies, roughly speaking, the following two claims which in combination yield the desired “chaotic behaviour”.

1. The total volume of \( \mathcal{O}_d \) grows quickly enough. In fact, the total volume is \( d^{2+o(1)} \) as one can attach to any \( \mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p) \)-orbit in \( \mathcal{O}_d \) a uniquely determined integer point of length \( \sqrt{d} \) (cf. Section 2.1) and then apply the result mentioned earlier.

2. There are not too many pairs of orbits in \( \mathcal{O}_d \) that lie close together. This is the content of Linnik’s basic lemma (Proposition 2.12), for which we shall use a bound on the number of representations of a binary quadratic form by \( x^2 + y^2 + z^2 \) (cf. Theorem 2.14).

1.2. CM points. Given a fundamental discriminant \( d \) (see Section 3.3 for the definitions), there are \( h_d = [d]^{2+o(1)} \) “inequivalent” solutions to the equation \( d = \)

\[^1\]In fact, note that \(-d \in (\mathbb{F}_p^2)^2\) if and only if \( p \) is split in the field \( \mathbb{Q}(\sqrt{-d}) \).
Linnik’s Theorem B (Equidistribution of CM points). Let $p$ be an odd prime. Let $H_d$ be the set of CM points associated to a fundamental discriminant $d < 0$. Then

$$\frac{1}{|\text{SL}_2(\mathbb{Z}), H_d|} \sum_{z \in \text{SL}_2(\mathbb{Z}), H_d} \delta_z \rightarrow m_{\text{SL}_2(\mathbb{Z})|\mathbb{H}}$$

as $d \rightarrow -\infty$ satisfying Linnik’s condition $\left(\frac{d}{p}\right) = 1$.

This theorem should again be seen as an analogue to Duke’s theorem (see Theorem 1.3 in [ELMV12]) which treats the case of positive discriminants. For positive discriminants the CM points are replaced by the geodesic connecting the two points on the real axis. For a reformulation of Linnik’s Theorem B in a fashion similar to Linnik’s Theorem A see Theorem 1.1 in [ELMV12].

In Section 3 of this paper, we prove Linnik’s Theorem B using maximal entropy methods again. Up to some complications due to non-compactness of the homogeneous space $\text{PGL}_2(\mathbb{Z}[\frac{1}{p}]) \backslash \text{PGL}_2(\mathbb{R} \times \mathbb{Q}_p)$ we consider, the proof is largely analogous to the proof of Linnik’s Theorem A. We note that it also studies collections of orbits under certain tori (given by stabilizer subgroups constructed in Section 3.1).

1.3. Some notation and facts. Throughout this paper $G$ will usually denote either of the $\mathbb{Q}$-algebraic groups $\text{SO}_3$ or $\text{PGL}_2$. Also write $A$ for the ring of adeles of $\mathbb{Q}$, $A_f$ for the ring of finite adeles of $\mathbb{Q}$ and $\hat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$. For $G = \text{SO}_3$ or $G = \text{PGL}_2$ the group $G(\mathbb{Z}[\frac{1}{p}])$ (resp. $G(\mathbb{Q})$) embedded diagonally into $G(\mathbb{R} \times \mathbb{Q}_p)$ resp. $G(\hat{\mathbb{A}})$ defines a lattice and we obtain the $p$-adic resp. adelic extensions

$$G(\mathbb{Z}[\frac{1}{p}]) \backslash G(\mathbb{R} \times \mathbb{Q}_p), \quad G(\mathbb{Q}) \backslash G(\hat{\mathbb{A}})$$

of the real quotient $G(\mathbb{Z}) \backslash G(\mathbb{R})$. The groups $\text{SO}_3$ and $\text{PGL}_2$ have class number one, that is for any set $S \subset \mathcal{V}_\mathbb{Q}$ of finite places we have

$$G(\mathbb{R} \times \mathbb{Q}_p) = G(\mathbb{R} \times \mathbb{Z}_p)G(\mathbb{Z}[\frac{1}{p}]), \quad G(\hat{\mathbb{A}}) = G(\mathbb{Q})G(\mathbb{R} \times \hat{\mathbb{Z}})$$

$^2$Proving this is elementary for $G = \text{PGL}_2$. For $G = \text{SO}_3$ it means that the genus of the quadratic form $x^2 + y^2 + z^2$ consists of one $\mathbb{Z}$-equivalence class – see Section 7.1 in [EMV13] or p.138 in [Cas78].
In particular one obtains projection maps from the adelic to the $p$-adic and from the $p$-adic to the real quotient by taking quotients with compact groups (for instance by $G(\mathbb{Z}_p)$ in the latter case).

For any dimension $d$ we equip $\mathbb{Q}_p^d$ with the norm $\|a\|_p = \max(|a_1|_p, \ldots, |a_d|_p)$. For any $A \in \text{Mat}_d(\mathbb{Q}_p)$ and $x \in \mathbb{Q}_p^d$ we have that $\|Ax\|_p \leq \|A\|_p \|x\|_p$ and that $A \in \text{GL}_d(\mathbb{Z}_p)$ if and only if $\|A\|_p = 1$. The groups $G(\mathbb{R})$, $G(\mathbb{Q}_p)$, $G(A)$ will always be equipped with a left-invariant metric $d$, which for $G(\mathbb{R})$ may be obtain from a left-invariant Riemannian metric and for $G(\mathbb{Q}_p)$ from the norm $\|\cdot\|_p$ on $\text{Mat}_d(\mathbb{Z}_p)$ (see [Rüh16]).

Acknowledgements. This is a shortened version of the author’s master thesis. We would like to thank Manfred Einsiedler for suggesting the topic and for many enthusiastic discussions as well as Menny Akka and Manuel Lüthi for commenting on preliminary versions of this paper.

Contents

1. Introduction 1
   1.1. Integer points on spheres 1
   1.2. CM points 2
   1.3. Some notation and facts 3
2. Equidistribution of integer points on spheres 4
   2.1. Stabilizer orbits and producing integer points 5
   2.2. Conjugacy of stabilizer subgroups 7
   2.3. Equidistribution on the $p$-adic extension and maximal entropy 8
   2.4. Proof of Linnik’s Theorem A 9
   2.5. Exponential map and horospherical subgroups 10
   2.6. Uniqueness of measures of maximal entropy 12
   2.7. Maximal entropy 13
3. Equidistribution of CM points 18
   3.1. Algebraic tori associated to quadratic number fields 18
   3.2. Compact torus orbits 19
   3.3. Obtaining CM points and the proof of Linnik’s Theorem B 23
   3.4. Ideal classes and heights 24
   3.5. Maximal Entropy 26
   3.6. Visiting the cusp 29
References 32

2. Equidistribution of integer points on spheres

Denote the $p$-adic and the adelic extension of $X_\infty = \text{SO}_3(\mathbb{Z})\backslash\text{SO}_3(\mathbb{R})$ by

$$X_{p,\infty} = \text{SO}_3(\mathbb{Z}_p)\backslash\text{SO}_3(\mathbb{R} \times \mathbb{Q}_p), \quad X_A = \text{SO}_3(\mathbb{Q})\backslash\text{SO}_3(A)$$

and observe that all these quotients are compact, since $X_\infty$ is compact and is obtained from $X_{p,\infty}$ or $X_A$ by taking a quotient with a compact group (given by $\text{SO}_3(\mathbb{Z}_p)$ and $\prod_{q \text{ prime}} \text{SO}_3(\mathbb{Z}_q)$ respectively). Furthermore, set $\Gamma := \text{SO}_3(\mathbb{Q})$ and $Q_0(x, y, z) = x^2 + y^2 + z^2$. 

References
2.1. Stabilizer orbits and producing integer points. Let \( v \in \mathbb{Z}^3 \) be a primitive integer point and denote by \( \mathbb{H}_v \), \( < \mathbb{S}_3 \) its stabilizer subgroup. The orbit \( \Gamma \mathbb{H}_v(\hat{A}) \) of the identity coset under \( \mathbb{H}_v(\hat{A}) \) in \( X_\mathbb{A} \) is closed\(^3\) and thus compact. We may hence write it as a finite union of orbits under the open subgroup \( \mathbb{H}_v(\hat{R} \times \hat{Z}) < \mathbb{H}_v(\hat{A}) \):

\[
\Gamma \mathbb{H}_v(\hat{A}) = \bigcup_{\rho \in \mathcal{R}_v} \Gamma \rho \mathbb{H}_v(\hat{R} \times \hat{Z})
\]

for a finite set of representatives \( \mathcal{R}_v \subset \mathbb{S}_3(\hat{R} \times \hat{Z}) \) (using again the fact that \( \mathbb{S}_3 \) has class number one – see (1.1)).

As we shall see now, each of the compact stabilizer orbits on the right hand side of Equation (2.1) produces a primitive integer point, which will allow us to retrieve from it the integer points considered in Linnik’s Theorem A. Let \( \gamma \) be an integer point of the same length as \( v \) with \( \gamma v = v \) so that \( \gamma \) is an integer point of length \( v \). Then \( \gamma v = \gamma hv = g\bar{v} \in \mathbb{Q}^3 \cap (\hat{R} \times \hat{Z})^3 = \mathbb{Z}^3 \)

is an integer point of the same length as \( v \). We say that an integer point \( w \) is produced by the adelic stabilizer orbit of \( v \), if it there exists \( h \in \mathbb{H}_v(\hat{A}) \) and \( \gamma \in \Gamma \) so that \( \gamma v = w \) and \( \gamma h \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \). Equivalently, there exists \( \gamma \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \) with \( \gamma h = g \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \) by Equation (1.1). Then

\[
\gamma v = \gamma hv = g\bar{v} \in \mathbb{Q}^3 \cap (\hat{R} \times \hat{Z})^3 = \mathbb{Z}^3
\]

Remark 2.1. Ellenberg, Michel and Venkatesh [EMV13] show that all primitive integer points of length \( \sqrt{Q_0(v)} \) are produced by the adelic stabilizer orbit of \( v \).

We will not need this result for our purposes.

In the following we will collect a few important facts about the integer points generated by the above procedure.

- Any point \( \Gamma h \in \Gamma \mathbb{H}_v(\hat{A}) \) gives rise to a unique point in \( \mathbb{S}_3(\mathbb{Z})^3 \).

  Indeed, choose \( \gamma, \gamma' \in \Gamma \) such that \( g = \gamma h, g' = \gamma' h \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \). Then

  \[
g'g^{-1} = \gamma'\gamma^{-1} \in \mathbb{S}_3(\mathbb{Q}) \cap \mathbb{S}_3(\hat{R} \times \hat{Z}) = \mathbb{S}_3(\mathbb{Z}).
\]

  This shows that \( \mathbb{S}_3(\mathbb{Z})\gamma'v = \mathbb{S}_3(\mathbb{Z})\gamma'\gamma^{-1}v = \mathbb{S}_3(\mathbb{Z})g\bar{v} \).

- Let \( h, h' \in \mathbb{H}_v(\hat{A}) \). The cosets \( \Gamma h, \Gamma h' \) produce the same integer point up to \( \mathbb{S}_3(\mathbb{Z}) \) if and only if \( \Gamma h, \Gamma h' \) lie in the same \( \mathbb{H}_v(\hat{R} \times \hat{Z}) \)-orbit.

  If for \( \Gamma h, \Gamma h' \) there are \( \gamma, \gamma' \in \Gamma \) with \( \gamma v = \gamma' v \) and \( \gamma h, \gamma'h' \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \) then the element \( h = h^{-1}\gamma^{-1}\gamma h' \in \mathbb{H}_v(\hat{R} \times \hat{Z}) \) satisfies \( \Gamma h\bar{h} = \Gamma h' \).

  For the converse assume that there exists \( h \in \mathbb{H}_v(\hat{R} \times \hat{Z}) \) so that \( \Gamma h\bar{h} = \Gamma h' \) and choose \( \gamma, \gamma' \in \Gamma \) with \( \gamma h \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \) and \( \gamma' \in \Gamma \) with \( \gamma h\bar{h} = \gamma'h' \). Then \( \gamma v = \gamma h\bar{h}v = \gamma' h'v = \gamma'v \) and \( \gamma' h' \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \).

- Every integer point \( w \) produced by the adelic stabilizer orbit of \( v \) is primitive.

  Let \( g \in \mathbb{S}_3(\hat{R} \times \hat{Z}) \) satisfy \( \gamma v = g \bar{v} \). Then, as \( v \) is primitive we have

\[
1 = \|v\|_p \leq \|g^{-1}\|_p\|g\bar{v}\|_p = \|g\bar{v}\|_p = \|w\|_p \leq \|g\bar{v}\|_p\|v\|_p = 1
\]

for any prime \( p \).

\(^3\)If \( (h_n)_n \) is a sequence in \( \mathbb{H}_v(\hat{A}) \) such that \( \gamma_n h_n \to g \) for some \( g \in \mathbb{S}_3(\hat{A}) \) and some sequence \( (\gamma_n)_n \) in \( \mathbb{S}_3(\hat{Z}) \), we must also have \( \gamma_n v = \gamma_n h_n v \to g\bar{v} \) for \( n \to \infty \). But \( \gamma_n v \in \mathbb{Q}^3 \) for all \( n \) and \( \mathbb{Q}^3 \subset A^3 \) is discrete. Thus, \( \gamma_n v = g\bar{v} \) for all large enough \( n \) and in particular \( \gamma_n^{-1}g \in \mathbb{H}_v(\hat{A}) \) and \( \Gamma g = \Gamma \gamma_n^{-1}g \in \Gamma \mathbb{H}_v(\hat{A}) \).
By the discussion at the beginning of this section, we conclude in particular that the number of integer points in $SO_3(\mathbb{Z})\mathbb{Z}^3$ produced by the adelic stabilizer orbit of $v$ is exactly $|\mathcal{R}_v|$ where $\mathcal{R}_v$ is as in Equation (2.1). The integer point produced by $\Gamma \rho h$ for $\rho \in \mathcal{R}_v$ and $h \in \mathbb{H}_v(\mathbb{R} \times \hat{\mathbb{Z}})$ is $\rho v = \rho_p v = \rho_\infty v$ for all primes $p$ up to an element in $SO_3(\mathbb{Z})$.

We now project the adelic stabilizer orbit $\Gamma \mathbb{H}_v(A)$ onto the $p$-adic extension $X_{p,\infty}$ for some odd prime $p$. Its projection is a finite union of $\mathbb{H}_v(\mathbb{R} \times \mathbb{Q}_p)$-orbits and more explicitly given by\(^4\)

\[
\bigcup_{\rho \in \mathcal{R}_v} SO_3(\mathbb{Z}[\frac{1}{p}])/(\rho_\infty, \rho_p)\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p) = \bigcup_{\rho \in \mathcal{R}_v} SO_3(\mathbb{Z}[\frac{1}{p}])\mathbb{H}_{p,v}(\mathbb{R} \times \mathbb{Z}_p)(\rho_\infty, \rho_p)
\]
as

\[(\rho_\infty, \rho_p)\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)(\rho_\infty, \rho_p)^{-1} = \mathbb{H}_{p,v}(\mathbb{R} \times \mathbb{Z}_p).
\]

In what follows, we will consider this collection of orbits for growing $D = Q_0(v)$.

For any primitive integer point $w$ let $m_{\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}$ be the normalized Haar measure on the stabilizer subgroup $\mathbb{H}_w(\mathbb{R} \times \mathbb{Z}_p)$. An orbit $x\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p) \simeq \text{Stab}_{\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}(x)\backslash \mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ is naturally equipped with a unique $\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$-invariant measure induced by $m_{\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}$, which we will denote by $m_{x\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}$ and refer to as the Haar measure on $x\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$. As $m_{\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}$ is a probability measure, the total mass of $m_{x\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}$ is $|\text{Stab}_{\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)}(x)|^{-1}$. Summing over the different orbits in (2.2) and normalizing by the total volume one obtains a probability measure $\mu_v$ on the projection of $\Gamma \mathbb{H}_v(A)$ onto the $p$-adic extension.

**Theorem 2.2 (Equidistribution of collections).** Let $p$ be an odd prime and let $(v_k)$ be sequence of primitive integer points with $Q_0(v_k) \in \mathbb{D}(p)$ for all $k$ and with $Q_0(v_k) \to \infty$ as $k \to \infty$. Then the probability measures $\mu_{v_k}$ converge in the weak\(^\ast\)-topology to the normalized Haar measure on $X_{p,\infty}$ as $k \to \infty$.

In Section 2.4 we will see that this theorem implies Linnik’s Theorem A. We now discuss the measures $\mu_v$ in greater detail, verifying in particular that $\mu_v$ is $\mathbb{H}_v(\mathbb{R} \times \mathbb{Q}_p)$-invariant. Thus, the measures $\mu_{v_k}$ in the theorem do not have a common invariance; we shall get rid of this dynamical inconvenience in Section 2.2.

**Lemma 2.3.** Let $v \in \mathbb{Z}^3$ be a primitive integer point. Then all orbits in the collection (2.2) have equal volume. Furthermore, $\mu_v$ is $\mathbb{H}_v(\mathbb{R} \times \mathbb{Q}_p)$-invariant.

**Proof.** Let $\rho \in \mathcal{R}_v$. Note that $SO_3(\mathbb{Z}[\frac{1}{p}])/(\rho_\infty, \rho_p)$ is stabilized by $g \in \mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ if and only if $SO_3(\mathbb{Q})\rho = SO_3(\mathbb{Q})h$ is stabilized by $g$. However, as $\mathbb{H}_v$ is abelian, the latter is equivalent to $g$ lying in $\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p) \cap SO_3(\mathbb{Q}) = \mathbb{H}_v(\mathbb{Z})$. This shows the first claim.

We now show that $\mu_v$ is $\mathbb{H}_v(\mathbb{R} \times \mathbb{Q}_p)$-invariant: Let $h \in \mathbb{H}_v(\mathbb{Q}_p)$. By linearity of the pushforward it suffices to prove that

$$R_h : SO_3(\mathbb{Z}[\frac{1}{p}])/(\rho_\infty, \rho_p)\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p) \to SO_3(\mathbb{Z}[\frac{1}{p}])/(\rho_\infty, \rho_p)h\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$$

\(^4\)The union is still disjoint: Suppose that there are $\rho \neq \rho'$ in $\mathcal{R}_v$, $\gamma$ in $SO_3(\mathbb{Z}[\frac{1}{p}])$ and $h$ in $\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ such that $\gamma(\rho_\infty, \rho_p)h = (\rho'_\infty, \rho'_p)$. Then $\gamma \in SO_3(\mathbb{Z})$ and $\gamma \rho_\infty v = \rho'_\infty v$ contradicting the choice of $\rho, \rho'$. 


is measure-preserving for some fixed $\rho \in \mathcal{R}_v$. However, observe that the push-forward of the Haar measure on $SO_3(\mathbb{Z}_p^1)\backslash \mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ is $\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$-invariant and has the same total mass as $SO_3(\mathbb{Z}_p^1)\backslash \mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ by the first claim. Here we have used that $SO_3(\mathbb{Z}_p^1)(\rho_\infty, \rho_p)h\mathbb{H}_v(\mathbb{R} \times \mathbb{Z}_p)$ is an orbit appearing in (2.2).

A consequence\(^5\) of Lemma 2.3 is that the total volume of the collection (2.2) is equal to the number of orbits $|\mathcal{R}_v|$ in the collection for large enough lengths. This number, which is the class number of $\mathbb{Q}$ with respect to the bilinear form associated to $Q_v$ can be described asymptotically: Viewing $v$ as a quaternion, one can relate it to the class number of $\mathbb{Q}(\sqrt{-d})$ where $d = Q_0(v)$ (Proposition 3.5 will provide more insight).

Lemma 2.4. For a primitive integer point $v \in \mathbb{Z}^3$ and $d = Q_0(v)$ the total volume of the collection (2.2) is $d^{\frac{5}{2}} + o(1)$.

This follows from Dirichlet’s class number formula and Siegel’s lower bound (see for instance [IK04, Theorem 5.28]).

2.2. Conjugacy of stabilizer subgroups. Let $p$ be an odd prime. In this subsection, we will in particular show that given two primitive integer points $v, v'$ with $Q_0(v), Q_0(v') \in \mathbb{D}(p)$ there exists some $k \in SO_3(\mathbb{Z}_p)$ so that the conjugacy relations

$$k\mathbb{H}_v(\mathbb{Z}_p)k^{-1} = \mathbb{H}_{v'}(\mathbb{Z}_p), \quad k\mathbb{H}_v(\mathbb{Q}_p)k^{-1} = \mathbb{H}_{v'}(\mathbb{Q}_p)$$

hold.

Proposition 2.5. Let $v \in \mathbb{Z}_p^3$ be primitive with $Q_0(v) =: D \in -\left(\mathbb{Z}_p^\times\right)^2$. Then there exists a basis of $\mathbb{Z}_p^3$ of the type $(v, w_1, w_2)$, in which $Q_0$ takes the form $Dx^2 + yz$. In particular, the following statements hold.

(a) For any two primitive integer points $v_1, v_2$ with $Q_0(v_1), Q_0(v_2) \in \mathbb{D}(p)$ there exists $k \in SO_3(\mathbb{Z}_p)$ with

$$k\frac{v_1}{\sqrt{-Q_0(v_1)}} = \frac{v_2}{\sqrt{-Q_0(v_2)}}.$$ 

(b) Let $v$ be a primitive integer point with $Q_0(v) \in \mathbb{D}(p)$. There exists an isomorphism $f : PGL(\mathbb{Z}_p) \to SO_3(\mathbb{Z}_p)$ such that

$$f(PGL(\mathbb{Z}_p)) = SO_3(\mathbb{Z}_p), \quad f^{-1}(\mathbb{H}_v(\mathbb{Q}_p)) = \left\{ \begin{pmatrix} x \\ 1 \end{pmatrix} : x \in \mathbb{Q}_p^\times \right\}.$$ 

Proof. Let $v \in \mathbb{Z}_p^3$ be primitive with $Q_0(v) =: D \in -\left(\mathbb{Z}_p^\times\right)^2$. As $v$ is primitive and $D$ is invertible, one can apply the Gram-Schmidt process to obtain $\mathbb{Z}_p^3 = \mathbb{Z}_p^\perp \cap \mathbb{Z}_p^3$ with respect to the bilinear form associated to $Q_0$. Denote $W = v^\perp \cap \mathbb{Z}_p^3$ and $q = Q_0|_W$. By compactness of $W$ we can find some integral vector $w_1 \in W$ with $|q(w_1)|_p = \max_{w \in W}|q(w)|_p$. By the polarization identity $q(w_1)$ divides all products $(w, w')_q$ for $w, w' \in W$ and therefore also the discriminant $D$ of $q$. Hence $q(w_1)$ is a unit and applying Gram-Schmidt again yields that there is a basis $(v, w_1, w_2)$ of

\(^5\)In fact, an elementary linear algebra argument shows that $\mathbb{H}_v(\mathbb{Z})$ is trivial if $Q_0(v) > 3$. 
\[ Z_p^3 \] so that \( Q_0 \) is of the form \( Dx^2 + \alpha y^2 + \beta z^2 \) where \( \alpha, \beta \in Z_p^* \). Using \( D \in -(Z_p^*)^2 \) the chain of \( Z_p \)-equivalences
\[
\alpha y^2 + \beta z^2 \sim \alpha(y^2 + \alpha \beta y^2) \sim \alpha(y^2 + Dz^2) \sim \alpha(y^2 - z^2) \sim \alpha yz \sim yz
\]
concludes the first part of the proposition.

For (a) pick \( g_1, g_2 \in SL_3(Z_p) \) with \( g_1 e_1 = u_1, g_2 e_2 = u_2 \) as in the first part of the proposition applied to \( u_i = v_i \sqrt{-Q_0(v_i)}^{-1} \) for \( i \in \{1, 2\} \). Then \( g = g_2 g_1^{-1} \) is in \( SO_3(Z_p) \) and satisfies \( gu_1 = u_2 \).

For (b) one uses the fact that there is an isomorphism \( PGL_2(Q_p) \cong SO_{disc}(Q_p) \) defined over \( Z_p \) obtained by letting \( PGL_2(Q_p) \) act on \( sl_2(Q_p) \) by conjugation. \( \square \)

2.3. **Equidistribution on the \( p \)-adic extension and maximal entropy.** Using Proposition 2.5 we shall rewrite the orbits from the last subsection on the \( p \)-adic level and reformulate Theorem 2.2 in order to obtain a sequence of measures commonly invariant under a single subgroup.

Let \( p \) be an odd prime. In what follows we will often write \( G \) for \( SO_3 \). Let \( v \in Z^3 \) be a fixed primitive integer point with \( D = Q_0(v) \in \mathbb{D}(p) \). Furthermore, let for any \( d \in \mathbb{D}(p) \) a primitive integer point \( w_d \in Z^3 \) of length \( \sqrt{d} \) be given. By Proposition 2.5 we may choose for any \( d \) a rotation \( k_d \in G(R \times Z_p) \) such that
\[
k_{d,p} \frac{v}{\sqrt{d}} = \frac{w_d}{\sqrt{d}} \quad k_{d,\infty} \frac{v}{\sqrt{d}} = \frac{w_d}{\sqrt{d}}.
\]
In particular,
\[
k_d \mathbb{H}_v(R \times Z_p)k_d^{-1} = \mathbb{H}_{w_d}(R \times Z_p).
\]
The projection of the adelic stabilizer orbit \( G(Q) \mathbb{H}_{w_d}(A) \) onto the \( p \)-adic extension \( X_{p,\infty} = G(Z \frac{1}{p})|G(R \times Q_p) \) is denoted after right-multiplication with \( k_d \) as
\[
O_d := \bigsqcup_{\rho \in \mathcal{R}_d} G(Z \frac{1}{p}) \rho k_d \mathbb{H}_v(R \times Z_p) \subset X_{p,\infty} := G(Z \frac{1}{p})|G(R \times Q_p)
\]
for a finite set \( \mathcal{R}_d \subset G(R \times Z_p) \).

For any \( d \) we equip the collection \( O_d \) with the pushforward of the measure \( \mu_{w_d} \) by \( k_d \), which we denote for simplicity by \( \mu_d \). It is not hard to verify that all the statements from Lemma 2.3 transfer to this slightly adapted situation, showing in particular that \( \mu_d \) is \( \mathbb{H}_v(R \times Q_p) \)-invariant. By compactness of the group \( SO_3(R \times Z_p) \) and the definitions of \( \mu_d \) and \( \mu_{w_d} \) respectively Theorem 2.2 is equivalent to the following statement.

**Theorem 2.6** (Equidistribution of \( O_d \)). As \( d \to \infty, d \in \mathbb{D}(p), \) the measures \( \mu_d \) equidistribute to the Haar measure on \( X_{p,\infty} \).

Let \( a \in \mathbb{H}_v(Q_p) \) be the element associated to \( \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \) under the isomorphism \( PGL_2(Q_p) \cong G(Q_p) \) from Proposition 2.5. We will reduce Theorem 2.6 to showing that any weak*-limit of the sequence \( (\mu_d) \) has maximal entropy.

**Theorem 2.7** (Maximal Entropy). Any weak*-limit of the sequence \( (\mu_d) \) has entropy \( \geq \log(p) \) with respect to right-multiplication by \( a^{-1} \).

For the reduction of Theorem 2.6 to Theorem 2.7 (see Section 2.6) and also for later purposes we will recall a few facts about the exponential map on (certain) \( p \)-adic Lie groups and about horospherical subgroups in Section 2.5. Before doing so, we will prove Linnik’s Theorem A in the next subsection using Theorem 2.6.
2.4. Proof of Linnik’s Theorem A. We keep the notation from that last subsection and use Theorem 2.6 to obtain Linnik’s Theorem on the sphere \( S^2 \), which we identify with \( \text{SO}(3) / \mathbb{H}_v(\mathbb{R}) \). Let \( I_d \) be the set of primitive integer points in \( \mathbb{R}^3 \) with norm \( \sqrt{d} \) projected onto the sphere \( S^2 \). Pushing the measures \( \mu_d \) and the normalized Haar measure \( m_{X_{p,\infty}} \) forward under the projection \( X_{p,\infty} \to X_{\infty} \), we obtain measures \( \nu_d \), which equidistribute to the (normalized) Haar measure on the real quotient \( X_{\infty} \). Using yet another projection, we could obtain a similar statement on \( G(\mathbb{Z}) S^2 =: Y \). We prefer to avoid this double quotient though. Instead, we will use the following characterization of probability measures on \( X_{\infty} \) (resp. \( Y \)) to obtain a “lift”. Denote by \( \pi_{X_{\infty}} : G(\mathbb{R}) \to X_{\infty} \) the natural projection.

**Lemma 2.8.** The map \( (\pi_{X_{\infty}})_\ast \) restricted to the set of \( G(\mathbb{Z}) \)-invariant probability measures on \( G(\mathbb{R}) \) is a homeomorphism

\[
\{ G(\mathbb{Z}) \text{-invariant prob. measures on } G(\mathbb{R}) \} \to \{ \text{prob. measures on } X_{\infty} \}.
\]

The analogous statement holds on the sphere.

Given a probability measure \( \mu \) on \( X_{\infty} \) (resp. on \( Y \)), we shall refer to the unique \( G(\mathbb{Z}) \)-invariant probability measure \( \tilde{\mu} \) on \( G(\mathbb{R}) \) (resp. \( S^2 \)) as the lift of \( \mu \). The lift of the Haar measure \( m_{X_{\infty}} \) is the normalized Haar measure on \( G(\mathbb{R}) \).

**Proof.** Abbreviate \( \pi = \pi_{X_{\infty}} \), \( G = G(\mathbb{R}) \), \( \Gamma = G(\mathbb{Z}) \) and \( X = X_{\infty} \). We will identify measures \( \mu \) with the associated positive linear functionals \( \mu(\varphi) = \int \varphi d\mu \). Notice that continuous functions \( \varphi \) on \( X \) correspond linearly to left-\( \Gamma \)-invariant continuous functions \( \varphi \) on \( G \). For an arbitrary continuous function \( \varphi \) on \( G \) we introduce its \( \Gamma \)-mean

\[
\varphi_\Gamma(x) := \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \varphi(\gamma^{-1}x)
\]

which is a left-\( \Gamma \)-invariant function. Observe that if \( \varphi \) is left-\( \Gamma \)-invariant, then \( \varphi = \varphi_\Gamma \). Given a probability measure \( \mu \) on \( X \) define a measure \( f(\mu) \) on \( G \) through

\[
f(\mu)(\varphi) = \mu(\varphi_\Gamma).
\]

One now verifies directly that \( f \) is a two-sided inverse of \( \pi_\ast \). \( \square \)

**Proof of Linnik’s Theorem A.** Given two primitive integer points \( w, w' \) of equal length, we say that \( w \) is equivalent to \( w' \) if for all \( p \) there exists \( g_p \in G(\mathbb{Z}_p) \) with \( g_p w = w' \). In Subsection 2.1, we have seen that the primitive integer points equivalent to a fixed primitive integer point \( w \) are exactly the integer points produced by the adelic stabilizer orbit of \( w \).

We first make the following claim: For any \( d \in \mathbb{D}(p) \), let \( S_d \) be an equivalence class of integer points of norm \( \sqrt{d} \) projected onto \( S^2 \). Then

\[
\frac{1}{|S_d|} \sum_{y \in S_d} \delta_y \to m_{S^2}
\]

as \( d \to \infty \), \( d \in \mathbb{D}(p) \) in the weak* topology.

To see this, take the lift of \( \nu_d \) and project again under \( \text{SO}(3) \to \text{SO}(3) / \mathbb{H}_v(\mathbb{R}) = S^2 \) to obtain a sequence of measures \( \tilde{\nu}_d \) converging to the Haar measure by Lemma 2.8.
We need to show that \( \hat{\nu}_d = \frac{1}{|S_d|} \sum_{y \in S_d} \delta_y \). Let \( \tilde{S}_d = \mathbb{G}(\mathbb{Z})/S_d \) and note that by Lemma 2.3

\[
(\pi_Y)_*(\frac{1}{|S_d|} \sum_{y \in S_d} \delta_y) = \frac{1}{|\tilde{S}_d|} \sum_{y \in \tilde{S}_d} \delta_y.
\]

On the other hand, since the diagram of projections

\[
\begin{array}{ccc}
SO(3) & \xrightarrow{\pi_X} & \mathbb{S}^2 \\
\pi_X & \downarrow & \pi_Y \\
X_\infty & \xrightarrow{\pi} & Y
\end{array}
\]

commutes, the measure \((\pi_Y)_* \hat{\nu}_d\) is the pushforward under the projection \(X_\infty \to Y\), which is also \(\frac{1}{|\tilde{S}_d|} \sum_{y \in \tilde{S}_d} \delta_y\). We thus conclude the claim using Lemma 2.8.

The equidistribution statement along equivalence classes can now easily be upgraded to equidistribution of all integer points. For any finite set \(F \subset S^2\) set \(\nu_F = \frac{1}{|F|} \sum_{x \in F} \delta_x\) for simplicity. Suppose by contradiction that \(\nu_{I_d} \to \nu \neq m_{S^2}\) as \(\ell \to \infty\) and choose \(f \in C(S^2)\) so that \(\int f \, d\nu \neq \int f \, dm_{S^2}\). For any \(d \in \mathbb{D}(p)\) write \(I_d\) as a finite union of equivalence classes \(I^1_d, \ldots, I^k_d\) for the equivalence relation defined above. In particular, we may view \(\nu_{I_d}\) as a convex combination

\[
\nu_{I_d} = \sum_{j=1}^{k_d} \frac{|I^j_d|}{|I_d|} \nu_{I^j_d}.
\]

Choose \(\varepsilon > 0\) so that for all large enough \(\ell\), we have \(|\int f \, d\nu_{I_d} - \int f \, dm_{S^2}| \geq \varepsilon\). In particular, there must exist some \(1 \leq j_\ell \leq k_d\) with \(|\int f \, d\nu_{I^j_d} - \int f \, dm_{S^2}| \geq \varepsilon\) for every \(\ell\). This contradicts the claim in (2.4) which implies that \(\nu_{I^j_d} \to m_{S^2}\) as \(\ell \to \infty\).

\[\square\]

### 2.5. Exponential map and horospherical subgroups

Let \(p\) be an odd prime and \(G < SL_d\) be a \(\mathbb{Q}\)-algebraic group. The norm \(\|A\|_p = \max_{ij} |A_{ij}|_p\) on \(\text{Mat}_d(\mathbb{Q}_p)\) induces a norm on the Lie algebra \(\mathfrak{g}\) of \(G = \mathbb{G}(\mathbb{Q}_p)\) by restriction. For simplicity, denote by \(B^G_K\) resp. \(B^\mathfrak{g}_K\) the ball of radius \(p^{-K}\) in \(G\) resp. \(G\) around \(0\) resp. the identity \(I\). Just as for real linear groups, one can define a matrix exponential on \(\text{Mat}_d(\mathbb{Q}_p)\) by the formula

\[
\exp(A) = \sum_{n \geq 0} \frac{A^n}{n!}.
\]

We take the following facts for granted; proofs may be found in [PR94], [Rüh16] and [Ser92]:

(i) The exponential map \(\exp\) is defined on \(B^G_p\) and forms an isometric bijection \(\exp : B^G_p \to B^G_1\). It maps Lie subalgebras to Lie subgroups.

(ii) The image of a \(\mathbb{Z}_p\)-subalgebra of \(\mathfrak{g}\) (a \(\mathbb{Z}_p\)-submodule which is stable under taking commutators) is a subgroup of \(G\). In particular, every ball of radius \(\leq p^{-1}\) is a subgroup of \(G\), since \(B^\mathfrak{g}_K\) is a \(\mathbb{Z}_p\)-subalgebra of \(\mathfrak{g}\) for every \(K\).
Let $a \in G$ be a diagonalizable element. Define the stable/unstable horospherical subgroups associated to $a$ as

$$G_a^- = \left\{ g \in G \mid a^n ga^{-n} \to e \text{ as } n \to \infty \right\}$$

$$G_a^+ = \left\{ g \in G \mid a^n ga^{-n} \to e \text{ as } n \to -\infty \right\}$$

and let $G_a^0 = C_G(a)$ be the centralizer of $a$. The groups $G_a^-, G_a^+, G_a^0$ are closed subgroups and the Lie algebras corresponding to the horospherical subgroups are

$$\mathfrak{g}_a^\pm = \left\{ X \in \mathfrak{g} \mid \text{Ad}_a^n(X) \to 0 \text{ as } n \to \pm \infty \right\}.$$

Moreover, the Lie algebra $\mathfrak{g}_a^+$ is the direct sum of the eigenspaces of $\text{Ad}_a$ associated eigenvalues of norm (strictly) less than one and the analogous statements hold for $\mathfrak{g}_a^-$ and $\mathfrak{g}_a^0$ as $a$ is diagonalizable. We have the decomposition

$$\mathfrak{g} = \mathfrak{g}_a^+ + \mathfrak{g}_a^- + \mathfrak{g}_a^0.$$

The main example relevant for our purposes (compare Proposition 2.5) is the following:

**Example 2.9.** Let $G = \operatorname{PGL}_2$ and $a = \left( \begin{array}{cc} p & 1 \\ 0 & 1 \end{array} \right) \in \operatorname{PGL}_2(\mathbb{Q}_p)$. A direct computation shows that

$$G_a^- = \left\{ \left( \begin{array}{cc} 1 & x \\ 0 & 1 \end{array} \right) : x \in \mathbb{Q}_p \right\}, \quad G_a^+ = \left\{ \left( \begin{array}{cc} 1 & 0 \\ x & 1 \end{array} \right) : x \in \mathbb{Q}_p \right\},$$

$$G_a^0 = \left\{ \left( \begin{array}{cc} x & 0 \\ 0 & 1 \end{array} \right) : x \in \mathbb{Q}_p \right\}$$

as well as the identities in $\mathfrak{pgl}_2 = \mathfrak{sl}_2$

$$\mathfrak{g}_a^- = \left\{ \left( \begin{array}{cc} 0 & x \\ 0 & 0 \end{array} \right) : x \in \mathbb{Q}_p \right\}, \quad \mathfrak{g}_a^+ = \left\{ \left( \begin{array}{cc} 0 & 0 \\ x & 0 \end{array} \right) : x \in \mathbb{Q}_p \right\},$$

$$\mathfrak{g}_a^0 = \left\{ \left( \begin{array}{cc} x & 0 \\ 0 & -x \end{array} \right) : x \in \mathbb{Q}_p \right\}.$$

The Lie algebra $\mathfrak{g}_a^-$ is the eigenspace of $\text{Ad}_a$ for the eigenvalue $p$ and $\mathfrak{g}_a^+$ is the eigenspace of $\text{Ad}_a$ for the eigenvalue $p^{-1}$.

We will usually consider open rectangles of the kind $B_{K_+,K_0}^0 + B_{K_0}^0 + B_{K_0}^0$ for $K_+, K_0 \geq 1$ instead of balls in $\mathfrak{g}$ as these are well-behaved with respect to conjugation by $a$ (see Lemma 2.10). These sets are open and induce the topology on $\mathfrak{g}$. In fact, by equivalence of norms there exists some $L \geq 0$ so that for all $K$

$$B_{K+L}^0 \subset B_{K}^0 + B_{K}^0 + B_{K}^0,$$

(2.6) Also, note that $B_{K_+}^0 + B_{K_-}^0 + B_{K_0}^0$ is a $\mathbb{Z}_p$-subalgebra; its image is thus a subgroup, which is explicitly given by

$$\exp \left( B_{K_+}^0 + B_{K_-}^0 + B_{K_0}^0 \right) = B_{K+}^{G^+} B_{K-}^{G^-} B_{K_0}^{G_0},$$

(2.7) where we may permute the factors on the right hand side. A proof of this fact based on the $p$-adic version of the Baker-Campbell-Hausdorff formula may be found in [Rüh16]. This together with (2.6) implies that there is some $L \geq 0$ so that

$$B_{K+L}^G \subset B_{K}^0 B_{K}^{G^+} B_{K}^{G^-} \subset B_{K-L}^G,$$

(2.8) for all large enough $K$. 
Lemma 2.10. For $G = \text{PGL}_2$ and any $N_1, N_2 \geq 0$ we have

$$\bigcap_{k=-N_1}^{N_2} a^{-k} B_K^{G_+} B_K^{G_0} B_K^{G_0} a^k = B_{K+N_1}^{G_+} B_K^{G_0} B_{K+N_2}^{G_0}.$$ 

Note that Lemma 2.10 holds in greater generality (see Lemma 4.1 in [Rüh16]).

Proof. The statement is true on the Lie-algebra level as is also $G$ and $X$ and recalling that $\text{PGL}_2$.

The claim.

by Example 2.9. Applying the exponential map and using Equation 2.7, one obtains

Proof of Theorem 2.6 assuming maximal entropy for $G$ is a Borel probability measure invariant under a diagonalizable element $\sigma \in \Gamma$. Suppose that $\mu$ is a topological entropy of the dynamical system $(X_{p,\infty}, R_{a^{-1}})$. This can be verified using the formula

$$h_{\text{top}}(R_{a^{-1}}) = \lim_{K \to \infty} \limsup_{n \to \infty} \frac{-\log(m_G(B_K^{G_+} B_K^{G_0} B_{K+n-1}^{G_0})))}{n}$$

which in turn follows from Lemma 2.10 and Equation (2.8).

2.6. Uniqueness of measures of maximal entropy. In this subsection, we will see how Theorem 2.7 implies Theorem 2.6 using the following theorem (a special case of Theorem 7.9 in [EL07]) to characterize measures of maximal entropy.

Theorem 2.11 (Additional invariance for measures of maximal entropy). Let $G$ be a $\mathbb{Q}$-algebraic group, $G = G(\mathbb{R} \times \mathbb{Q}_p)$, $\Gamma < G$ a lattice and $X = G(\mathbb{Q}_p)$. Suppose that $\mu$ is a weak$^*$-limit of the sequence $(\mu_d)$. By Theorem 2.7 and Theorem 2.11, $\mu$ is invariant under the subgroups $G(\mathbb{Q}_p)_a^+$ and $G(\mathbb{Q}_p)^-$. However, note that $\mu$ is also $G(\mathbb{Q}_p)_a^+$ and $G(\mathbb{Q}_p)^-$. This follows using the isomorphism from Proposition 2.5 and recalling that $\text{PGL}_2(\mathbb{Q}_p)$ is generated by upper triangular, lower triangular and diagonal matrices. We have thus shown that $\mu$ is $G(\mathbb{Q}_p)$-invariant.

It remains to verify that any such measure $\mu$ is also $G(\mathbb{R})$-invariant. For this note first that the subgroup $G(\mathbb{Z}_p^{1,\infty})$ is dense in $G(\mathbb{R})$. This can be proven either directly or using the Borel density theorem. Abbreviate $G_\sigma = G(\mathbb{Q}_p)$ for $\sigma \in \{p, \infty\}$, $G = G_\infty \times G_p$ and $\Gamma = G(\mathbb{Z}_p^{1,\infty})$. We have the following one-to-one correspondences

right $G_p$-invariant finite measures on $X_{p,\infty}$

$\leftrightarrow$ right $G_p$-invariant and left $\Gamma$-invariant locally finite measures on $G$

$\leftrightarrow$ left $\Gamma$-invariant finite measures on $G/G_p \cong G_\infty$

A left $\Gamma$-invariant finite measure $\nu$ on $G_\infty$ is left $G_\infty$-invariant: For any $g_\infty \in G_\infty$ choose a sequence $(\gamma_k)$ in $\Gamma$ with $\gamma_k \to g_\infty$. Then $(L_{g})_* \nu \leftarrow (L_{\gamma_k})_* \nu = \nu$. □
2.7. **Maximal entropy.** In this subsection, we prove Theorem 2.7 using a further number theoretic input (Theorem 2.14). We keep the notation of Subsection 2.3 and proceed as follows: We first show that in order to prove Theorem 2.7 the real place may be “forgotten” which allows us to somewhat simplify the procedure and in particular to use properties of ultrametric spaces. This is possible as the action of $a$ does not influence the real place (at least morally speaking).

2.7.1. **A reduction step.** Taking the right quotient by $G(\mathbb{R})$ induces a projection 

$$
\pi : X_{p,\infty} = G(\mathbb{Z}\frac{1}{p})\backslash G(\mathbb{R} \times \mathbb{Q}_p) \rightarrow X := G(\mathbb{Z})\backslash G(\mathbb{Z}_p)
$$

which is given by $G(\mathbb{Z}\frac{1}{p})(g_\infty, g_p) \mapsto G(\mathbb{Z})g_p$ for $(g_\infty, g_p) \in G(\mathbb{R} \times \mathbb{Z}_p)$. Since $a$ commutes with $G(\mathbb{R})$, the action of $a$ on $X_{p,\infty}$ by $R_a^{-1}$ induces a unique map $T : X \rightarrow X$, which satisfies the commutative diagram

$$
\begin{array}{ccc}
X_{p,\infty} & \xrightarrow{R_a^{-1}} & X_{p,\infty} \\
\downarrow \pi & & \downarrow \pi \\
X & \xrightarrow{T} & X
\end{array}
$$

However, the map $T$ is not given by right-multiplication with $a^{-1}$ as $a$ does not lie in $G(\mathbb{Z}_p)$. Nevertheless, the local dynamics is still given by conjugation: For any $x \in X$ and any $h \in G(\mathbb{Z}_p)$ with $aha^{-1} \in G(\mathbb{Z}_p)$ we have

$$
T(xh) = T(x)aha^{-1}
$$

as is straightforward to verify. The projection of the collection of orbits $\mathcal{O}_d$ defined in Equation (2.3) is given by

$$
\pi(\mathcal{O}_d) = \bigsqcup_{\rho \in \mathcal{R}_d} G(\mathbb{Z})\rho_p k_d \mathbb{H}_v(\mathbb{Z}_p).
$$

We equip $\pi(\mathcal{O}_d)$ with the pushforward of the measure $\mu_d$ on $\mathcal{O}_d$ under the projection and denote it by $\bar{\mu}_d$. The measure $\bar{\mu}_d$ is $\mathbb{H}_v(\mathbb{Z}_p)$- and $T$-invariant and is given by the normalized sum of the Haar measures on the $\mathbb{H}_v(\mathbb{Z}_p)$ orbits in $\pi(\mathcal{O}_d)$.

Let now $\mu$ be a weak$^\ast$-limit of the sequence $(\mu_d)$ and let $\bar{\mu} := \pi_* \mu$. By the Abramov-Rokhlin formula (see [AR65])

$$
h_\mu(a) = h_{\bar{\mu}}(T) + h_\mu(a|\mathcal{A})
$$

where $\mathcal{A} = \pi^{-1}\mathcal{B}_X$. In view of Theorem 2.7 it thus suffices to prove that $h\bar{\mu}(T)$ is bounded from below by $\log(p)$.

2.7.2. **Linnik’s basic lemma and maximal entropy.** We re-introduce some notation first. Let $v \in \mathbb{Z}^3$ be a primitive integer point with $D := Q_0(v) \in \mathbb{D}(p)$. Let $(w_d)$ be a sequence of primitive integer points with $d = Q_0(w_d) \in \mathbb{D}(p)$, $d \rightarrow \infty$. By Proposition 2.5, choose $k_d \in G(\mathbb{Z}_p)$ with

$$
k_d \frac{v}{\sqrt{-D}} = \frac{w_d}{\sqrt{-d}}
$$

Let $\Gamma := G(\mathbb{Z})$ and let

$$
\mathcal{O}_d := \bigsqcup_{\rho \in \mathcal{R}_d} \Gamma \rho k_d \mathbb{H}_v(\mathbb{Z}_p) \subset X := \Gamma \backslash G(\mathbb{Z}_p)
$$
be the projection of the adelic stabilizer orbit of \( w_d \) onto \( X \) multiplied with \( k_d \) where \( \mathcal{R}_d \subset \mathbb{G}(\mathbb{Z}_p) \) is a finite set. Let \( \mu_d \) be the measure on \( \mathcal{O}_d \) given by the normalized sum of the Haar measures on the orbits in \( \mathcal{O}_d \) which is invariant under the map \( T : X \to X \) defined in the last subsection. Set for simplicity \( G = \mathbb{G}(\mathbb{Q}_p) \). The central ingredient of the proof of Theorem 2.7 is the following lemma:

**Proposition 2.12** (Linnik’s basic lemma). For any \( \delta > 0 \) with \( d^{-\frac{1}{3}} \leq \delta < r \)

\[
\mu_d \times \mu_d \left( \{ (x, y) \in X^2 \mid d(x, y) \leq \delta \} \right) \ll \varepsilon \delta^3 d^5
\]

where \( \varepsilon > 0 \) is arbitrary and \( r < 1 \) is a uniform injectivity radius on \( X \).

The proof of Linnik’s basic lemma uses a theorem on representations of binary quadratic forms by ternary quadratic forms. Recall that a representation of an integral quadratic form \( q \) on \( \mathbb{Z}^n \) by an integral quadratic form \( Q \) on \( \mathbb{Z}^m \) is a structure-preserving \( \mathbb{Z} \)-linear map \( \iota : \mathbb{Z}^n \to \mathbb{Z}^m \) i.e. \( \iota \) satisfies \( Q(\iota(x)) = q(x) \) for all \( x \in \mathbb{Z}^n \). Let \( R_Q(q) \) be the set of representations of \( q \) by \( Q \) and observe that \( \text{SO}_Q(\mathbb{Z}) \) acts on \( R_Q(q) \) by post-composition.

**Example 2.13.** Consider the quadratic forms \( q(z) = dz^2 \) and \( Q(x, y) = xy \) for some integer \( d \). A representation \( \iota : \mathbb{Z} \to \mathbb{Z}^2 \) corresponds to a choice of image \( \iota(1) \in \mathbb{Z}^2 \), that is, a point \( (x, y) \in \mathbb{Z}^2 \) with \( xy = d \). The number \( |R_Q(q)| \) is thus exactly the number of divisors of \( d \). The divisor function \( \chi(n) := \sum_{d|n} 1 \) satisfies \( \chi(n) \ll n^\varepsilon \) for any \( \varepsilon > 0 \).

The proof of Proposition 2.12 needs the following number-theoretic input:

**Theorem 2.14.** Let \( Q \) be a non-degenerate integral ternary quadratic form and let \( q(x, y, z) = ax^2 + bxy + cy^2 \) be a non-degenerate integral binary quadratic form. Let \( \text{gcd}(a, b, c) \) be the greatest common square divisor of \( a, b, c \). The number of embeddings \( (\mathbb{Z}^2, q) \) into \( (\mathbb{Z}^3, Q) \) modulo the action of \( \text{SO}_Q(\mathbb{Z}) \) is \( \ll_Q \varepsilon \max(|a|, |b|, |c|)^\varepsilon \) where \( \varepsilon > 0 \) is arbitrary.

Venkov [Ven31] provided a first proof of Theorem 2.14 in the special case \( Q = Q_0 \) (which is one of interest to us). A conceptual proof of Theorem 2.14 by counting on the tree \( \text{SO}_3(\mathbb{Q}_p)/\text{SO}_3(\mathbb{Z}_p) \) may be found in [ELMV12, Appendix A].

**Proof of Proposition 2.12.** We first claim that the finite set

\[
I_{d, \delta} = \{ (w_1, w_2) \in \mathbb{Z}_p^3 \mid Q_0(w_1) = Q_0(w_2) = d, 0 < \|w_1 - w_2\|_p \leq \delta \}
\]

satisfies

\[
|I_{d, \delta}| \ll \varepsilon d^2 d^{1+\varepsilon}.
\]

From this we will deduce the proposition by attaching to any pair of \( \delta \)-close points in \( \mathcal{O}_d \) their associated integer points.

Let \( (w_1, w_2) \in I_{d, \delta} \) be given and set \( Q \) to be the integral quadratic form

\[
Q(x, y) := Q_0(xw_1 + yw_2) = dx^2 + \ell xy + dy^2
\]

for some \( \ell \in \mathbb{Z} \). Note that its coefficients satisfy

\[
|2d - \ell|_p = |Q(1, -1)|_p \leq \|w_1 - w_2\|_p^2 \leq \delta^2.
\]

Since \( w_1 \) and \( w_2 \) both have Euclidean norm \( d^{\frac{1}{4}} \) we also have

\[
|2d - \ell| = |Q(1, -1)| = \|w_1 - w_2\|_{\text{eucl}}^2 \leq 4d.
\]
For convenience, set $m = \lfloor -2 \log_p (\delta) \rfloor$ so that by (2.11) we have $p^m \leq (2d - \ell)$. The quadratic form $Q$ is non-degenerate: By Equation (2.12) $\ell \neq 2d$, since otherwise $w_1 = w_2$ and by Equation (2.11) $\ell \neq -2d$, since otherwise $1 = |d|_p = \delta^2$ which contradicts $\delta < 1$. Denote by $N_{\ell,d}$ the number of inequivalent ways of representing of $dx^2 + \ell xy + dy^2$ by $Q_0$ which satisfies by Theorem 2.14

$$N_{\ell,d} \ll f \max(|d|, |\ell|)^{2} \leq f \max(|d|, |2d - \ell| + |2d|)^{2} \ll f d^\varepsilon$$

where $f^2 = \gcd(d, \ell)$ is the greatest common square divisor of $\ell$ and $d$. For $L \leq 4d$ compute

$$|I_{d,\delta}| \leq \sum_{\ell:|2d-\ell| \leq L, \ell \notin \pm 2d, \ell \neq 2d} N_{\ell,d} = \sum_{\ell:|\ell| \leq L, \ell \neq 0, d} N_{2d-\ell,\delta} \leq \sum_{f^2 | d} \sum_{\ell':|\ell'| \leq L, p^m|\ell', f^2 = \gcd(\ell', d), \ell' \neq 0, d} f d^\varepsilon \ll \sum_{f^2 | d} \sum_{\ell':|\ell'| \leq L, p^m|\ell', f^2 = \gcd(\ell', d)} f d^\varepsilon = \sum_{f^2 | d} f d^\varepsilon \sum_{\ell':|\ell'| \leq L, p^m|\ell', f^2 = \gcd(\ell', d)} 1.$$ 

Now observe that the number of $\ell'$ satisfying $p^m | \ell'$, $f^2 | \ell'$ and $|\ell'| \leq L$ is $\ll \frac{L}{p^m f^2}$, since $f^2$ and $p^m$ are coprime ($p$ does not divide $d$). Thus, by Example 2.13

$$|I_{d,\delta}| \ll \frac{L}{p^m f^2} \ll \frac{d^\varepsilon \delta^2 L}{d} \leq d^{1+\varepsilon} \delta^2 \frac{1}{d} \ll d^{1+2\varepsilon} \delta^2$$

which finishes the proof of the claim in (2.10).

Now let $x_1 = \Gamma g_1$, $x_2 = \Gamma g_2 \in O_d$ be $\delta$-close, where we choose $g_1, g_2 \in G(\mathbb{Z}_p)$ with $d(g_1, g_2) \leq \delta$. By the discussion in Section 2.1 the points

$$w_i = \frac{\sqrt{d}}{\sqrt{\delta}} g_i v$$

are primitive integer points of Euclidean norm $\sqrt{d}$, which satisfy $\|w_1 - w_2\|_p \leq \delta$.

**Case 1** – equal integer points. If $w_1 = w_2$ then $x_1$ and $x_2$ lie on the same orbit. The volume of a $\delta$-ball in $H_1(\mathbb{Z}_p)$ is $\ll \delta$ and in particular, the set of $\delta$-close pairs $(x_1, x_2) \in O_d$ that lie on the same orbit has volume $\ll \delta^{\frac{d}{2} + \varepsilon} \delta$ by Fubini’s theorem. After normalization, the contribution to the total mass is $\ll \delta^{\frac{d}{2} + \varepsilon} \delta \leq d^\varepsilon \delta^3$ in this case.

**Case 2** – distinct integer points. For fixed $(w_1, w_2) \in I_{d,\delta}$ the set of pairs $(\Gamma g_1, \Gamma g_2)$ as above with associated integer pair $(w_1, w_2)$ has volume $\ll \delta$ by Fubini’s theorem and thus the volume in total is $\ll |I_{d,\delta}| \delta \ll d^{1+2\varepsilon} \delta^3$. After normalization, the measure contribution in this case is therefore $\ll \delta^{3 \varepsilon} \delta^3$. □

As in Section 2.5 denote by $B_{K^\pm_0}^{\pm_0}$ the ball of radius $p^{-K}$ in $G_{a_0}^{\pm_0}$.

**Lemma 2.15** (A suitable partition). There exists a finite partition $\mathcal{P}$ of $X$ such that for any $x \in X$ and any $N \geq 0$ we have $[x]_{\mathcal{P}^N} = xB_{K^+_{K+N}}^{G_n^+}B_{K^-_{K+N}}^{G_n^-}B_{K^0_{K+N}}^{G_n^0}$ where $p^{-K}$ for $K \geq 1$ is a small enough uniform injectivity radius on $X$.

The atoms of the partition $\mathcal{P}$ above thus shrink exactly by a factor $p^{-N}$ along the $G_{a_0}^{\pm}$-directions when one refines the partition $N$ times under the dynamics.
**Proof.** Let \( p^{-K} \) be a uniform injectivity radius on \( X \) for \( K \geq 1 \). We abbreviate \( B := B_K^{G_0} B_K^{G_0} B_K^{G_0} \) and recall from Subsection 2.5 that \( B \) is a group. Also, note that \( B \in B_K^{G_0} \) since \( B_K^{G_0} \) is a group. By choice of \( K \), the map \( B \to X \), \( g \mapsto xg \) is injective for any \( x \).

There exists a finite partition \( \mathcal{P} \) of \( X \), whose atoms satisfy \( [x]_\mathcal{P} = xB \): Since \( B \) is a group, \( xB \), \( yB \) are either equal or disjoint for any \( x \), \( y \in X \). \( X \) is compact and \( xB \) is open for any \( x \), thus there exists such a partition. By shrinking \( K \), we may assume that \( B \subset B_K^{G_0} \cap a^{\pm 1} B_K^{G_0} a^{\mp 1} \) where \( K' \geq 1 \) such that \( p^{-K'} \) is a uniform injectivity radius on \( X \).

We now verify that the partition \( \mathcal{P} \) has the required properties. Let \( x \) be a uniform injectivity radius on \( K \) and \( N \geq 0 \). We first show that the \( \mathcal{P}_N \)-atom of \( x \) is contained in the set \( xB_K^{G_0} \). Let \( y = xg_0 \in [x]_{\mathcal{P}_N} \) for \( g_0 \in B \). Since \( T(y) = [T(x)]_\mathcal{P} \) we may write \( T(y) = T(x)g_1 \) with \( g_1 \in B \). On the other hand, \( T(y) = T(x)(ag_0a^{-1}) \) by Equation (2.9). But both \( g_1 \) and \( ag_0a^{-1} \) lie inside the injective ball \( B_K^{G_0} \), and hence \( g_1 = ag_0a^{-1} \). Proceeding this way, we find elements \( g_1, g_2, \ldots, g_N \) in \( B \) where \( g_n = a^ng_0a^{-n} \) for each \( n \). Applying the same method to \( T^{n-1} \) instead of \( T \) we obtain \( g_0 \in \bigcap_{n=0}^{N} g_{a^n}B_K^{G_0} \) by Lemma 2.10. For the remaining inclusion let \( g \in B_K^{G_0} \) and \( y = xg \). Then \( a^nga^{-n} \) is in \( B \) for all \( n \) with \( -N \leq n \leq N \) by Lemma 2.10 and thus \( T^n(y) = T^n(x)a^nga^{-n} \) for \( -N \leq n \leq N \) by Equation (2.9).

**Proof of Theorem 2.7.** Recall that for any partition \( \mathcal{P}' \) of \( X \) the inequality

\[
H_{\mu_d}(\mathcal{P}') \geq -\log \left( \sum_{P \in \mathcal{P}'} \mu_d(P)^2 \right)
\]

holds. Let \( \mathcal{P} \) be the finite partition constructed in Lemma 2.15. We first claim that there exist \( a_1, \ldots, a_{p^N} \in \mathbb{H}_o(Z_p) \) with

\[
\bigcup_{S \in \mathcal{P}_N^{p^N}} S \times S \subset \bigcup_{i=1}^{p^N} \{(x, ya_i) \in X^2 \mid d(x, y) < p^{-(K+N)}\}.
\]

To see this, choose \( a_1, \ldots, a_{p^N} \in B_K^{G_0} \) with \( B_K^{G_0} = \bigsqcup_{i=1}^{p^N} B_K^{G_0}a_i \) and observe that

\[
B_{K+N}^{G_0} B_K^{G_0} B_{K+N}^{G_0} = B_{K+N}^{G_0} B_{K+N}^{G_0} B_K^{G_0} = \bigcup_{i=1}^{p^N} B_{K+N}^{G_0} B_{K+N}^{G_0} B_{K+N}^{G_0} a_i \subset \bigcup_{i=1}^{p^N} B_{K+N}^{G_0} a_i.
\]

Given any \( S \in \mathcal{P}_N^{p^N} \) and \( (x, y) \in S \times S \) we have \( y \in xB_K^{G_0} B_K^{G_0} B_{K+N}^{G_0} \subset xB_{K+N}^{G_0} a_i \) for some \( i \).

By the above claim, \( \mathbb{H}_o(Z_p) \)-invariance of \( \mu_d \) and Linnik’s basic lemma for the choice \( \delta = p^{-(K+N)} \) we obtain

\[
\sum_{S \in \mathcal{P}_N^{p^N}} \mu_d(S)^2 \ll \varepsilon p^{-2N} \ll \varepsilon^2 p^{-2N}
\]
if \( d^{-\frac{1}{4}} \leq \delta \) or equivalently, \( N \leq \frac{1}{4} \log_p(d) - K \). Thus, set \( N_d = \lfloor \frac{1}{4} \log_p(d) \rfloor \). Let \( C(\varepsilon) \) be the implicit constant appearing in the estimate above. Then

\[
H_{\mu_d}(\mathcal{P}_{-N_d}^N) \geq -\log \left( \sum_{S \in \mathcal{P}_{-N}^N} \mu_d(S)^2 \right) \geq -\log(C(\varepsilon)) - \varepsilon \log(d) + 2N_d \log(p)
\]

Note that \( \log(d) \leq 5N_d \log(p) + 5 \log(p) \) and \( \log(C(\varepsilon)) + 5\varepsilon \log(p) \leq \varepsilon N_d \log(p) \) if \( d \) is large enough. Hence,

\[
H_{\mu_d}(\mathcal{P}_{-N_d}^N) \geq (2 - 6\varepsilon)N_d \log(p).
\]

We eliminate the dependency on \( d \) in the refinement of the partition \( \mathcal{P} \): Choose \( \ell' \) such that \( 2N_d + \ell \geq \ell\ell' \geq 2N_d + 1 \). Now observe that the partition

\[
\bigvee_{j=0}^{\ell'-1} T^{-j\ell} \mathcal{P}_0^{\ell-1} = \mathcal{P}_0^{\ell\ell'-1}
\]

is finer than the partition \( \mathcal{P}_0^{2N_d} \). Hence

\[
H_{\mu_d}(\mathcal{P}_{-N_d}^N) = H_{\mu_d}(\mathcal{P}_0^{2N_d}) \leq \ell' H_{\mu_d}(\mathcal{P}_0^{\ell-1}) \leq \frac{2N_d + \ell}{\ell} H_{\mu_d}(\mathcal{P}_0^{\ell-1})
\]

and therefore

\[
H_{\mu_d}(\mathcal{P}_0^{\ell-1}) \geq \frac{\ell}{2N_d + \ell} (2 - 6\varepsilon)N_d \log(p).
\]

As the partition \( \mathcal{P}_0^{\ell-1} \) consists of clopen sets, letting \( d \to \infty \) shows that

\[
H_\mu(\mathcal{P}_0^{\ell-1}) \geq \frac{\ell}{2} (2 - 6\varepsilon) \log(p)
\]

and since \( \varepsilon \) was arbitrary

\[
\frac{1}{\ell} H_\mu(\mathcal{P}_0^{\ell-1}) \geq \log(p).
\]

This yields the theorem when taking the limit as \( \ell \to \infty \). \( \square \)
3. Equidistribution of CM points

3.1. Algebraic tori associated to quadratic number fields. Consider the imaginary quadratic number field $\mathbb{K} := \mathbb{Q}(\sqrt{d})$ for a negative square-free integer $d$ and let $\mathcal{O}_\mathbb{K} = \mathcal{O}_d$ be its ring of integers. Denote by $\text{Cl}(\mathcal{O}_d)$ the class group of $\mathbb{K}$ and by $h_d$ the class number. Furthermore, consider a fractional ideal $a$ in $\mathbb{K}$ and choose a $\mathbb{Z}$-basis $(a_1, a_2)$ of $a$. Given an element $\lambda \in \mathbb{K}$, we represent the multiplication $R_\lambda$ on $\mathbb{K}$ in the basis $(a_1, a_2)$ to obtain a matrix $\psi_a(\lambda) \in \text{Mat}_2(\mathbb{Q})$ and choose $\psi_a(\lambda) \in \text{Mat}_2(\mathbb{Q})$ to act on row vectors in $\mathbb{Q}^2$ from the right. This yields an embedding of $\mathbb{Q}$-algebras

$$
\psi_a : \mathbb{K} \hookrightarrow \text{Mat}_2(\mathbb{Q}).
$$

Denoting by $\iota_a : \mathbb{Q}^2 \rightarrow \mathbb{K}$ the isomorphism induced by the choice of basis of $a$ we obtain the commutative diagram

\[
\begin{array}{ccc}
\mathbb{Q}^2 & \xrightarrow{\psi_a(\lambda)} & \mathbb{Q}^2 \\
\downarrow{\iota_a} & & \downarrow{\iota_a} \\
\mathbb{K} & \xrightarrow{R_\lambda} & \mathbb{K}
\end{array}
\]

Observe that $\psi_a(\lambda)$ has integer entries if and only if $R_\lambda$ preserves $a = \iota_a(\mathbb{Z}^2)$. That is,

(3.1) \hspace{1cm} \psi_a(\lambda) \in \text{Mat}_2(\mathbb{Z}) \iff \lambda \in \mathcal{O}_\mathbb{K}

by properness of $a$ and in particular

(3.2) \hspace{1cm} \psi_a(\lambda) \in \text{GL}_2(\mathbb{Z}) \iff \lambda \in \mathcal{O}_\mathbb{K}^\times.

**Definition 3.1.** Set $v_a := \psi_a(\sqrt{d}) \in \text{Mat}_2(\mathbb{Z})$ (by (3.1)). The $\mathbb{Q}$-algebraic torus associated to $a$ is given by

$$
T_a(\mathbb{R}) := \{ h \in \text{PGL}_2(\mathbb{R}) \mid hv_a h^{-1} = v_a \} = \{ h \in \text{PGL}_2(\mathbb{R}) \mid \forall \lambda \in \mathbb{K} : h\psi_a(\lambda)h^{-1} = \psi_a(\lambda) \}
$$

for any ring $\mathbb{R}$.

Note that the characteristic polynomial of $v_a$ is $x^2 - d$. The eigenvalues of $v_a$ (or more precisely its conjugacy class) yield a lot of information about the group $T_a$ as we shall presently see. Denote by $\overline{\psi_a}$ the composition of $\psi_a : \mathbb{K}^\times \rightarrow \text{GL}_2(\mathbb{Q})$ and the projection $\text{GL}_2(\mathbb{Q}) \rightarrow \text{PGL}_2(\mathbb{Q})$.

**Claim 3.2.** We have $\psi_a(\mathbb{K}) = \{ h \in \text{Mat}_2(\mathbb{Q}) \mid hv_a = v_a h \}$ and $T_a(\mathbb{Q}) = \overline{\psi_a}(\mathbb{K}^\times)$. Furthermore, $T_a(\mathbb{Z}) = \overline{\psi_a}(\mathcal{O}_\mathbb{K}^\times)$.

**Proof.** The dimension of $\{ h \in \text{Mat}_2(\mathbb{Q}) \mid hv_a = v_a h \}$ over $\mathbb{Q}$ is the same as the dimension of $\{ h \in \text{Mat}_2(\mathbb{Q}) \mid hv_a = v_a h \}$, the latter being 2. The second statement follows from the observation we made in (3.2). \qed

**Lemma 3.3.** We have

(i) The group of $\mathbb{R}$-points $T_a(\mathbb{R})$ is conjugate to the compact group $\text{PO}(2)$.
(ii) Let \( p \) be a rational prime, which is split (in \( K \)). Then \( \mathbb{T}_a(\mathbb{Q}_p) \) is conjugate to the diagonal subgroup

\[
\left\{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \mid a \in \mathbb{Q}_p^\times \right\} < \text{PGL}_2(\mathbb{Q}_p).
\]

Proof. (i): The matrix \( v_a \) is conjugate over \( \mathbb{R} \) to the matrix

\[
\begin{pmatrix} 0 & \sqrt{|d|} \\ -\sqrt{|d|} & 0 \end{pmatrix} =: v_{d,\infty}
\]

and the subgroup of matrices in \( \text{PGL}_2(\mathbb{R}) \) commuting with \( v_{d,\infty} \) is indeed given by

\[
\left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in \text{PGL}_2(\mathbb{R}) \mid a, b \in \mathbb{R} \right\} = \text{PO}(2).
\]

(ii): By Hensel’s lemma, the polynomial \( x^2 - d \) splits over \( \mathbb{Q}_p \) and thus \( v_a \) is diagonalizable. Let \( \varepsilon, -\varepsilon \) be the eigenvalues of \( v_a \). The matrix \( v_a \) is therefore conjugate to

\[
\begin{pmatrix} \varepsilon & 0 \\ 0 & -\varepsilon \end{pmatrix} =: v_{d,p}
\]

and its commutator subgroup is conjugate to

\[
\left\{ g \in \text{PGL}_2(\mathbb{Q}_p) \mid gv_{d,p} = v_{d,p}g \right\} = \left\{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \mid a \in \mathbb{Q}_p^\times \right\}.
\]

\[\square\]

3.2. Compact torus orbits. In this subsection, we proceed just as in Sections 2.1 and 2.2 to obtain a collection of compact orbits in the \( p \)-adic extension

\[
\text{PGL}_2(\mathbb{Z}[\frac{1}{p}]) \backslash \text{PGL}_2(\mathbb{R} \times \mathbb{Q}_p)
\]

which, when projected to the modular surface \( \text{PGL}_2(\mathbb{Z}) \backslash \mathbb{H} \), will yield the CM points. We keep the notation from the last subsection.

Lemma 3.4. The orbit \( \text{PGL}_2(\mathbb{Q})\mathbb{T}_a(\mathbb{A}) \subset \text{PGL}_2(\mathbb{Q})\text{PGL}_2(\mathbb{A}) \) is compact.

The fact that the orbit is closed is proven using standard methods; compactness will follow from Proposition 3.5. As the subgroup \( \mathbb{T}_a(\mathbb{R} \times \hat{\mathbb{Z}}) < \mathbb{T}_a(\mathbb{A}) \) is open and the orbit \( \text{PGL}_2(\mathbb{Q})\mathbb{T}_a(\mathbb{A}) \) is compact, we can write

\[
\text{PGL}_2(\mathbb{Q})\mathbb{T}_a(\mathbb{A}) = \bigsqcup_{\rho \in \mathcal{R}_a} \text{PGL}_2(\mathbb{Q})\rho\mathbb{T}_a(\mathbb{R} \times \hat{\mathbb{Z}})
\]

where \( \mathcal{R}_a \subset \text{PGL}_2(\mathbb{R} \times \hat{\mathbb{Z}}) \) is a finite set of representatives (using the fact that \( \text{PGL}_2 \) has class number one – see (1.1)).

Proposition 3.5 (Cardinality of \( \mathcal{R}_a \)). There is a one-to-one correspondence between the double quotient

\[
\mathbb{T}_a(\mathbb{Q})\big/\mathbb{T}_a(\mathbb{A}) \to \text{Cl}(\mathcal{O}_K)
\]

and ideal classes in \( \text{Cl}(\mathcal{O}_K) \).
Note that there is a bijection
\[
\mathcal{T}_a(\mathbb{Q})/\mathcal{T}_a(A_f) \cong \mathcal{T}_a(\mathbb{Q})/\mathcal{T}_a((\mathbb{R} \times \hat{\mathbb{Z}}))
\]
\[
\mathcal{T}_a(\mathbb{Q})(h_2,h_3,h_5,\ldots)\mathcal{T}_a(\hat{\mathbb{Z}}) \mapsto \mathcal{T}_a(\mathbb{Q})(I,h_2,h_3,h_5,\ldots)\mathcal{T}_a((\mathbb{R} \times \hat{\mathbb{Z}}))
\]
In particular, this concludes the proof of Lemma 3.4 as the class number is finite. Further, the number of \(\mathcal{T}_a((\mathbb{R} \times \hat{\mathbb{Z}}))\)-orbits \(|\mathcal{R}_a|\) is exactly the cardinality of the right hand side. Hence, there are exactly \(h_d\) many \(\mathcal{T}_a((\mathbb{R} \times \hat{\mathbb{Z}}))\)-orbits in (3.5) by Proposition 3.5.

The main step in the proof of Proposition 3.5 is the following refinement of Lemma 3.3:

**Proposition 3.6.** Let \(p\) be a rational prime. Then
\[
\mathcal{T}_a(\mathbb{Q}_p)/\mathcal{T}_a(\mathbb{Z}_p) \cong \begin{cases} 
\{0\} & \text{if } p \text{ is inert} \\
\mathbb{Z} & \text{if } p \text{ is split} \\
\mathbb{Z}/2\mathbb{Z} & \text{if } p \text{ is ramified}
\end{cases}
\]

**Proof.** First, observe that since the quotient group \(\mathcal{T}_a(\mathbb{Q}_p)/\mathcal{T}_a(\mathbb{Z}_p)\) is discrete and \(\mathcal{T}_a(\mathbb{Q})\) is dense in \(\mathcal{T}_a(\mathbb{Q}_p)\) we have \(\overline{\psi_a}(\mathbb{K}^\times)\mathcal{T}_a(\mathbb{Z}_p) = \mathcal{T}_a(\mathbb{Q}_p)\mathcal{T}_a(\mathbb{Z}_p)\). The proof uses the trivial observation that for \(x,y \in \mathbb{K}^\times\)
\[
(3.6) \quad \overline{\psi_a}(x)\mathcal{T}_a(\mathbb{Z}_p) = \overline{\psi_a}(y)\mathcal{T}_a(\mathbb{Z}_p) \iff \exists \alpha \in \mathbb{Q}_p^\times : \psi_a(xy^{-1})\alpha \in \text{GL}_2(\mathbb{Z}_p)
\]
and the following fact

**Claim.** Let \(x \in \mathbb{K}\) with \((x)\) coprime to \((p)\). Then \(\psi_a(x) \in \text{GL}_2(\mathbb{Z}_p)\).

**Proof.** Note that the statement is clear if \(x \in \mathcal{O}_\mathbb{K}\). Indeed, we have \(\psi_a(x) \in \text{Mat}_2(\mathbb{Z})\) and \(\det(\psi_a(x)) = \text{Nr}(x) \in \mathbb{Z}_p^\times\). Write \(xb = c\) for two integral ideals \(b,c\) coprime to \((p)\). If there is \(b \in b\) with \((b)\) coprime to \((p)\), then \(c := xb \in c\) also satisfies that \((c)\) is coprime to \((p)\). In particular, \(\psi_a(x) = \psi_a(c)\psi_a(b)^{-1} \in \text{GL}_2(\mathbb{Z}_p)\). If \(p\) is inert, then one can choose any \(b \in b \setminus (p)\). If \(p\) is ramified with \((p) = p^2\), one can choose any \(b \in b \setminus (p)\). Thus assume that \(p\) is split and write \((p) = p_1p_2\). We show that there is some \(b \in b \setminus p_1 \cup p_2\). Choose \(x_1 \in b \setminus p_1\) and \(x_2 \in b \setminus p_2\). If \(x_1 \notin p_2\) or \(x_2 \notin p_1\) we are done. Otherwise, \(x_1 + x_2 \notin p_1 \cup p_2\) does the job. \(\Box\)

**Case 1:** Assume that \(p\) is inert. Set \(\ell := \text{ord}_p(x)\) for \(x \in \mathbb{K}^\times\). Then \(x' := \overline{x}^{\ell} \in \mathbb{K}^\times\) satisfies that \((x')\) is coprime to \((p)\). By the claim \(\psi_a(x') \in \text{GL}_2(\mathbb{Z}_p)\) and by the characterization (3.6), \(\overline{\psi_a}(x)\mathcal{T}_a(\mathbb{Z}_p) = \mathcal{T}_a(\mathbb{Z}_p)\) as \(\psi_a(x)p^{-\ell} = \psi_a(x')\). This shows that \(\mathcal{T}_a(\mathbb{Q}_p) = \mathcal{T}_a(\mathbb{Z}_p)\) in this case.

**Case 2:** Assume that \(p\) is split and write \((p) = p_1p_2\). Let \(y \in \mathbb{K}\) be such that \(\text{ord}_{p_1}(y) = 1\) and \(\text{ord}_{p_2}(y) = 0\). Set \(\ell := \text{ord}_{p_1}(x) - \text{ord}_{p_2}(x)\) for \(x \in \mathbb{K}^\times\).

Now observe that \(x' = \overline{x}^{p^{-\text{ord}_{p_2}(x)}} \in \mathbb{K}^\times\) satisfies \(\text{ord}_{p_1}(x') = \text{ord}_{p_2}(x') = 0\) and therefore \(\psi_a(x') \in \text{GL}_2(\mathbb{Z}_p)\) by the claim. By the characterization (3.6) we obtain \(\overline{\psi_a}(x)\mathcal{T}_a(\mathbb{Z}_p) = \overline{\psi_a}(y)^i\mathcal{T}_a(\mathbb{Z}_p)\). It follows from Lemma 3.3 that there is no \(k \neq 0\) with \(\overline{\psi_a}(y)^k \in \mathcal{T}_a(\mathbb{Z}_p)\) as otherwise \(\mathcal{T}_a(\mathbb{Q}_p)\) would be compact.

**Case 3:** Assume that \(p\) is ramified and write \((p) = p^2\). Choose \(y \in \mathbb{K}\) with \(\text{ord}_p(y) = 1\). Let \(x \in \mathbb{K}^\times\) and set \(\ell := \text{ord}_p(x)\). As before, we obtain \(\overline{\psi_a}(x)\mathcal{T}_a(\mathbb{Z}_p) = \overline{\psi_a}(y)\ell \mathcal{T}_a(\mathbb{Z}_p)\).
We claim that $\psi(y)T_a(Z_p) \neq T_a(Z_p)$ but $\overline{\psi(y)^2T_a(Z_p)} = T_a(Z_p)$. The first assertion follows from the observation that $\det(\psi(y)) = \text{Nr}(y)$ is $p$ times a unit in $Z_p$. The second follows from the claim, which yields that $\psi(y^2/p) \in \text{GL}_2(Z_p)$. \hfill $\Box$

**Proof of Proposition 3.5.** The bijections in Proposition 3.6 are explicitly given by

$$
\psi_a(x)T_a(Z_p) \mapsto \begin{cases} 
0 & \text{if } p \text{ is inert} \\
\text{ord}_p(x) - \text{ord}_p'(x) & \text{if } p \text{ is split and } (p) = pp'
\end{cases}
$$

for $x \in K^\times$ as the proof shows. The induced map

$$
T(A_f)/T(\hat{Z}) \to \bigoplus_{p \text{ ramified}} (\mathbb{Z}/2\mathbb{Z}) \oplus \bigoplus_{p \text{ split}} \mathbb{Z}
$$

is a bijection. The image of $T(Q)$ in the target of (3.7) is thus exactly the image of the subgroup of principal ideals under the map

$$
F_K \to \bigoplus_{p \text{ ramified}} (\mathbb{Z}/2\mathbb{Z}) \oplus \bigoplus_{p \text{ split}} \mathbb{Z},
$$

where $F_K$ is the group of fractional ideals in $K$. As the projection $F_K \to \text{Cl}(O_K)$ factors through this map, one obtains a bijection

$$
T(Q)/T(A_f)/T(\hat{Z}) \approx \text{Cl}(O_K).
$$

By letting the orbit of the torus “act” on $K$, we may generate further ideals:

**Lemma 3.7.** For every $\rho \in R_a$ there is an ideal $a_{\rho}$ so that $\rho v_{a_{\rho}}^{-1} = v_{a_{\rho}}$ (with respect to a specific basis) and in particular $\rho T_a v_{a_{\rho}}^{-1} = T_{a_{\rho}}$.

**Proof.** Write $\rho \in R_a$ as $\rho = \gamma h$ for $\gamma \in \text{PGL}_2(Q)$ and $h \in T_a(A)$ and choose a representative $\gamma \in \text{GL}_2(Q)$. We claim that $Z^2\gamma$ is preserved under right-multiplication with $\psi_\gamma(O_K)$, which then implies that $a_{\rho} := \iota(Z^2\gamma)$ is an ideal. Let $b \in O_K$ and note that $Z^2\gamma\psi_\gamma(b) = Z^2\gamma\psi_\gamma(b)\gamma^{-1}\gamma$. By the choice of $\gamma$,

$$
\text{Mat}_2(Q) \ni \gamma\psi_\gamma(b)\gamma^{-1} = \rho\psi_\gamma(b)\rho^{-1} \in \text{Mat}_2(\mathbb{R} \times \hat{Z})
$$

Therefore, $\gamma\psi_\gamma(b)\gamma^{-1} \in \text{Mat}_2(Z)$ and $Z^2\gamma\psi_\gamma(b) \subset Z^2\gamma$. Observe that by definition of $a_{\rho}$, we have $v_{a_{\rho}} = \gamma v_{a_{\rho}}^{-1} = \rho v_{a_{\rho}}^{-1}$ in the basis $\iota(e_1\gamma), \iota(e_2\gamma)$. \hfill $\Box$

We project everything onto the $p$-adic extension

$$
\text{PGL}_2(\mathbb{Z}_p) \to \text{PGL}_2(\mathbb{R} \times Q_p).
$$

Consider a fixed odd prime $p$ and the field $K = \mathbb{Q}(\sqrt{d})$ where $d$ is a negative square-free integer with $\left(\frac{d}{p}\right) = 1$. The orbit $\text{PGL}_2(Q)T_a(A)$ in the adelic extension projects onto the disjoint union

$$
\bigcup_{\rho \in R_a} \text{PGL}_2(\mathbb{Z}_p)\rho T_a(\mathbb{R} \times Z_p),
$$

Compare to the proof of Lemma 3.7.
which is thereby also $T_a(\mathbb{Q}_p)$-invariant. Denote

$$M := \text{PO}(2) \times \left\{ \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \in \text{PGL}_2(\mathbb{Z}_p) \mid x \in \mathbb{Z}_p^\times \right\}$$

and $(v_{d,\infty}, v_{d,p}) =: v_d$ where $(v_{d,\infty}, v_{d,p})$ was defined in Equations (3.3) and (3.4). An elementary computation shows that for any ideal $b$ there is a $g_b \in \text{PGL}_2(\mathbb{R} \times \mathbb{Z}_p)$ with

$$g_b^{-1}v_b g_b = v_d.$$  

Furthermore, the choice of $g_b$ is unique up to a right factor in $M$ and we have $g_a^{-1}T_a(\mathbb{R} \times \mathbb{Z}_p)g_a = M$. Observe that for $\rho \in \mathcal{R}_a$ the element $(\rho_\infty, \rho_p)g_a =: g_{a,\rho}$ satisfies $g_{a,\rho}^{-1}v_a g_{a,\rho} = v_d$ and therefore

$$\bigcup_{\rho \in \mathcal{R}_a} \text{PGL}_2(\mathbb{Z}_p^{1/2})((\rho_\infty, \rho_p) T_a(\mathbb{R} \times \mathbb{Z}_p)) g_a = \bigcup_{\rho \in \mathcal{R}_a} \text{PGL}_2(\mathbb{Z}_p^{1/2}) g_{a,\rho} M$$

The orbit PGL$_2(a)$$_{\mathbb{Z}_p}$ associated to an ideal $b$ is independent of the choice of basis on $b$. If $b, b'$ are two equivalent ideals, then respective bases may be chosen so that $v_b = v_{b'}$ and in particular, PGL$_2(\mathbb{Z}_p^{1/2}) g_b M = PGL_2(\mathbb{Z}_p^{1/2}) g_{b'} M$. Thus, we will write the projection of orbit PGL$_2(\mathbb{Q})$$_{\mathbb{Z}_p}$($\mathbb{A}$) onto the $p$-adic extension after right multiplication with $g_a$ as

$$\mathcal{G}_d := \bigcup_{[b]} \text{PGL}_2(\mathbb{Z}_p^{1/2}) g_b M$$

where the union runs over all ideal classes by Proposition 3.5. Note that $\mathcal{G}_d$ is independent of the initial choice of ideal $a$ and is not only invariant under $M$ but also under the non-compact group

$$A := \left\{ \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \in \text{PGL}_2(\mathbb{Q}_p) \mid x \in \mathbb{Q}_p^\times \right\}$$

since $\bigcup_{\rho \in \mathcal{R}_a} \text{PGL}_2(\mathbb{Z}_p^{1/2})((\rho_\infty, \rho_p) T_a(\mathbb{R} \times \mathbb{Z}_p)$ was invariant under $T_a(\mathbb{Q}_p)$.

**Remark 3.8.** Let $\phi: K \rightarrow \mathbb{C} \cong \mathbb{R}^2$ be a Galois embedding. For any ideal $a$ and a $\mathbb{Z}$-basis $a_1, a_2$ of $a$ we may choose $g_{a,\infty}$ as

$$g_{a,\infty} = \begin{pmatrix} \phi(a_1) \\ \phi(a_2) \end{pmatrix}.$$  

The set $\mathcal{G}_d$ is naturally equipped with an $M$-invariant probability measure as follows: Let $m_M$ be the normalized Haar measure $M$ and denote by $m_x M$ the induced Haar measure on an orbit $x M$ in $\mathcal{G}_d$. The total volume of $x M$ is

$$m_x M(x M) = |\mathbb{T}_a(\mathbb{Z})|^{-1} = |O_x^\times|^{-1} =: \omega_d^{-1}$$

and in particular independent of $x$. Therefore, the total volume of $\mathcal{G}_d$ is $\frac{h_d}{\omega_d}$ where $h_d$ is the class number of $O_d$. One verifies that there are exactly two units in $O_d$ if $d > 3$. This implies the asymptotics

$$\text{vol}(\mathcal{G}_d) = \frac{h_d}{\omega_d} = |d|^{\frac{1}{2} + o(1)}$$  

by Siegel’s lower bound (see [IK04]) and Dirichlet’s class number formula. Define

$$\mu_d := \frac{1}{\text{vol}(\mathcal{G}_d)} \text{vol}$$
Proof. Let $PGL_2(\mathbb{Z})\backslash GL_2(\mathbb{R})$ and consider the traceless matrices $g \in GL_2(\mathbb{R})$ under this correspondence. The analogous statement holds over $\mathbb{Q}_p$ for more generally any field of characteristic not 2.

Lemma 3.10 (Points in $\mathcal{G}_d$ and quadratic forms). To a point $PGL_2(\mathbb{Z})\backslash GL_2(\mathbb{R})$ associate the quadratic form corresponding to the traceless matrix $g v_d g^{-1}$. This quadratic form is integral, has discriminant $d$ and is uniquely determined up to $GL_2(\mathbb{Z})$-equivalence. Furthermore, the quadratic forms associated to two points on different $M$-orbits in $\mathcal{G}_d$ are inequivalent.

Proof. Let $PGL_2(\mathbb{Z})\backslash GL_2(\mathbb{R})$ with $g \in PGL_2(\mathbb{R} \times \mathbb{Z}_p)$ and consider the traceless matrix $g v_d g^{-1}$. This matrix has integral entries: Writing $g = \gamma g_a m$ for $m \in M$
and \( \gamma \in \text{PGL}_2(\mathbb{Z}) \), we see that

\[
gv_a g^{-1} = \gamma g_a v_a g_a^{-1} \gamma^{-1} = \gamma v_a \gamma^{-1} \in \text{Mat}_2(\mathbb{Z})
\]

The discriminant of the quadratic form associated to \( gv_a g^{-1} \) is \(-\det(gv_a g^{-1}) = d\). Now let \( \text{PGL}_2(\mathbb{Z}_p) \) and \( \text{PGL}_2(\mathbb{Z}_p) \) with \( g, \bar{g} \in \text{PGL}_2(\mathbb{R} \times \mathbb{Z}_p) \) so that there exists \( \gamma \in \text{PGL}_2(\mathbb{Z}) \) with \( gv_a g^{-1} = \gamma g \bar{g} \gamma^{-1} \). By replacing \( \bar{g} \) we may assume that \( \gamma = I \). Notice that \( h := \bar{g}^{-1} g \) commutes with \( v_a \) and thus lies in \( M \). \( \square \)

**Proof of Linnik’s Theorem B assuming Theorem 3.9.** Consider a CM point

\[
x = \frac{-b + \sqrt{-d}}{2a} \in \mathbb{H}, \quad d = b^2 - 4ac.
\]

where we assume \( a \geq 0 \) for the sake of concreteness. The matrix

\[
g_x := \frac{1}{|d|^\frac{1}{2} \sqrt{2a}} \begin{pmatrix} \sqrt{-d} & -b \\ 0 & 2a \end{pmatrix} \in \text{SL}_2(\mathbb{R})
\]

yields \( x \) as \( g_x \cdot x = x \) and satisfies the equation

\[
g_x v_a \gamma g_x^{-1} = \begin{pmatrix} b & 2c \\ -2a & -b \end{pmatrix}.
\]

By Lemma 3.10 and the correspondence in \([\text{ELMV12}, \text{Section 2}]\), there is an ideal \( \mathfrak{a} \) in \( \mathbb{Q}(\sqrt{d}) \), \( k \in \text{SO}(2) \) and \( \gamma \in \text{GL}_2(\mathbb{Z}) \) so that \( \gamma g_{\mathfrak{a} \gamma} k = g_x \). This proves that \( \text{PSL}_2(\mathbb{Z}) \cdot x \in \mathcal{H}_d \). On the other hand, the equation above shows that the CM point reproduces its underlying quadratic form. As there are exactly \( h_d \text{PSL}_2(\mathbb{Z}) \)-equivalence classes of primitive integral quadratic forms there are \( h_d \text{PSL}_2(\mathbb{Z}) \)-equivalence classes of CM points, which proves the other inclusion. \( \square \)

### 3.4. Ideal classes and heights.

We keep the notation from the last subsection and derive two important estimates concerning the measures \( \mu_d \). The first roughly states that there are not too many orbits in \( \mathcal{G}_d \) “high” in the cusp. The second estimate answers the question as to how many “low lying” orbits in \( \mathcal{G}_d \) are close together (Linnik’s basic lemma). The height \( \text{ht}(x) \) of a point \( x = [\Lambda] \in X \) is defined by

\[
\frac{1}{\text{ht}(\Lambda)} = \min_{\lambda \in \Lambda \setminus \{0\}} \| \lambda \|_{\infty} \| \lambda |_{\mathbb{P} \mathbb{R}} \|_{\mathbb{P}} \frac{1}{\text{covol}(\Lambda)^\frac{1}{2}}
\]

where \( \Lambda \) is a lattice\(^7\) representing \( x \). It is straightforward to verify that the height of a point in \( X \) is equal to the height of its image under the projection to \( \text{PGL}_2(\mathbb{Z}) \). Let \( X_{\geq H} \) be the set of points in \( X \) of height bigger or equal than \( H \) and similarly define \( X_{< H} \), \( X_{\leq H} \) and \( X_{\geq H} \). Set \( d_K := \text{disc}(\mathcal{O}_K) \) where \( K = \mathbb{Q}(\sqrt{d}) \) and recall that \( |d| \leq |d_K| \leq 4|d| \).

**Proposition 3.11** (Orbits high in the cusp). Let \( \mathfrak{a} \) be an ideal in \( K = \mathbb{Q}(\sqrt{d}) \) and choose \( g_a \) as in (3.8). The following are equivalent:

(i) \( \text{PGL}_2(\mathbb{Z}_p) | g_a M \cap X_{\geq H} \) is non-empty.

(ii) There exists \( \lambda \in \mathfrak{a} \) with \( \text{Nr}(\lambda^{-1}) \leq \frac{1}{2} \sqrt{|d_K| H^{-2}} \).

\(^7\)Recall that a point in \( X \) is naturally identified with a homothety class of lattices in \((\mathbb{R} \times \mathbb{Q}_p)^2\), where a lattice is a \( \mathbb{Z}[\frac{1}{p}] \)-submodule of the form \( \mathbb{Z}[\frac{1}{p}] g \) with \( g \in \text{GL}_2(\mathbb{R} \times \mathbb{Q}_p) \).
In particular, \( G_d \) does not contain a point of height \( > |d|^{\frac{3}{4}} \). Furthermore, the number of orbits in \( G_d \), which intersect \( X_{\geq H} \), is bounded by the number of integral ideals of norm \( \leq \frac{1}{2} \sqrt{|d_K|} H^{-2} \).

Note that for any \( a \) all points in \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])g_aM \) have the same height. In particular, \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])g_aM \cap X_{\geq H} \) is either empty or equal to \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])g_aM \).

**Proof.** Let \( \phi : \mathbb{K} \to \mathbb{C} \cong \mathbb{R}^2 \) be a Galois embedding and choose \( g_{a,\infty} \) as in Remark 3.8. Now observe that \( \text{PGL}_2(\mathbb{Z})g_{a,\infty} \) has height \( \geq H \) if and only if \( \phi(a) \) contains an element of norm \( \leq \sqrt{\text{covol}(\phi(a))} H^{-1} \). The latter is equivalent to the condition that \( a \) contains an element \( \lambda \) with \( \text{Nr}(\lambda) \leq \frac{1}{2} \sqrt{|d_K|} H^{-2} \text{Nr}(a) \). In particular, \( G_d \) does not contain a point of height \( > |d|^{\frac{3}{4}} \), since the ideal \( \lambda a^{-1} \) is integral by definition of the inverse \( a^{-1} \) and integral ideals have integral norm.

Any primitive \( \lambda \in a \) as in (ii) is unique up to a sign. To any \( a \) which satisfies (i) thus corresponds the unique integral ideal \( \lambda a^{-1} \) where \( \lambda \in a \) is chosen to be primitive and as in (ii). \( \square \)

**Proposition 3.12** ("Not too much mass high in the cusp"). For all \( \varepsilon > 0 \) and \( H > 1 \) we have

\[
\mu_d(X_{\geq H}) \ll |d|^{\varepsilon} H^{-2}
\]

**Proof.** By Proposition 3.11, the number of orbits in \( G_d \) which intersect \( X_{\geq H} \) is bounded by the number of integral ideals in \( \mathbb{K} = \mathbb{Q}(\sqrt{d}) \) of norm \( \leq \frac{1}{2} \sqrt{|d_K|} H^{-2} \).

Counting lattice points shows that the latter is

\[
\ll \left( \frac{h_d}{\sqrt{|d_K|}} \sqrt{|d_K|} H^{-2} \right)^{1+\varepsilon} \ll (h_d H^{-2})^{1+\varepsilon}.
\]

The same bound holds for the volume of \( G_d \cap X_{\geq H} \), as all \( M \)-orbits have length \( \ll 1 \). This yields the right estimate after normalization by the total volume (see Equation (3.9)). \( \square \)

We now turn to the following analog of Proposition 2.12.

**Proposition 3.13** (Linnik’s basic lemma). For any \( \delta > 0 \) with \( |d|^{-\frac{1}{4}} \leq \delta \leq \frac{1}{2} H^{-2} \) and any \( \varepsilon > 0 \) we have

\[
\mu_d^2(\{(x,y) \in (X_{\leq H})^2 \mid d_X(x,y) \leq \delta\}) \ll \varepsilon H^4 \delta^3 |d|^\varepsilon.
\]

**Proof.** Set \( S := \{(x,y) \in (X_{\leq H} \cap \mathcal{G}_d)^2 \mid d_X(x,y) \leq \delta\} \), let \( (x_1, x_2) \in S \) and choose \( g_1, g_2 \in \text{PGL}_2(\mathbb{R} \times \mathbb{Z}_p) \) with \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])g_i = x_i \) for \( i = 1, 2 \) and \( d(g_1, g_2) \leq \delta \). Then \( ||g_i|| \ll H \). We attach to both points the integral quadratic form \( q_i \) constructed in Lemma 3.10 and distinguish as in the proof of Proposition 2.12 two cases:

**Case 1:** Assume that \( q_1, q_2 \) are equivalent or in other words that \( x_1, x_2 \) lie on the same \( M \)-orbit in \( \mathcal{G}_d \). As the volume of a \( \delta \)-ball in \( M \) is \( \ll \delta \), the set of such \( x_1, x_2 \) has total volume \( \ll \varepsilon \delta |d|^{\frac{1}{4}+\varepsilon} \) which yields a contribution of \( \ll \varepsilon \delta |d|^{-\frac{1}{4}} |d|^\varepsilon \leq \delta^3 |d|^\varepsilon \) to the measure of \( S \) in this case.

**Case 2:** Assume that \( q_1, q_2 \) are inequivalent and write \( q_i = a_i x_i^2 + b_i x_i y_i + c_i y_i^2 \) for \( i = 1, 2 \). The bound \( ||g_i|| \ll H \) yields \( \max(|a_i|, |b_i|, |c_i|) \ll |d|^{\frac{1}{2}} H^2 \). On the other hand, the bound \( d(g_1, g_2) \leq \delta \) implies

\[
\max(|a_1 - a_2|, |b_1 - b_2|, |c_1 - c_2|) \ll |d|^{\frac{1}{4}} H^2 \delta.
\]
Consider the integral quadratic form
\[ Q(x, y) = \text{disc} \left( x(a_1, b_1, c_1) + y(a_2, b_2, c_2) \right) = dx^2 + \ell xy + dy^2 \]
for some \( \ell \). The bound on the difference between the coefficients of \( q_1 \) and \( q_2 \) yields
\[ |2d - \ell| = |Q(1, -1)| \ll |d|H^4\delta^2. \]
and also \( |\ell| \ll |d| + |d|H^4\delta^2 \ll |d| \). We claim that \( Q \) is non-degenerate. Indeed, if we had \( \ell = \pm 2d \) this would contradict the assumption that \( d < 0 \) as
\[ d(a_2 \mp a_1)^2 = Q(a_2, -a_1) = \text{disc} \left( a_2(a_1, b_1, c_1) - a_1(a_2, b_2, c_2) \right) = (a_2b_1 - a_1b_2)^2. \]
Let \( N_{\ell,d} \) be the number of inequivalent ways to represent the binary quadratic form \( dx^2 + \ell xy + dy^2 \) by the ternary quadratic form disc up to \( \text{SO}_{\text{disc}}(\mathbb{Z}) \)-equivalence. By Theorem 2.14
\[ N_{\ell,d} \ll \epsilon f \max(|d|, |\ell|) \epsilon \ll \epsilon |d|^\epsilon \]
where \( f^2 = \text{gcd}(d, \ell) \) is the greatest common square divisor. By commensurability we may replace \( \text{SO}_{\text{disc}}(\mathbb{Z}) \) by \( \text{PGL}_2(\mathbb{Z}) \) above. Let
\[ \text{PGL}_2(\mathbb{Z})(q_1^{(1)}, q_2^{(1)}), ... \text{PGL}_2(\mathbb{Z})(q_1^{(k)}, q_2^{(k)}) \]
be a complete list of pairs of inequivalent quadratic forms, where \( q_i^{(j)} \) is obtained from \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])q_i^{(j)} \) as in the beginning. The number \( k \) satisfies the bound
\[ k \leq \sum_{\ell : |\ell| \leq L, \ell \neq \pm 2d} N_{\ell,d} = \sum_{f^2 | d} \sum_{\ell' : |\ell'| \leq L, \ell' \neq 0, 4d} N_{2d-\ell',d} \leq \sum_{f^2 | d} \sum_{\ell' : |\ell'| \leq L, \ell' \neq 0, 4d} \frac{L}{f^2} \ll \epsilon |d|^{1+2\epsilon}H^4\delta^2 \]
for \( L \ll |d|H^4\delta^2 \) where the implicit constant is as in Equation (3.10). If now \((x_1, x_2), (g_1, g_2)\) and \((q_1, q_2)\) are as in the beginning of the proof, there is some \( j \) and \( \gamma \in \text{PGL}_2(\mathbb{Z}) \) so that \( (\gamma q_1, \gamma q_2) = (q_1^{(j)}, q_2^{(j)}) \). Therefore, \( x_i \) lies on the same \( M \)-orbit as \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])q_i^{(j)} \) for \( i = 1, 2 \). For fixed \( j \), the set of \( \delta \)-close pairs \((x_1, x_2)\) lying on the orbit \( \text{PGL}_2(\mathbb{Z}[\frac{1}{p}])q_i^{(j)} \) has measure \( \ll \delta \). Thus, the total volume is
\[ \ll \delta k \ll \epsilon \delta |d|^{1+2\epsilon}H^4\delta^2 = |d|^{1+2\epsilon}H^4\delta^3 \]
before normalization. After normalization, we obtain that the contribution to the measure of \( S \) is \( \ll \epsilon |d|^{3\epsilon}H^4\delta^3 \) in this case.

### 3.5. Maximal Entropy.
In this subsection, we prove Theorem 3.9 along the lines of [ELMV12] using the estimates derived in the last subsection. Consider the map \( T: X \to X, x \mapsto xa \), where
\[ a := \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \in \text{PGL}_2(\mathbb{Q}_p). \]
Let \( d_i \) be a sequence of negative, square-free integers satisfying Linnik’s condition for \( p \) and write for simplicity \( \mu_i \) for the measure \( \mu_{d_i} \) defined in Section 3.2. Denote \( \delta_i = |d_i|^{-\frac{1}{2}} \) so that Proposition 3.13 applies for any height \( \ll \delta_i^{-1/2} = |d_i|^{1/8} \). By restricting to a subsequence, we may assume that \( \mu_i \) converges to some finite measure \( \mu \) on \( X \) with total mass at most 1.
Proposition 3.14 (Maximal Entropy). The weak*-limit $\mu$ of the sequence $(\mu_n)$ is a probability measure and has maximal entropy with respect to $T$, $h_\mu(T) = \log(p)$.

Proof of Theorem 3.9 using Proposition 3.14. Naturally, $\mu$ is invariant under the diagonal subgroup $A$. However, by Theorem 2.11 the measure $\mu$ is also invariant under the horospherical subgroups $G_a^-$ and $G_a^+$ where $G_a^-$ and $G_a^+$ were computed in Example 2.9. Therefore, $\mu$ is $\text{PGL}_2(\mathbb{Q}_p)$-invariant as the subgroup generated by $A, G_a^-$ and $G_a^+$ is $\text{PGL}_2(\mathbb{Q}_p)$. Since $\text{PGL}_2(\mathbb{Z}[1/2])$ is dense in $\text{PGL}_2(\mathbb{R})$, one shows as in Section 2.6 that $\mu$ is also $\text{PGL}_2(\mathbb{R})$-invariant. Thus, $\mu$ is the normalized Haar measure on $X$.

We will use the following proposition and postpone the proof to the next subsection.

Proposition 3.15. Let $H > 1$ be a height. For $N \geq 1$ and a set of times $V$ in $[-N,N]$ let
\[ Z(V) := \{ x \in X \mid T^{\pm N}(x) \in X_{\leq H}, \forall n \in [-N,N]: T^n(x) \in X_{\geq H} \iff n \in V \}. \]
Then $Z(V)$ can be covered by $\ll_H p^{2N-H\frac{1}{2}|V|}$ Bowen $N$-balls and is non-empty for $\ll_H e^{2\log(p)\log(\log(H))cN}$ many subsets $V \subset [-N,N]$ where $c$ is an absolute constant.

In this context, a (two-sided) Bowen $N$-ball in $X$ will always be a set of the kind $xB_N$ where $x$ is a point in $X$ and
\[ B_N = \bigcap_{n=-N}^{N} a^{-n}B_n a^n \]
is a Bowen ball in the group $\text{PGL}_2(\mathbb{R} \times \mathbb{Q}_p)$. The statement in Proposition 3.15 is independent of the choice of radius $\eta > 0$: Given two radii $0 < \eta' < \eta$, the ball $B_\eta$ in $\text{PGL}_2(\mathbb{R} \times \mathbb{Q}_p)$ is covered by $\ll_{\eta,\eta'} 1$ shifts of the ball $B_{\eta'}$. For the purposes of this subsection, one fixed choice of radius $\eta > 0$ usually suffices.

Lemma 3.16. For all large enough heights $H$
\[ \mu(X_{< H}) \geq 1 - 2\log(p)\frac{\log(\log(H))}{\log(H)}. \]
In particular, $\mu$ is a probability measure.

The proof is up to minor details the proof of Lemma 4.4 in [ELMV12] and uses the geometric interpretation provided by the Hecke tree (see Section 3.6) – we will omit it here. The same conclusion applies to the following lemma.

Lemma 3.17. For any height $H > 1$ there is a finite partition $\mathcal{P}$ of $X$ such that for every $0 < \kappa < 1$ and every $N$ there is a measurable subset $X' \subset T^{-N}X_{< H}$ satisfying the following conditions.
\begin{enumerate}
\item $\nu(X') \geq 1 - 2\nu(X_{\geq H})\kappa^{-1}$ for any $T$-invariant probability measure $\nu$.
\item $X'$ is a union of partition elements $S_1, \ldots, S_\ell \in \mathcal{P}_N$, each of which is covered by at most $p^{2N+1}$ Bowen $(N,\eta)$-balls. Here, $\eta$ is assumed to be smaller than $1/p$ times an injectivity radius on $X_{< H}$.
\end{enumerate}
Fixing an invariant measure $\nu$ with $\nu(\partial X_{\geq H}) = 0$ the partition $\mathcal{P}$ may be constructed so that all partition elements have boundaries of measure zero.
Proof of Proposition 3.14. Let $H > 1$ be a fixed height so that the boundary of $X_{\geq H}$ has $\mu$-measure zero and let $\mathcal{P}$ be the partition from Lemma 3.17. Define $\kappa = \mu(X_{\geq H})^{\frac{1}{2}}$, $N_i = [-\log_p(\delta_i)]$ and choose $X_i \subset X$ according to Lemma 3.17.

We define a new partition $\mathcal{Q}_i$, which is finer than $\mathcal{P}_N^{-N_i}$, by splitting all the $S$ in $\mathcal{P}_N^{-N_i}$, which are contained in $X_i$, into at most $p^{N_i}$ sets, which are contained in Bowen $N_i$-balls. As $\mathcal{Q}_i$ is finer than $\mathcal{P}_N^{-N_i}$, we have

$$|H_{\mu_\iota}(\mathcal{Q}_i) - H_{\mu_\iota}(\mathcal{P}_N^{-N_i})| = H_{\mu_\iota}(\mathcal{Q}_i | \mathcal{P}_N^{-N_i}) = \sum_{S \in \mathcal{P}_N^{-N_i}, S \subset X_i} \mu(S)H_{\mu_\iota|S}(\mathcal{Q}_i) \leq \kappa(2N_i + 1) \log(p).$$

Claim. $H_{\mu_\iota}(\mathcal{Q}_i) \geq (1 - 2\kappa^{-1}\mu(X_{\geq H}))(2 + 6\epsilon) \log(p) N_i$

The claim implies the proposition as follows: By the claim and the computation above the claim

$$H_{\mu_\iota}(\mathcal{P}_N^{-N_i}) \geq (1 - 2\kappa^{-1}\mu(X_{\geq H}))(2 + 6\epsilon) \log(p) N_i - \kappa(2N_i + 1) \log(p)$$

Proceeding as in the proof of Theorem 2.7, we obtain that for $\epsilon > 0$ and all large enough $N_0$

$$H_{\mu_\iota}(\mathcal{P}_N^{-N_0}) \geq (1 - 2\kappa^{-1}\mu(X_{\geq H}))(2 + 6\epsilon) \log(p) N_0 - \kappa(2N_0 + 1) \log(p) - \epsilon N_0.$$ 

By Lemma 3.17, we may assume that boundaries of all partition elements in $\mathcal{P}$ are $\mu$-null sets. Thus, taking the limit as $i \to \infty$

$$H_{\mu_\iota}(\mathcal{P}_N^{-N_0}) \geq (1 - 2\kappa)(2 + 6\epsilon) \log(p) N_0 - \kappa(2N_0 + 1) \log(p) - \epsilon N_0.$$ 

Dividing by $2N_0 + 1$ and letting $N_0$ go to infinity

$$h_{\mu}(T) \geq (1 - 2\mu(X_{\geq H})\frac{1}{2})(1 + 3\epsilon) \log(p) - \mu(X_{\geq H})\frac{1}{2} \log(p) - \frac{\epsilon}{2}.$$ 

Taking the limit $H \to \infty$ and $\epsilon \to 0$, we have $\mu(X_{\geq H}) \to 0$ and thus $h_{\mu}(T) \geq \log(p)$ as desired.

To the proof of the claim: The entropy of $\mathcal{Q}_i$ satisfies

$$H_{\mu_\iota|X_i}(\mathcal{Q}_i) \geq H_{\mu_\iota}(\mathcal{Q}_i | \{X_i, X \setminus X_i\}) \geq \mu_i(X_i)H_{\mu_\iota|X_i}(\mathcal{Q}_i).$$

The right hand side is bounded from below by

$$H_{\mu_\iota|X_i}(\mathcal{Q}_i) \geq -\log \left( \sum_{S \in \mathcal{Q}_i, S \subset X_i} \frac{\mu_i(S)^2}{\mu_i(X_i)^2} \right) = 2\log(\mu_i(X_i)) - \log \left( \sum_{S \in \mathcal{Q}_i, S \subset X_i} \mu_i(S)^2 \right).$$

As any atom of $\mathcal{Q}_i$, which lies in $X_i$, is contained in a Bowen $N_i$-ball we obtain

$$\bigcup_{S \in \mathcal{Q}_i, S \subset X_i} S \times S \subset \bigcup_{j=1}^{k} \{ (x, y a_j) \mid d(x, y) < \eta p^{-N_i} \}$$

for $k \ll p^{N_i}$ and $a_1, \ldots, a_k \in A$. By Linnik’s basic lemma (Proposition 3.13)

$$\sum_{S \in \mathcal{Q}_i, S \subset X_i} \mu_i(S)^2 \ll \epsilon p^{-(3 - 5\epsilon)N_i} p^{N_i} = p^{(-2 + 5\epsilon)N_i}.$$
for all large enough \(i\). Let \(C_e\) by the implicit constant. Overall, we obtain

\[
H_{\mu_i|X_i}(Q_i) \geq 2\mu_i(X_i) \log(\mu_i(X_i)) - \mu_i(X_i) \log(p)(-2 + 5\varepsilon)N_i - \mu_i(X_i) \log(C_e).
\]

Observe that only the middle term is unbounded as \(\mu_i(X_i)\) is bounded from below. Thus, for \(i\) large enough

\[
H_{\mu_i|X_i}(Q_i) \geq \mu_i(X_i) \log(p)(2 - 6\varepsilon)N_i \geq (1 - 2\kappa^{-1}\mu_i(X_{\geq H})) \log(p)(2 - 6\varepsilon)N_i.
\]

\[\square\]

3.6. Visiting the cusp. In this subsection, we provide a proof of Proposition 3.15 using the geometric picture supplied by the Hecke tree and adapting [ELMV12] correspondingly. Notice that under the projection

\[\pi : X \rightarrow Y := \text{PGL}_2(\mathbb{Z})/\text{PGL}_2(\mathbb{R})\]

a \(\text{PGL}_2(\mathbb{Q}_p)\)-orbit gets mapped to a set isomorphic to \(\text{PGL}_2(\mathbb{Q}_p)/\text{PGL}_2(\mathbb{Z}_p)\). This follows from the fact that the action of \(\text{PGL}_2(\mathbb{Q}_p)\) on \(X\) has trivial stabilizers. The quotient \(\text{PGL}_2(\mathbb{Q}_p)/\text{PGL}_2(\mathbb{Z}_p)\) is equipped with the structure of a \((p+1)\)-regular tree; we refer to Section 3.2 in Einsiedler and Ward [EW10] for the details. Given

\[\mathcal{N} := \left\{ \left( \begin{array}{c} 1 \\ 0 \\ p \end{array} \right), \left( \begin{array}{c} p \\ 0 \\ 1 \end{array} \right), \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right), \ldots, \left( \begin{array}{c} p \\ 0 \\ 1 \end{array} \right) \right\}\]

we declare the neighbours of \(g \text{PGL}_2(\mathbb{Z}_p)\) to be the points \(\{gh \text{PGL}_2(\mathbb{Z}_p) \mid h \in \mathcal{N}\}\). As is verified in Proposition 3.15 in [EW10], this really imposes the structure of a \((p+1)\)-regular tree on \(\text{PGL}_2(\mathbb{Q}_p)/\text{PGL}_2(\mathbb{Z}_p)\). Given a point \(y = \text{PGL}_2(\mathbb{Z})y \in Y\) the image of \(\text{PGL}_2(\mathbb{Z})/\text{PGL}_2(\mathbb{Z}_p)\) under \(\pi\) is called the embedded Hecke tree through \(y\) and is equipped with a tree-structure by identification with the tree \(\text{PGL}_2(\mathbb{Q}_p)/\text{PGL}_2(\mathbb{Z}_p)\).

Given a point \(x \in X\), the point \(\pi(T(x))\) is always a neighbour of \(\pi(x)\) in the Hecke tree by definition of \(\mathcal{N}\). Taking the right quotient by \(\text{PO}(2)\) of \(Y\), the neighbours of \(z = \pi(x)\) \(\text{PO}(2) \in \text{PSL}_2(\mathbb{Z})/\mathbb{H}\) are exactly

\[\text{PSL}_2(\mathbb{Z}) \left\{ \frac{z}{p}, \frac{z+1}{p}, \ldots, \frac{z+p-1}{p} \right\} \]

If the height of \(z\) is large enough (for instance \(\text{ht}(z) \geq p\)), then exactly one of its neighbours (that is, \(\frac{z}{p}\)) is “above” \(z\) (precisely \(\text{ht}(\frac{z}{p}) = p\text{ht}(z)\)) and the other neighbours (that is, \(\frac{z+1}{p}, \frac{z+2}{p}, \ldots, \frac{z+p-1}{p}\)) are “below” \(z\) (precisely of height \(\text{ht}(z)/p\)). Furthermore, the point \(T^2(x)\) cannot be equal to \(x\) due to the tree structure. We will use this observation as follows.

\textbf{Remark 3.18.} Let \(H > 1\) be large enough, let \(x \in X\) be a point with height \(\geq H\) and suppose that the height of \(T(x)\) is smaller than \(x\). Then the point \(T^2(x)\) cannot be “above” \(T(x)\) as it would equal to \(x\) in that case and is therefore “below” \(T(x)\). The only condition we need to impose here, is that all points are above height 1. In other words, the \(T\)-orbit of \(x\) moves downwards for at least \(\lfloor \log_p(H) \rfloor\) time steps (as \(\text{ht}(T^k(x)) = \text{ht}(T^{k-1}(x))/p\) for these \(k\)) until it “crosses” height one. The minimum time to reach height \(H\) from height one is also at least \(\lfloor \log_p(H) \rfloor\).

For the proof of Proposition 3.15 we proceed exactly as in Section 5.1 of [ELMV12] and begin with the second assertion as the proof only depends on the remark above.
Proof of the second assertion in Proposition 3.15. Consider the partition
\[ \mathcal{P}_{H,N} = \bigvee_{n=-N}^{N} T^{-n}([X_{<H}, X_{\geq H}]). \]
Every \( V \subset [-N,N] \) with \( Z(V) \neq \emptyset \) defines an atom of \( \mathcal{P}_{H,N} \) and thus it suffices to prove that \( \mathcal{P}_{H,N} \) contains \( \ll_H e^{\frac{2 \log(\log(H)) N}{\log(2)}} \) atoms. Consider first an atom of \( \mathcal{P}_{H,\lfloor \log_p(H) \rfloor} \) and a point \( x \) in it. If for some \( n \in \mathbb{Z} \) with \( |n| \leq \lfloor \log_p(H) \rfloor \) the point \( T^n(x) \) is above height \( H \) and \( T^{n+1}(x) \) is below height \( H \), then the orbit of \( x \) stays below height \( H \) for all times \( n \) in this interval by Remark 3.18. Thus, every point can leave \( X_{<H} \) at most once. In particular, the time interval contains at most one stretch of times for which the orbit of the point can be above \( H \). Therefore, the starting and the end point of that time interval uniquely determine an atom in \( \mathcal{P}_{H,\lfloor \log_p(H) \rfloor} \) and in particular there are \( \leq (2\lfloor \log_p(H) \rfloor + 1)^2 \) many atoms in \( \mathcal{P}_{H,\lfloor \log_p(H) \rfloor} \). The partition \( T^{-N}(\mathcal{P}_{H,N}) \) is coarser than a refinement over \( \frac{2N+1}{2 \lfloor \log_p(H) \rfloor + 1} \) many partitions of the kind \( T^{-j}(\mathcal{P}_{H,\lfloor \log_p(H) \rfloor}) \). Hence \( T^{-N}(\mathcal{P}_{H,N}) \) (and thus also \( \mathcal{P}_{H,N} \)) contains at most
\[ \left( (2\lfloor \log_p(H) \rfloor + 1)^2 \right) \frac{2N+1}{2 \lfloor \log_p(H) \rfloor + 1} \ll_H e^{4N \frac{\log(2\lfloor \log_p(H) \rfloor + 1)}{2 \lfloor \log_p(H) \rfloor + 1}} \leq e^{4N \frac{\log(2\lfloor \log_p(H) \rfloor)}{\log_p(H)}} \leq e^{2 \log(p) N \frac{\log(\log(H)) + \log(2) - \log(\log(p))}{\log_p(H)}} \]
many atoms.

The main geometric idea for the second assertion of Proposition 3.15 is the following.

Remark 3.19 (Moving up half of the time). Let \( x \in X \) be a point for which \( T^n(x) \) is below height \( H > 1 \) at some times \( n = N, N' \) for \( N < N' \) and for which \( T^n(x) \) is above height \( H \) for all times \( n \) with \( N < n < N' \). Then the orbit of \( x \) is “moving upwards” (the first) 50% of the time (in \([N, N']\)). This is a consequence of the fact that the “speed of moving up or down” is always \( p \).

Writing \( x = \text{PGL}_2(\mathbb{Z}_{\frac{1}{p}})((g_{\infty}, g_p) \in \text{PGL}_2(\mathbb{R} \times \mathbb{Z}_p) \) this means that
\[ T^n(x) = \text{PGL}_2(\mathbb{Z}_{\frac{1}{p}})((a^{-k}g_{\infty}, a^{-k}g_p a^k) \]
projects to the point \( p^k g_{\infty}.i = a^{-k}g_{\infty}.i \) on the modular surface and therefore \( a^{-k}g_p a^k \in \text{PGL}_2(\mathbb{R} \times \mathbb{Z}_p) \) for all \( k \) in the first half of the interval \([N, N']\).

Proof of the first assertion in Proposition 3.15. For simplicity we denote the horospherical subgroups \( G_a^+, G_a^- \) associated to \( a \) by \( U^+, U^- < \text{PGL}_2(\mathbb{Q}_p) \) respectively (see Example 2.9) and by \( U^0 := \text{PGL}_2(\mathbb{R}) \times G_a^0 \).

It suffices to show that given a set of times \( V \subset [0, N] \) and an open neighbourhood \( O \) of a point \( x_0 \in X \) of the form \( x_0 B_{\eta/2}^{U^+} B_{\eta/2}^{U^-} \) the set
\[ Z^+_O(V) = \{ x \in O \cap T^{-N}X_{<H} | \forall n \in [0, N] : T^n(x) \in X_{\geq H} \iff n \in V \} \]
can be covered by \( \ll p^{N-\frac{1}{2}|V|} \) forward Bowen balls. This follows by compactness of \( X_{\geq H} \). We partition the interval \([0, N]\) as follows. Decompose \( V \) into maximal intervals containing consecutive times in \( V \). By Remark 3.18 two such intervals in \( V \) have to be separated at least by \( 2\lfloor \log_p(H) \rfloor \). Therefore we may thicken the
above intervals in $V$ on both sides by $\lfloor \log_2(H) \rfloor$ to, obtain disjoint intervals $I_1, \ldots, I_k$ covering $V$. Note that

$$|I_1| + \ldots + |I_k| = 2k\lfloor \log_2(H) \rfloor + |V|.$$ 

By slightly enlarging the interval $[0, N]$ if necessary we may assume that $I_1, \ldots, I_k$ are contained in $[0, N]$. This does not effect the estimate we aim for as the difference in $N$ depends on $H$ only and can thus by taken into the multiplicative constant.

We fill the gaps between the intervals $I_i$ with maximal intervals $J_j$, $1 \leq j \leq \ell$ so that $[0, N] = I_1 \cup \ldots \cup I_k \cup J_1 \cup \ldots \cup J_{\ell}$ and prove the following claim by induction:

**Claim.** For any $K \leq N$ with $[0, K] = I_1 \cup \ldots \cup I_i \cup J_1 \cup \ldots \cup J_{\ell}$ the set $Z_O^+(V)$ can be covered by

$$\begin{equation}
\tag{3.11}
\text{preimages under } T^K \text{ of sets of the form}
\end{equation}$$

where $u^+ \in U^+$.

In particular, for $K = N$ this yields that $Z^+_O(V)$ is covered by

$$\leq p^{|J_1|+\ldots+|J_{\ell}|+k\lfloor \log_2(H) \rfloor + \frac{1}{2}(|I_1|+\ldots+|I_k|) = p^{N+k(\lfloor \log_2(H) \rfloor - \frac{1}{2}(|I_1|+\ldots+|I_k|) = p^{N-\frac{1}{2}|V|}
$$

sets of the form

$$T^{-N}(T^N(x_0)u^+B_{\eta/2}^{-N}B_{\eta/2} U^0) = x_0(a^N u^+ a^N B_{\eta/2}^{-N} B_{\eta/2} U^0)$$

contained in a Bowen $N$-ball. Thus the claim implies the proposition.

Assume that the claim holds for $K \leq N$. We distinguish two cases:

**Case 1:** Suppose that $[0, K] = [K+1, K+S]$. Taking a set $T^K(x_0)u^+B_{\eta/2}^{-N}B_{\eta/2} U^0 a^K$ obtained in the previous step, its image under $T^S$ splits into $p^S = p^{J_{\ell+1}}$ sets of the form (3.11) by properties of $U^+$, thus proving the claim.

**Case 2:** Suppose that $[0, K]$ is followed by $I_{i+1} = [K+1, K+S]$. Let

$$R_K := T^K(x_0)u^+B_{\eta/2}^{-N}B_{\eta/2} U^0 a^K$$

be a set of the kind (3.11) obtained in the previous step. As in the last case, we may split the image of $R_K$ into $\leq p^S$ sets of the kind (3.11). In this case, we claim that we may discard some of these sets as we are only interested in points in $y \in R_K$, which satisfy

$$T^n(y) \in X_{\geq H} \iff K + n \in V$$

for $1 \leq n \leq S$. Let $y_1, y_2 \in R_K \cap T^K(Z_O^+(V))$. Then $y_2 \in y_1B_{\eta/2}^{-N}B_{\eta/2} U^0 a^K$, say $y_2 = y_1(g_{\infty}, h^+ b)$. We claim that

$$h^+ \in B_{p^{-S/2}} U^+$$

By definition of $I_{i+1}$, the points $y_1, y_2$ satisfy

$$\text{ht}(y_1), \text{ht}(y_2), \ldots, \text{ht}(y_1a^{\lfloor \log_2(H) \rfloor}) < H$$

$$\text{ht}(y_1a^{\lfloor \log_2(H) \rfloor + 1}), \ldots, \text{ht}(y_1a^{S - \lfloor \log_2(H) \rfloor}) \geq H$$

$$\text{ht}(y_1a^{S - \lfloor \log_2(H) \rfloor + 1}), \ldots, \text{ht}(y_1a^S) < H$$
In particular, \( y_1 \) and \( y_2 \) move upwards during the first \( S/2 \) time steps by Remark 3.19 and therefore \( a^{-j}h+^j \in \text{PGL}_2(\mathbb{Z}_p) \) for all \( j \in [0, S/2] \) also by Remark 3.19. Conjugation by \( a^{-1} \) stretches \( h+ \) by a factor of \( p \); if the size of \( h+ \) is \( \leq 1 \) after \( S/2 \) conjugations, we must have had \( h+ \in B_{p^{-S/2}}^{U+} \) initially. This concludes the claim made in (3.12).

If \( R_K \cap T^K(Z^*_O(V)) \) is empty, there is nothing to do. Otherwise, choose a point of reference \( y \in R_K \cap T^K(Z^*_O(V)) \). The claim above implies that \( R_K \cap T^K(Z^*_O(V)) \) is contained in \( yB_{p^{-S/2}}^{U+}a^{-K}B_{\eta/2}^{U+}a^K \). The image of this set under \( T^S \) is covered by \( \ll p^{-S/2}p^S \) sets of the form (3.11) for \( K+S \) as the ball \( B_{\eta/2}^{U+} \) is a disjoint union of \( \ll p^{S/2} \) translates of the ball \( B_{\eta/2}^{U+} \).

**References**

[AES16] M. Aka, M. Einsiedler, and U. Shapira, Integer points on spheres and their orthogonal lattices, *Invent. Math.* **206** (2016), no. 2, 379–396.

[AR65] L.M. Abramov and V.A. Rokhlin, The entropy of a skew product of measure-preserving transformations, *Amer. Math. Soc. Transl.* **48** (1965), 225–265.

[Cas78] J.W.S Cassels, *Rational Quadratic Forms*, London Mathematical Society Monographs, vol. 13, Academic Press Inc., 1978.

[Duk88] W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms, *Invent. Math.* **92** (1988), no. 1, 73–90.

[EL07] M. Einsiedler and E. Lindenstrauss, *Diagonal actions on locally homogeneous spaces*, Homogeneous flows, moduli spaces and arithmetic, Clay Mathematics Proceedings, vol. 10, 2007, pp. 168–241.

[ELMV12] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, The distribution of closed geodesics on the modular surface and Duke’s theorem, *Enseign. Math.* **58** (2012), 249–313.

[EMV13] J.S. Ellenberg, P. Michel, and A. Venkatesh, Linnik’s ergodic method and the distribution of integer points on spheres, *Tata Inst. Fundam. Res. Stud.* **22** (2013), 119–185.

[EW10] M. Einsiedler and T. Ward, *Arithmetic quantum unique ergodicity on \( \Gamma \backslash \mathbb{H} \)*, Arizona Winter School, 2010.

[Gau86] C. F. Gauss, *Disquisitiones arithmeticae*, Springer-Verlag, 1986, Translated and with a preface by Arthur A. Clarke, Revised by William C. Waterhouse, Cornelius Grietzer and A.W.Grootendorst and with a preface by Waterhouse.

[IK04] H. Iwaniec and E. Kowalski, *Analytic number theory*, Colloquium Publications, vol. 53, American Mathematical Society, 2004.

[Lin68] Yu. V. Linnik, *Ergodic properties of algebraic fields*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 45, Springer-Verlag, New York, 1968, Translated from the Russian by M.S. Keane.

[MT94] G.A. Margulis and G. Tomanov, Invariant measures for actions of unipotent groups over local fields on homogeneous spaces, *Invent. Math.* **116** (1994), no. 1, 347–392.

[PR04] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press, Inc., 1994, Translated from the 1991 Russian original by R. Rowen.

[Rüh16] R. Rühr, Effectivity of uniqueness of the maximal entropy measure on \( p \)-adic homogeneous spaces, *Ergod. Theory Dyn. Syst.* **36** (2016), no. 6, 1972–1988.

[Ser73] J.P. Serre, *A course in arithmetic*, Graduate Texts in Mathematics, vol. 7, Springer, 1973.

[Ser92] J.-P. Serre, *Lie algebras and Lie groups*, second ed., Lecture Notes in Mathematics, Springer, 1992, 1964 Lectures given at Harvard University.

[Ven31] B.A. Venkov, Über die Klassenzahl positiver binärer quadratischer Formen, *Math. Z.* **33** (1931), 350–374.

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

**E-mail address:** andreas.wieser@math.ethz.ch