IDEALS, PERIODIC ORBITS AND DUKE'S THEOREM

CONSTANTIN KOGLER

Abstract. In these expository notes, we present the connection between (for example fractional) ideals in real quadratic number fields and periodic orbits for the group of diagonal matrices on the space of unimodular lattices, or equivalently closed geodesics on the modular surface. Moreover, we state Duke's Theorem. Towards the ergodic theoretic proof [ELMV12] of Duke's Theorem by Einsiedler, Lindenstrauss, Michel and Venkatesh, we introduce the height of lattices and relate it to the norm of ideals.

Contents

1. Introduction

Fix a positive non-square integer d and write $K = \mathbb{Q}(\sqrt{d})$. Throughout these notes we only consider so called **discriminants**, i.e. integers of the form

$$d \equiv 0, 1 \mod 4.$$

The central aim of these notes is to relate algebraic information contained in K to interesting curves in the three dimensional space $X = \mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})$. More precisely, the latter curves will be **orbits** of the diagonal subgroup $A < \mathrm{PGL}_2(\mathbb{R})$, i.e. an A-orbit in X is a set of form

$$x.A = \{xa : a \in A\} \subset X$$

for some $x \in X$. As we only consider A-orbits in this text, we will sometimes drop the A and just refer to A-orbits as **orbits**.

In Section 6, a finite collection of A-orbits, which we denote $\mathcal{G}_d$, will be constructed. The orbits $\mathcal{G}_d$ will come from arithmetical information contained in K, for example in some cases fractional ideals (see Section 5 for

the definition of the term fractional ideal). We are now in a suitable position to state Duke's Theorem in a vague form.

Theorem 1.1. *(Duke's Theorem, vague form, see Theorem 3.2) As $d \to \infty$ among the non-square discriminants, the sets $\mathcal{G}_d$ become uniformly distributed.*

In Section 3 we will make more precise what is meant by $\mathcal{G}_d$ becoming *uniformly distributed*. This will rely upon defining natural measures on X as well as on the orbits in $\mathcal{G}_d$.

We next comment on the history of this result. Already around 1950 Linnik [Lin68] proved Duke's Theorem assuming certain congruence conditions. In 1988, William Duke [Duk88] proved Theorem 1.1 unconditionally via techniques from analytic number theory, especially the theory of modular forms. Finally, in 2012 [ELMV12] Einsiedler, Lindenstrauss, Michel and Venkatesh gave an proof of Duke's Theorem giving a modern dynamical interpretation of Linnik's ideas.

1.1. **Notations.** We denote

$$\begin{aligned}
\mathrm{GL}_2(\mathbb{R}) &= \{g \in \mathrm{M}_2(\mathbb{R}) \,:\, \det(g) \neq 0\} \\
\mathrm{GL}_2(\mathbb{Z}) &= \{g \in \mathrm{M}_2(\mathbb{Z}) \,:\, \det(g) = \pm 1\} \\
\mathrm{SL}_2(\mathbb{R}) &= \{g \in \mathrm{M}_2(\mathbb{R}) \,:\, \det(g) = 1\} \\
\mathrm{SL}_2(\mathbb{Z}) &= \{g \in \mathrm{M}_2(\mathbb{Z}) \,:\, \det(g) = 1\} \\
\mathrm{PGL}_2(\mathbb{R}) &= \mathrm{GL}_2(\mathbb{R})/\{\lambda \cdot \mathrm{Id}_2 \,:\, \lambda \in \mathbb{R}_{\neq 0}\} \\
\mathrm{PGL}_2(\mathbb{Z}) &= \mathrm{GL}_2(\mathbb{Z})/\{\pm \mathrm{Id}_2\} \\
\mathrm{PSL}_2(\mathbb{R}) &= \mathrm{SL}_2(\mathbb{R})/\{\pm I_2\} \\
\mathrm{PSL}_2(\mathbb{Z}) &= \mathrm{SL}_2(\mathbb{Z})/\{\pm I_2\}
\end{aligned}$$

and write

$$X = \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}) = \mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R}) = \mathrm{PGL}_2(\mathbb{Z})\backslash \mathrm{PGL}_2(\mathbb{R}),$$

where we explain the equivalence of these spaces in Section 2.

We denote by A the diagonal subgroup of $\mathrm{PGL}_2(\mathbb{R})$ which is parametrized by

$$A = \left\{ a_t^{\pm} = \begin{pmatrix} \pm e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \,:\, t \in \mathbb{R} \right\}.$$

We furthermore use the decomposition $A = A^+ \cup A^-$ where $A^+ = \{a_t^+ \,:\, t \in \mathbb{R}\}$ and $A^- = \{a_t^- \,:\, t \in \mathbb{R}\}$. The subgroup A^+ can viewed as the diagonal subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

1.2. **Acknowledgment.** I thank Menny Aka, Manfred Einsiedler and Andreas Wieser for helping me understand this material. Moreover, I express my gratitude to Noah Held for a useful remark concerning fractional $\mathcal{O}_d$-ideals. Finally, I thank the attendees of my talks on this topic for their presence.

2. The Space of Lattices

We explain in this section how any of the spaces

$$\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}), \quad \mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R}) \quad \text{and} \quad \mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})$$

can be viewed as the space of unimodular lattices of $\mathbb{R}^2$.

A **lattice** in $\mathbb{R}^2$ is a discrete subgroup of $\mathbb{R}^2$ of $\mathbb{Z}$-rank 2. In order words, a lattice L is the $\mathbb{Z}$ of two linearly independent vectors $v_1, v_2 \in \mathbb{R}^2$. For $g = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ it can be written in the form

$$L = \mathbb{Z}v_1 \oplus \mathbb{Z}v_2 = \mathbb{Z}^2 g = \{(m,n)g \,:\, m,n \in \mathbb{Z}\}.$$

We call

$$\mathrm{vol}(L) = |\det(g)|$$

the **volume** of L, and the lattice L **unimodular** if $\mathrm{vol}(L) = 1$.

On the space of all lattices there is a canonical action of $\mathrm{GL}_2(\mathbb{R})$ given by simply multiplying each lattice vector by the given element of $\mathrm{GL}_2(\mathbb{R})$. This action is transitive and the stabilizer at the lattice $\mathbb{Z}^2 < \mathbb{R}^2$ is given by $\mathrm{GL}_2(\mathbb{Z})$. This yields an identification of $\mathrm{GL}_2(\mathbb{Z})\backslash\mathrm{GL}_2(\mathbb{R})$ as the space of all lattices.

In analogy to this we can view $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$ as the space of unimodular lattices. Moreover, $\mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})$ is viewed as the space of all lattices up to homothety, which again can be identified as the space of unimodular lattices.

3. Measures and Orbits in X

We first explain how to arrive at a natural measure on the space $X = \mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})$.

Denote $G = \mathrm{PGL}_2(\mathbb{R})$ and let μ_G be a Haar measure on G, i.e. the unique up to scalar multiples G-invariant Radon measure on G. Let $\Gamma < G$ be a discrete subgroup of G and write $X = \Gamma\backslash G$. Then there exists a G-invariant measure μ_X on the quotient $X = \Gamma\backslash G$ that is uniquely characterized by the property that for any $f \in C_c(G)$ we have

$$\int_G f(g)\, d\mu_G(g) = \int_X \sum_{\gamma\in\Gamma} f(\gamma x)\, d\mu_X(x).$$

We call a discrete subgroup $\Gamma < G$ a **lattice** if the measure μ_X is finite and hence up to normalization of the Haar measure μ_G we can choose μ_X to be a probability measure. We call this probability measure μ_X the **Haar measure** on X.

Returning to $\Gamma = \mathrm{PGL}_2(\mathbb{Z})$ and $X = \mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})$ we note that $\mathrm{PGL}_2(\mathbb{Z})$ is a lattice in $\mathrm{PGL}_2(\mathbb{R})$ and so we fix μ_X the Haar measure on X.

The collection of A-orbits $\mathcal{G}_d$ for which Duke's Theorem is formulated consists of a finite number of **periodic** A-orbits, which we discuss next. In order to give a geometric definition of a periodic orbit, it will be useful to distinguish between the A^+ and the A^--part of the orbit.

A point $x \in X$ is called **A^+-periodic** if there is $t_0 > 0$ such that

$$x.a_0^+ = x = x.a_{t_0}^+ \tag{3.1}$$

as an equality in the space X, or equivalently

$$x = x.(a_{t_0}^+)^n = x.a_{nt_0}^+$$

for all $n \in \mathbb{Z}$. If $x \in X$ is an A-periodic point, then the set

$$\mathcal{O}_x^{A^+} = x.A^+ = \{x.a_t^+ \,:\, t \in \mathbb{R}\} = \{x.a_t^+ \,:\, t \in [0, t_0]\}$$

is called a **periodic A^+-orbit**. By pushing forward the Lebesgue measure on $[0, t_0]$, the last equation allows us to give a natural A-invariant finite measure on X that is supported the periodic orbit $\mathcal{O}_x^{A^+}$. By normalizing this measure, we arrive at a probability measure $\mu_{\mathcal{O}_x^{A^+}}$, which we call the **periodic orbit measure**.

Finally, the smallest $t_0 > 0$ such that (3.1) holds is called the **length**[1] of the orbit $x.A^+$.

In analogy to (3.1) a point $x \in X$ is called **A^--periodic** if there is $t_0 > 0$ such that

$$x.a_0^- = x.a_{t_0}^- \tag{3.2}$$

All of the above carries over to the notion of an A^--periodic point.

Lemma 3.1. *An element $x \in X$ is A^+ periodic if and only if it is A^--periodic and the associated orbits have the same length.*

Proof. For $t_0 > 0$ we have

$$x = x.a_{t_0}^+$$

if and only if

$$x.a_{t_0}^- = x.a_{t_0}^+.a_0^- = x.a_0^-.$$

□

In spite of Lemma 3.1, a point $x \in X$ is called **A-periodic** if it is either A^+- or A^--periodic. As above, each periodic orbit $x.A$ supports an A-invariant probability measure.

[1]This definition fits to the geometric viewpoint discussed in Section 4, more precisely, with this definition the length of the orbit $x.A^+$ corresponds to the hyperbolic length of the associated closed geodesic in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$.

We are now ready to make Duke's Theorem more precise. Namely, to the finite collection of periodic A-orbits $\mathcal{G}_d$ (which will be defined in Section 6) we attach the following A-invariant probability measure:

$$\mu_d = \frac{1}{|\mathcal{G}_d|} \sum_{\mathcal{O}_x^A \in \mathcal{G}_d} \mu_{\mathcal{O}_x^A}.$$

Theorem 3.2. *(Duke's Theorem) As $d \to \infty$ among the non-square discriminants, the measures μ_d converge in the weak* topology to the Haar measure μ_X on X, i.e. for all $f \in C_c(X)$ we have*

$$\int_X f(x)\, d\mu_d(x) \longrightarrow \int_X f(x)\, d\mu_X(x),$$

as $d \to \infty$ among the non-square discriminants.

However, we haven't yet explained of which periodic A-orbits the collection $\mathcal{G}_d$ consists. This will be done in Section 6.

The above definition of a periodic point generalizes to the action of any one-parameter subgroup on any (homogeneous) space. We next investigate the condition in our concrete setting more closely. One first observation the reader could make is that the A action on X is **ergodic**. This has the following consequence.

Corollary 3.3. *We have*

$$\mu_X(\{x \in X \,:\, x \text{ is an } A\text{-periodic point}\}) = 0.$$

Proof. Recall that the A-action on X is ergodic (see Chapter 9 of [EW11]), i.e. for almost all $x \in X$ the orbit of x equidistributes. More precisely, this means that for almost all $x \in X$ and all $f \in L^1(X)$ we have that

$$\frac{1}{T} \int_0^T f(x.a_t)\, dt \to \int_X f(x)\, d\mu_X(x)$$

as $T \to \infty$. Since

$$\{x \in X \,:\, x.A \text{ equidistributes}\} \subset \{x \in X \,:\, x.A \text{ is dense}\},$$

this implies that

$$\mu_X(\{x \in X \,:\, x.A \text{ equidistributes}\}) = \mu_X(\{x \in X \,:\, x.A \text{ is dense}\}) = 1.$$

As a periodic orbit is compact and hence not dense, the claim follows. □

It turns out that we can describe the periodic points in X in a more direct way, which will show that there are only countably many orbits. In order to achieve this, we first make the condition (3.1) more explicit. In the following, we work in the space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and with the A^+ orbit. If $x = \Gamma g$ for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ then x is an A^+-periodic point if and only if there is a matrix $\gamma = \left(\begin{smallmatrix} m & n \\ p & q \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and a smallest $t_0 > 0$ such that

$$\gamma g = \begin{pmatrix} m & n \\ p & q \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{t_0/2} & 0 \\ 0 & e^{-t_0/2} \end{pmatrix} = g a_{t_0}.$$

Inverting g yields $\gamma = ga_{t_0}g^{-1}$. This last equation gives a diagonalization of γ. The characteristic polynomial of γ is $X^2 - \mathrm{tr}(\gamma)X + 1$ and thus the eigenvalues of γ have the form

$$\lambda_\pm = \frac{1}{2}(\mathrm{tr}(\gamma) \pm \sqrt{\mathrm{tr}(\gamma)^2 - 4}) \tag{3.3}$$

and so in particular

$$e^{\pm t_0/2} = \frac{1}{2}(\mathrm{tr}(\gamma) \pm \sqrt{\mathrm{tr}(\gamma)^2 - 4}).$$

By assumption, we have that a_{t_0} is a real non-trivial matrix and hence we have that $|\mathrm{tr}(\gamma)| > 2$ (i.e. the element is **hyperbolic**). We now attach to this periodic orbit a number field. Namely we consider $K = \mathbb{Q}(\sqrt{\mathrm{tr}(\gamma)^2 - 4})$ and then note that $\lambda_\pm$ are **units** in K (for a definition of a unit see Section 5).

To summarize, for each periodic point $x = \Gamma g$ we can associate an element $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $|\mathrm{tr}(\gamma)| > 2$ such that the eigenvectors of γ are given by the columns of g. We now change the viewpoint and focus on the element $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

In the following, we only consider **hyperbolic** elements $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, i.e. elements with $|\mathrm{tr}(\gamma)| > 2$. This last condition characterizes the elements in $\mathrm{SL}_2(\mathbb{Z})$ uniquely as those which are diagonalizable over $\mathbb{R}$ and are not diagonal or unipotent matrices. So let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ be a hyperbolic element with eigenvalues $\lambda_\pm$, where $\lambda_\pm$ is of the form (3.3). Then let g be some element of $\mathrm{SL}_2(\mathbb{R})$ of which the first column is an eigenvector of λ_+ and the second column is one of λ_-. If g' is another such element then there is $a \in A$ so that

$$g' = ga.$$

Hence we have that the orbits associated to g and g', namely $\Gamma g.A$ and $\Gamma g'.A$, are equal.

If we replace next γ by a some other element of the conjugacy class $\gamma' = \gamma_1\gamma\gamma_1^{-1}$ for $\gamma_1 \in \mathrm{SL}_2(\mathbb{Z})$ then we obtain from the diagonalization $\gamma = gag^{-1}$ the diagonalization of γ' by simply multiplying g by γg. Hence we see that the periodic orbit we attach to γ does not depend on the conjugacy class of γ. This yields a map:

$$\{\text{conjugacy classes of hyperbolic elements}\} \longrightarrow \{A\text{-periodic orbits}\}$$

Proposition 3.4. *We have a bijection:*

$$\{\textit{conjugacy classes of hyperbolic elements in } \mathrm{SL}_2(\mathbb{Z})\}$$
$$\longleftrightarrow$$
$$\{A\textit{-periodic orbits}\}$$

Proof. Above, we described the map from conjugacy classes of hyperbolic elements to A-periodic orbits. The inverse map is given as follows. Let $\mathcal{O}$ is an A-periodic orbit, then choose some $x \in \mathcal{O}$, so $\mathcal{O} = x.A$. Write $x = \Gamma g$ for $g \in \mathrm{SL}_2(\mathbb{R})$. As $x.A$ is periodic, there is $t > 0$ so that $\gamma = ga_t g^{-1} \in \mathrm{SL}_2(\mathbb{Z})$

is a hyperbolic element. Similar arguments to the above show that the conjugacy class of γ only depends on the orbit $\mathcal{O}$. It remains to check that the maps are inverse to each other, which again we omit here for the brevity of this exposition. □

Corollary 3.5. *The set of all A-periodic orbits in X is countable.*

Proof. This follows from the last proposition, as $\mathrm{SL}_2(\mathbb{Z})$ is countable and hence in particular the set of conjugacy classes of hyperbolic elements in $\mathrm{SL}_2(\mathbb{Z})$ is countable. □

4. Periodic Orbits and Closed Geodesics

In this section we give a geometric description of X and of periodic A-orbits.

Denote $\mathbb{H} = \{x + iy : y > 0\}$ together with the hyperbolic Riemannian metric

$$\frac{dx^2 + dy^2}{y^2}.$$

The Riemannian metric defines the length of smooth curves $\phi : [a, b] \to \mathbb{H}$ as

$$L(\phi) = \int_a^b \frac{||\phi'(t)||_2}{\mathrm{Im}(\phi(t))}\, dt$$

and hence induces a metric on $\mathbb{H}$ given by

$$d(x, y) = \inf_{\phi} L(\phi)$$

where ϕ varies over all the curves from x to y. A **geodesic** is a *locally length minimizing curve* on $\mathbb{H}$ (for a more precise definition see any book on Riemannian geometry). It turns out (see Section 1 of [Kog18]) that the geodesics on $\mathbb{H}$ are precisely given by semicircles with center on $\partial\mathbb{H} = (\mathbb{R}, 0) \subset \mathbb{R}^2$ or by lines perpendicular to $\partial\mathbb{H}$.

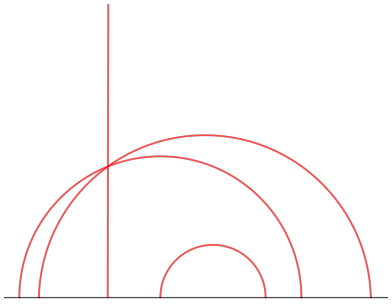

Figure 4.1. Geodesics on $\mathbb{H}$

An object that we are interested in is the **unit tangent bundle**, which consists of the collection of all points (z, v) where $z \in \mathbb{H}$ and v is a tangent vector at z of length one.

Proposition 4.1. *The isometry group is given by*

$$\mathrm{Iso}(\mathbb{H}) = \left\{ f_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(z) = \frac{az+b}{cz+d} \,:\, \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \right\} \cong \mathrm{SL}_2(\mathbb{R})$$

and acts transitively on $\mathbb{H}$*. Moreover,*

$$\mathrm{Stab}_{(i)}(\mathbb{H}) \cong \mathrm{SO}_2(\mathbb{R}).$$

This yields an identification

$$\mathbb{H} \cong \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R}).$$

Finally, we note that the action

$$\mathrm{SL}_2(\mathbb{R}) \ni g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \left((z,v) \in T^1(\mathbb{H}) \mapsto \left(\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2} \right) \right)$$

is transitive and $\mathrm{Stab}_{(i,i)}(T^1(\mathbb{H})) = \{\pm I_2\}$*, yielding the identification*

$$T^1(\mathbb{H}) \cong \mathrm{SL}_2(\mathbb{R})/\{\pm I_2\} =: \mathrm{PSL}_2(\mathbb{R}).$$

Proof. See Chapter 9 of [EW11]. □

We next describe the geodesic flow. Fix some time $t \in \mathbb{R}$. The **geodesic flow** at time t for an element $(z,v) \in T^1(\mathbb{H})$ is defined by following the uniquely characterized parametrization of the geodesic determined by (z,v) for time t. Under the above identification of $T^1(\mathbb{H})$, the geodesic flow is simply given as left matrix multiplication by a_t. This is captured by the next proposition.

Proposition 4.2. *For* $t \in \mathbb{R}$ *the following diagram commutes:*

$$\begin{array}{ccc} T^1(\mathbb{H}) & \xrightarrow{g_t} & T^1(\mathbb{H}) \\ \downarrow\cong & & \downarrow\cong \\ \mathrm{PSL}_2(\mathbb{R}) & \xrightarrow{\cdot a_t^+} & \mathrm{PSL}_2(\mathbb{R}) \end{array}$$

Proof. See Chapter 9 of [EW11]. □

Now let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be discrete, for instance $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Identifying Γ as part of the isometry group of $\mathbb{H}$, we get a surface $\Gamma\backslash\mathbb{H}$. The above description of the unit tangent bundle carries over to $\Gamma\backslash\mathbb{H}$. More precisely, we can identify the unit tangent bundle $T^1(\Gamma\backslash\mathbb{H})$ with $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ and we also get the commutative diagram:

$$\begin{array}{ccc} T^1(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) & \xrightarrow{g_t} & T^1(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) \\ \downarrow\cong & & \downarrow\cong \\ \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}) & \xrightarrow{\cdot a_t^+} & \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}) \end{array}$$

A geodesic is called **closed**, if the orbit closes itself and hence one arrives at a compact orbit of finite length. In order to make this definition more precise it is useful to phrase it in terms of the unit tangent bundle. In analogy to the definition of an A-periodic point, we call a point $(x, v) \in T^1(\Gamma\backslash\mathbb{H})$ **periodic** for the geodesic flow if there is some time $t_0 > 0$ such that

$$g_{t_0}(x, v) = (x, v) \qquad \text{in } T^1(\Gamma\backslash\mathbb{H}).$$

So a **closed** geodesic is just the orbit of x under the geodesic flow, if (x, v) is a periodic point. Choosing the smallest such t_0, it is by construction precisely the hyperbolic length of the geodesic $\{g_t(z, v) : t \in [0, t_0]\}$.

The above diagram consequently yields bijections:

$$\{A\text{-periodic points in } \Gamma\backslash\mathrm{SL}_2(\mathbb{R})\} \longleftrightarrow \{\text{periodic points in } \Gamma\backslash\mathbb{H}\}$$

$$\{A\text{-periodic orbits on } \Gamma\backslash\mathrm{SL}_2(\mathbb{R})\} \longleftrightarrow \{\text{closed geodesics on } \Gamma\backslash\mathbb{H}\}$$

In order to visualize the surface $\Gamma\backslash\mathbb{H}$, we use a **fundamental domain**, i.e. a subset $S \subset \mathbb{H}$ such that

$$\mathbb{H} = \bigsqcup_{\gamma\in\Gamma} \gamma S,$$

where we assume that the union is disjoint up to sets of measure zero (see [EW11] Section 11 for a more careful treatment). For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, a fundamental domain is given by

$$S = \{z \in \mathbb{H} : -\tfrac{1}{2} \le \mathrm{Im}(z) \le \tfrac{1}{2} \text{ and } |z| \ge 1\}. \tag{4.1}$$

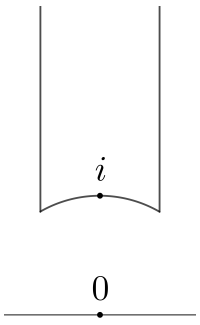

FIGURE 4.2. Fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$

5. A Quick Review of Algebraic Number Theory

In this section, some definitions and central results of algebraic number theory are briefly reviewed. A good reference is the book of Neukirch [Neu92].

Let K be a number field of degree $n = [K : \mathbb{Q}]$. We denote by $\mathcal{O}_K$ the *integral closure* of $\mathbb{Z}$ in K, i.e. the elements of K that are a zero of a monic polynomial with coefficients in $\mathbb{Z}$. We also sometimes call $\mathcal{O}_K$ the

ring of integers of K. For example if d is a square-free positive integer and $K = \mathbb{Q}(\sqrt{d})$, then

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\frac{d+\sqrt{d}}{2}] & \text{if } d \equiv 1 \mod 4, \\ \mathbb{Z}[\sqrt{d}] & \text{if } d \equiv 2, 3 \mod 4. \end{cases}$$

Theorem 5.1. *Any non-zero finitely generated $\mathcal{O}_K$-submodule of K is a free $\mathbb{Z}$-module of rank n.*

Proof. See Theorem 2.10 of Chapter 1 of [Neu92]. □

Definition 5.2. *A **fractional ideal** of K is a non-zero finitely generated $\mathcal{O}_K$-submodule of K. The **class group** of K is the group*

$$\mathrm{Cl}_K = \{\textit{fractional ideals of } K\}/K^\times,$$

*whose cardinality is called the **class number** of K.*

Roughly speaking, the class number measures how far away $\mathcal{O}_K$ is from being from a principal ideal domain.

Theorem 5.3. *The class number of any number field is a finite group. Moreover, the inverse of any ideal class $[\mathfrak{a}] \in \mathrm{Cl}_K$ is given by the ideal class defined by*

$$\mathfrak{a}^{-1} = \{\lambda \in K \,:\, \lambda\mathfrak{a} \subset \mathcal{O}_K\}$$

Proof. See Chapters 1.3 and 1.6 of [Neu92]. □

It will turn out to be important to understand the units of $\mathcal{O}_K$. We first make the observation that there are exactly n field embeddings $\tau : K \to \mathbb{C}$ that fix $\mathbb{Q}$, which we denote $\tau_1, \ldots, \tau_n$. Then we define the **norm** of an element $\lambda \in K$ to be

$$\mathrm{N}(\lambda) = \prod_{i=1}^{n} \tau_i(\lambda).$$

Observe

$$\mathcal{O}_K^\times = \{\lambda \in \mathcal{O}_K \,:\, \mathrm{N}(\lambda) = \pm 1\}.$$

We next discuss the discriminant of a fractional ideal $\mathfrak{a}$. If $a_1, \ldots, a_n$ is a $\mathbb{Z}$-basis of $\mathfrak{a}$, then the discriminant of $\mathfrak{a}$ is defined as

$$d(\mathfrak{a}) = \det(\tau_i a_j)^2.$$

The discriminant does not depend on the choice of integral basis and is moreover not zero. As a general fact, we note that for fractional ideals $\mathfrak{a} \subset \mathcal{O}_K$ (see Chapter 1.1. of [Neu92])

$$d(\mathfrak{a}) = (\mathcal{O}_K : \mathfrak{a})^2 d(\mathcal{O}_K) = \mathrm{N}(\mathfrak{a})^2 d(\mathcal{O}_K), \tag{5.1}$$

where we call $\mathrm{N}(\mathfrak{a}) = (\mathcal{O}_K : \mathfrak{a})$ the **norm** of $\mathfrak{a}$.

The field embeddings $K \to \mathbb{C}$ are grouped into two classes, namely first the **real embeddings** $\tau : K \to \mathbb{C}$ that satisfy $\tau(K) \subset \mathbb{R}$ and second the **complex embeddings** $\tau : K \to \mathbb{C}$ that are not real embeddings, i.e. $\tau(K) \not\subset \mathbb{R}$. The complex embeddings appear in pairs, as for each complex

embedding $\tau : K \to \mathbb{C}$, the conjugate embedding $\overline{\tau} : K \to \mathbb{C}$ forms a separate complex embedding $K \to \mathbb{C}$. If r denotes the number of real embeddings and $2s$ the number of complex ones, then $r + 2s = n$. The number field K is then called of **type** (r, s). Now we are ready to describe the units of K.

Theorem 5.4. *(Dirichlet's Unit Theorem) Let K be a number field of type (r, s). Then the units of the ring of integers can be expressed as a product*

$$\mathcal{O}_K^\times = \mu(K) \times F,$$

where $\mu(K)$ are the roots of unity of K and F is a free $\mathbb{Z}$-module of rank $r + s - 1$.

Proof. See Chapter 1.7 of [Neu92]. □

We next want to generalize some of the above concepts to a more general class of rings than just the ring $\mathcal{O}_K$, namely so-called **orders**, which we define next.

Definition 5.5. *A subring $\mathcal{O} \subset \mathcal{O}_K$ is called an **order** if it is a finitely generated $\mathbb{Z}$-module of rank n, i.e. a ring of the form*

$$\mathbb{Z}[\alpha_1, \ldots, \alpha_n],$$

where $\alpha_1, \ldots, \alpha_n$ are integral elements that form a basis of K over $\mathbb{Q}$.

Another important example of an order is given as follows. If $\alpha_1, \ldots, \alpha_n$ is a basis of K over $\mathbb{Q}$ and we denote $\mathfrak{a} = \mathbb{Z}\alpha_1 + \ldots + \mathbb{Z}\alpha_n$, then

$$\mathcal{O} = \{\lambda \in K \,:\, \lambda\mathfrak{a} \subset \mathfrak{a}\}$$

is an order in K (for a proof of this fact see Section 3.4 of [Kog]). We next generalize a few of the above notions for orders.

Definition 5.6. *Let $\mathcal{O}$ be an order. We call a non-zero finitely generated $\mathcal{O}$-submodule of K an **$\mathcal{O}$-fractional ideal**. We say that the fractional ideal $\mathfrak{a}$ is **invertible**, if there is another fractional ideal $\mathfrak{b}$ so that*

$$\mathfrak{a} \cdot \mathfrak{b} = \{ab \,:\, a \in \mathfrak{a} \text{ and } b \in \mathfrak{b}\} = \mathcal{O}.$$

*The **Picard group** of the order $\mathcal{O}$ is*

$$\mathrm{Pic}(\mathcal{O}) = \{\text{invertible fractional } \mathcal{O}\text{-ideals}\}/K^\times.$$

Theorem 5.7. *(Dirichlet's Unit Theorem for orders) Let K be a number field of type (r, s) and let $\mathcal{O}$ be an order. Then the units of $\mathcal{O}$ can be expressed as a product*

$$\mathcal{O} = T \times F,$$

where T is a torsion subgroup of $\mathcal{O}$ and F is a free $\mathbb{Z}$-module of rank $r+s-1$.

Another class of ideals that will be important in Section 6 are so called **proper $\mathcal{O}$-ideals**.

Definition 5.8. *Let $\mathcal{O}$ be an order in K. An $\mathcal{O}$-submodule $\mathfrak{a} \subset K$ is called a **proper $\mathcal{O}$-ideal** if it is a free $\mathbb{Z}$-module of rank n and satisfies*

$$\mathcal{O} = \{\lambda \in K \,:\, \lambda\mathfrak{a} \subset \mathfrak{a}\}.$$

For a proper $\mathcal{O}$-ideal $\mathfrak{a}$ we define the norm of $\mathfrak{a}$ with respect to $\mathcal{O}$ as

$$\mathrm{N}(\mathfrak{a}) = \frac{(\mathcal{O} : \mathcal{O} \cap \mathfrak{a})}{(\mathfrak{a} : \mathcal{O} \cap \mathfrak{a})}.$$

6. Ideals and Periodic Orbits

As before, let d be a discriminant and $K = \mathbb{Q}(\sqrt{d})$. Throughout this section write $\mathcal{O}_d = \mathbb{Z}[\frac{d+\sqrt{d}}{2}]$ and note that as d is a discriminant, $\mathcal{O}_d \subset \mathcal{O}_K$ and so $\mathcal{O}_d$ is an order. For the beginning of this section, we assume that d is a **fundamental** discriminant, meaning that $\mathcal{O}_K = \mathcal{O}_d$ which is equivalent to d being square-free in the case $d \equiv 1 \mod 4$ and that $\frac{d}{4}$ is square-free if $d \equiv 0 \mod 4$. Moreover, denote throughout the non-trivial embedding

$$\tau : \mathbb{Q}(\sqrt{d}) \to \mathbb{Q}(\sqrt{d}), \qquad x + y\sqrt{d} \mapsto x - y\sqrt{d}$$

for $x, y \in \mathbb{Q}$.

The aim of this section is to associate to fractional ideals $\mathfrak{a} \subset \mathcal{O}_K$, A-periodic points $x_\mathfrak{a} \in X$. We start by considering a $\mathbb{Z}$-basis a_1, a_2 of $\mathfrak{a}$ and set

$$g_{(a_1,a_2)} = \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \in \mathrm{PGL}_2(\mathbb{R}) \quad \text{and} \quad x_{(a_1,a_2)} = \Gamma g_{(a_1,a_2)} \in X. \tag{6.1}$$

We observe that the matrix $\left(\begin{smallmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{smallmatrix}\right)$ has non-zero determinant as the square of its determinant is the discriminant, which is non-zero. Thus $g_{(a_1,a_2)}$ is a well-defined element of $\mathrm{PGL}_2(\mathbb{R})$.

Lemma 6.1. *The element $x_{(a_1,a_2)} \in X$ does not depend on the choice of $\mathbb{Z}$-basis.*

Proof. The argument is analogous to showing that the discriminant of a fractional ideal does not depend on choice of integral basis. Let (b_1, b_2) be another $\mathbb{Z}$-basis. Then by definition, there are integers $n, m, p, q \in \mathbb{Z}$ so that

$$b_1 = ma_1 + na_2, \qquad b_2 = pa_1 + qa_2$$

and

$$g_{(b_1,b_2)} = \begin{pmatrix} b_1 & \tau(b_1) \\ b_2 & \tau(b_2) \end{pmatrix} = \begin{pmatrix} m & n \\ p & q \end{pmatrix} \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} = \begin{pmatrix} m & n \\ p & q \end{pmatrix} g_{(a_1,a_2)}.$$

Furthermore by the same argument, the inverse of $\left(\begin{smallmatrix} m & n \\ p & q \end{smallmatrix}\right)$ has again integer entries and so $\left(\begin{smallmatrix} m & n \\ p & q \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z})$. □

Lemma 6.1 allows us to define

$$x_\mathfrak{a} := x_{(a_1,a_2)} \in X$$

for any $\mathbb{Z}$-basis a_1, a_2 of $\mathfrak{a}$.

Lemma 6.2. *The element $x_{\mathfrak{a}} \in X$ is A-periodic.*

Proof. We show that $x_{\mathfrak{a}}$ is A^+-periodic. Choose a non-trivial unit $\varepsilon \in \mathcal{O}_K^\times$ with $\mathrm{N}(\varepsilon) = 1$. Such a unit can be found by choosing for instance a fundamental unit and then squaring it. Next choose an integral basis a_1, a_2. As ε is an element of $\mathcal{O}_K$ and $\mathcal{O}_K = \{\lambda \in K : \lambda\mathfrak{a} \subset \mathfrak{a}\}$, we can find integers $m, n, p, q \in \mathbb{Z}$,

$$\varepsilon a_1 = ma_1 + na_2 \qquad \text{and} \qquad \varepsilon a_2 = pa_1 + qa_2.$$

Hence

$$\begin{pmatrix} m & n \\ p & q \end{pmatrix} \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} = \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \begin{pmatrix} \varepsilon & 0 \\ 0 & \tau(\varepsilon) \end{pmatrix}. \tag{6.2}$$

As by assumption $\mathrm{N}(\varepsilon) = \varepsilon\tau(\varepsilon) = 1$, we have that $\left(\begin{smallmatrix} \varepsilon & 0 \\ 0 & \tau(\varepsilon) \end{smallmatrix}\right) \in A^+$ and this implies $\left(\begin{smallmatrix} m & n \\ p & q \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$. Thus $x_{\mathfrak{a}}$ is A^+-periodic. □

Lemma 6.3. *Let ε be a fundamental unit of $\mathcal{O}_K$. Set*

$$c_K = \begin{cases} 1 & \text{if } N(\varepsilon) = 1, \\ 2 & \text{if } N(\varepsilon) = -1. \end{cases}$$

Then the length of the orbit $x_{\mathfrak{a}}.A^+$ is equal to

$$2c_K |\log(|\varepsilon|)| = 2c_K \mathrm{Reg}(\mathcal{O}_K).$$

Proof. Let $\mu \in K$ and set $t = \left(\begin{smallmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{smallmatrix}\right)$. Assume that $x_{\mathfrak{a}} = x_{\mathfrak{a}}t$ in X. We first claim that this can only happen if $\mu \in \mathcal{O}_K^\times$. So let a_1, a_2 be an integral basis and then we have (6.2) with the only difference that we replace $\tau(\mu)$ by μ^{-1}. Thus, multiplication by ε is represented by an integer matrix and so $\mu \in \mathcal{O}_K$. Moreover, it follows directly by computing the matrix entries that $\tau(\mu) = \mu^{-1}$ and so μ is a unit in $\mathcal{O}_K$.

We next use Dirichlet's Unit Theorem. The field $\mathbb{Q}(\sqrt{d})$ is of type $(1, 0)$, so the unit group $\mathcal{O}_K^\times$ is the product of a torsion part and a free $\mathbb{Z}$-module of rank 1. Let $\varepsilon \in \mathcal{O}_K^\times$ be a fundamental unit. We note that the set of fundamental units is given by $\{\pm\varepsilon^\pm\}$ and so c_K is independent of the choice of fundamental unit. Thus the smallest non-trivial element of $\mathcal{O}_K$ such that (6.2) holds is ε^{c_K}, and so the claim follows. So the length $\ell_0 > 0$ of the orbit $x_{\mathfrak{a}}.A^+$ is determined by $e^{\ell_0/2} = |\varepsilon^{c_K}|$ so that $\ell_0 = 2c_K \log(|\varepsilon|)$, where we assume without loss of generality that $|\log(|\varepsilon|)| > 0$. □

The next lemma clarifies the difference between the A^+-part and the A^--part of the periodic A-orbit.

Lemma 6.4. *In the setting of Lemma 6.3, $x_{\mathfrak{a}}.A = x_{\mathfrak{a}}.A^+$ if and only if the norm of a fundamental unit is -1. Thus, the length of the A-orbit $x_{\mathfrak{a}}.A$ is $4\mathrm{Reg}(\mathcal{O}_K)$.*

Proof. Let $\varepsilon \in \mathcal{O}_K^\times$ be a fundamental unit and assume that $N(\varepsilon) = -1$. Then, $t^- = \operatorname{diag}(\varepsilon, \tau(\varepsilon)) \in A^-$ and $x_{\mathfrak{a}} = x_{\mathfrak{a}}.t^-$. Hence $x_{\mathfrak{a}}.A^+ = x_{\mathfrak{a}}.A^-$, showing that $x_{\mathfrak{a}}.A = x_{\mathfrak{a}}.A^+$.

Conversely assume that $x_{\mathfrak{a}}.A^+ = x_{\mathfrak{a}}.A^-$. Then there is some $t = \operatorname{diag}(\mu, \mu^{-1}) \in A^+$ so that $x_{\mathfrak{a}}.t = x_{\mathfrak{a}}.\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. By the same argument as in the first paragraph of the proof of Lemma 6.3, we conclude that $\varepsilon \in \mathcal{O}_K^\times$ and that $\varepsilon^{-1} = \tau(-\varepsilon)$. So we have that $N(\varepsilon) = -1$ and hence the norm of the fundamental unit must also be -1 as otherwise the norm of any unit would be 1. □

We next show that the periodic orbit only depends on the the ideal class of $\mathfrak{a}$.

Lemma 6.5. *Let $\mathfrak{a}, \mathfrak{b} \subset K$ be two fractional ideals. Then there exists $\lambda \in K^\times$ such that*

$$\mathfrak{b} = \lambda\mathfrak{a}$$

if and only if

$$x_{\mathfrak{a}}.A = x_{\mathfrak{b}}.A.$$

Proof. Assume there is $\lambda \in K^\times$ such that $\mathfrak{b} = \lambda\mathfrak{a}$ and set $a = \left(\begin{smallmatrix} \lambda & 0 \\ 0 & \tau(\lambda) \end{smallmatrix}\right) \in A$. Then for any integral basis a_1, a_2 we have that $\lambda a_1, \lambda a_2$ is an integral basis of $\mathfrak{b}$ and we have

$$\begin{pmatrix} \lambda_1 a_1 & \tau(\lambda_1 a_2) \\ \lambda_2 a_2 & \tau(\lambda_2 a_2) \end{pmatrix} = \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \tau(\lambda) \end{pmatrix} = \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} a.$$

Thus $x_{\mathfrak{b}} = x_{\mathfrak{a}}.a$, implying $x_{\mathfrak{b}}.A = x_{\mathfrak{a}}.A$.

Now assume that $x_{\mathfrak{a}}.A = x_{\mathfrak{b}}.A$ or equivalently that there exists $a = \operatorname{diag}(\pm t, t^{-1}) \in A$ for $t \neq 0$ so that

$$x_{\mathfrak{b}} = x_{\mathfrak{a}}.a. \tag{6.3}$$

Let b_1, b_2 be an integral basis of $\mathfrak{b}$ and a_1, a_2 one of $\mathfrak{a}$. We note that the determinant of the matrix $g_{(b_1,b_2)}$ is equal to the square-root of the discriminant. Then (6.3) is equivalent to

$$\begin{pmatrix} m & n \\ p & q \end{pmatrix} \frac{1}{d(\mathfrak{b})^{\frac{1}{4}}} \begin{pmatrix} b_1 & \tau(b_1) \\ b_2 & \tau(b_2) \end{pmatrix} = \frac{1}{d(\mathfrak{a})^{\frac{1}{4}}} \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \begin{pmatrix} \pm t & 0 \\ 0 & t^{-1} \end{pmatrix}$$

for $\left(\begin{smallmatrix} m & n \\ p & q \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z})$. By replacing b_1 by $mb_1 + nb_2$ and b_2 by $pb_1 + qb_2$ we again arrive at an integral basis and

$$\frac{1}{d(\mathfrak{b})^{\frac{1}{4}}} \begin{pmatrix} b_1 & \tau(b_1) \\ b_2 & \tau(b_2) \end{pmatrix} = \frac{1}{d(\mathfrak{a})^{\frac{1}{4}}} \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \begin{pmatrix} \pm t & 0 \\ 0 & t^{-1} \end{pmatrix}.$$

This implies

$$\begin{pmatrix} b_1 & \tau(b_1) \\ b_2 & \tau(b_2) \end{pmatrix} = \begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2) \end{pmatrix} \frac{d(\mathfrak{b})^{\frac{1}{4}}}{d(\mathfrak{a})^{\frac{1}{4}}} \begin{pmatrix} \pm t & 0 \\ 0 & t^{-1} \end{pmatrix}. \tag{6.4}$$

Set

$$t' = \frac{d(\mathfrak{b})^{\frac{1}{2}}}{d(\mathfrak{a})^{\frac{1}{2}}} \pm t.$$

Then (6.4) shows $b_1 = t'a_1$, and $b_2 = t'a_2$ or equivalently $t' = \frac{b_1}{a_1} = \frac{b_2}{a_2} \in K^\times$ and $\mathfrak{b} = t'\mathfrak{a}$. □

The collection of A-periodic orbits $\mathcal{G}_d$ for which Duke's Theorem is formulated is described next. Namely, we simply take all geodesics associated to all fractional ideals, arriving at a finite collection as the class number is finite. So define

$$\mathcal{G}_d = \bigcup_{\mathfrak{a} \in \mathrm{Cl}_K} x_{\mathfrak{a}}.A$$

and then

$$\mu_d = \frac{1}{|\mathrm{Cl}_K|} \sum_{\mathfrak{a} \in \mathrm{Cl}_K} \mu_{x_{\mathfrak{a}}.A}.$$

Recall that we assumed up to now that d is a fundamental discriminant. We explain next how to extend all of the above to the case where d is a non-fundamental discriminant. In this case, $\mathcal{O}_d = \mathbb{Z}[\frac{d+\sqrt{d}}{2}]$ is an order in $\mathcal{O}_K$ and no longer equal to $\mathcal{O}_K$. We want to associate to invertible $\mathcal{O}_d$-ideals A-periodic orbits. For all of the above, it was central that every fractional ideal is a free $\mathbb{Z}$-module of rank 2. For an invertible fractional $\mathcal{O}_d$-ideal $\mathfrak{a}$ this also easily follows: As $\mathcal{O}_d$ is finitely generated over $\mathbb{Z}$ and by definition $\mathfrak{a}$ is finitely generated over $\mathcal{O}_d$, it follows that $\mathfrak{a}$ is a finitely generated $\mathbb{Z}$-module. Moreover, as $\mathfrak{a}$ has no torsion, it is a free $\mathbb{Z}$-module and its rank can be computed by

$$\mathrm{rank}_{\mathbb{Z}}(\mathfrak{a}) = \dim_{\mathbb{Q}}(\mathfrak{a} \otimes_{\mathbb{Z}} \mathbb{Q}) = \dim_{\mathbb{Q}}(K) = 2,$$

where we used that

$$\mathfrak{a} \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathfrak{a} \otimes_{\mathbb{Z}} \mathbb{Z}_{\mathbb{Z}\setminus\{0\}} \cong \mathfrak{a}_{\mathbb{Z}\setminus\{0\}} \cong \mathfrak{a} \otimes_{\mathcal{O}_d} (\mathcal{O}_d)_{\mathbb{Z}\setminus\{0\}} \cong \mathfrak{a} \otimes_{\mathcal{O}_d} K \cong K.$$

The last isomorphism follows as $\mathfrak{a} \otimes_{\mathcal{O}_d} K$ is an invertible K-vector space and hence one-dimensional. More generally, the next proposition holds.

Proposition 6.6. *Let $\mathfrak{a} \subset K$ be a fractional $\mathcal{O}_d$-ideal. Then $\mathfrak{a}$ is invertible if and only if $\mathfrak{a}$ is a proper $\mathcal{O}_d$-ideal, i.e. $\mathfrak{a}$ is a free $\mathbb{Z}$-module of rank 2 and satisfies*

$$\mathcal{O}_d = \{\lambda \in \mathfrak{a} \,:\, \lambda\mathfrak{a} \subset \mathfrak{a}\}.$$

Proof. See [ELMV12] Section 2.2. □

The last proposition allows us to use exactly the same procedure as above to associate to each invertible fractional $\mathcal{O}_d$-ideal $\mathfrak{a} \subset K$ an A-periodic point $x_{\mathfrak{a}}$ that only depends on the ideal class $[\mathfrak{a}] \in \mathrm{Pic}(\mathcal{O}_d)$. This allows us to define for a general non-square integer d,

$$\mathcal{G}_d = \bigcup_{\mathfrak{a} \in \mathrm{Pic}(\mathcal{O}_d)} x_{\mathfrak{a}}.A \qquad \text{and} \qquad \mu_d = \frac{1}{|\mathrm{Pic}(\mathcal{O}_d)|} \sum_{\mathfrak{a} \in \mathrm{Pic}(\mathcal{O}_d)} \mu_{x_{\mathfrak{a}}.A}.$$

The formulation of Theorem 3.2 is hereby complete. In full generality, one can associated to proper $\mathcal{O}$-ideals for any order $\mathcal{O}$, periodic A-orbits on X (see Section 3.5 of [Kog]). One can then show an analogous result to Lemma 6.5. Namely, if $\mathfrak{a}$ is a proper $\mathcal{O}$-ideal and $\mathfrak{a}'$ is a proper $\mathcal{O}'$-ideal, then

$$x_{\mathfrak{a}}.A = x_{\mathfrak{a}'}.A$$

if and only if

$$\mathcal{O} = \mathcal{O}' \quad \text{and} \quad [\mathfrak{a}] = [\mathfrak{a}'].$$

Bases and Periodic Orbits. In this short subsection we give a slightly different viewpoint on the construction given above. Let $a, b \in K$ be a $\mathbb{Q}$-basis of K. We aim at associating to the basis a, b a periodic orbit. In particular, it will be shown that

$$x_{(a,b)} := \begin{pmatrix} a & \tau(a) \\ b & \tau(b) \end{pmatrix} \in \mathrm{PGL}_2(\mathbb{R})$$

is a periodic point. This will be done in a similar fashion as showing that $x_{\mathfrak{a}}$ is periodic.

Write $\mathfrak{a}_{(a,b)} = \mathbb{Z}a \oplus \mathbb{Z}b$ and set

$$\mathcal{O}_{(a,b)} = \{\lambda \in K \,:\, \lambda\mathfrak{a}_{(a,b)} \subset \mathfrak{a}_{(a,b)}\}.$$

Then $\mathcal{O}_{(a,b)}$ is an order in K, as we can represent multiplication by such a λ with respect to the basis (a, b) by an integer matrix and hence use the characteristic polynomial of this matrix to show that λ is an algebraic integer. Hence $\mathcal{O}_{(a,b)} \subset \mathcal{O}_K$ and so it is an order, as $\mathcal{O}_{(a,b)}$ is a free $\mathbb{Z}$-module of rank $[K : \mathbb{Q}]$. Moreover, $\mathfrak{a}_{(a,b)}$ is a proper $\mathcal{O}_{(a,b)}$-ideal. Thus we can again use the units of the order $\mathcal{O}_{(a,b)}$ to show that $x_{(a,b)}$ is periodic.

7. The Height of Lattices and the Norm of Ideals

In this section we explain the relation between the **height** of lattices and the norm of ideals. We closely follow [ELMV12] Section 3.1, making some details more explicit.

We define the **height** of a lattice $L = \mathbb{Z}^2 g$ for $g \in \mathrm{GL}_2(\mathbb{R})$ as

$$\mathrm{ht}(L) = \left(\frac{\min_{x \in L \setminus \{0\}} ||x||}{\mathrm{vol}(L)^{\frac{1}{2}}}\right)^{-1} = \left(\frac{\min_{x \in \mathbb{Z}^2 \setminus \{0\}} ||xg||}{\mathrm{vol}(L)^{\frac{1}{2}}}\right)^{-1}.$$

Observe that $\mathrm{ht}(L)$ only depends on the homothety class of L and hence the hight is well-defined for elements in $\mathrm{PGL}_2(\mathbb{R})$.

We can relate the height of a unimodular lattice to a geometric quantity on the modular surface, as stated in the next lemma. We denote as in Section 4 by S the fundamental domain for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ given by (4.1).

Lemma 7.1. *Let $x \in \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and assume $x \cong (z, v) \in T^1(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})$ for $z \in S$. Then*

$$\mathrm{Im}(z) = \mathrm{ht}(x)^2.$$

Proof. We can choose $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ with $x = \Gamma g$ such that $g.i \in S$, i.e. $|\mathrm{Re}(g.i)| \leq \frac{1}{2}$ and $|g.i| \geq 1$. As

$$g.i = \frac{ai+b}{ci+d} = \frac{ai+b}{ci+d}\frac{d-ci}{d-ci} = \frac{ac+bd}{c^2+d^2} + i\frac{1}{c^2+d^2}$$

the assumption $|\mathrm{Re}(g.i)| \leq \frac{1}{2}$ translates to

$$|\mathrm{Re}(z)| = \left|\frac{ac+bd}{c^2+d^2}\right| \leq \frac{1}{2}.$$

Moreover we can assume upon multiplying g by $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and thus replacing $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ by $g = \left(\begin{smallmatrix} -c & -d \\ a & d \end{smallmatrix}\right)$ that

$$c^2 + d^2 \leq a^2 + b^2.$$

Set $L = \mathbb{Z}^2 g$. Then

$$\begin{aligned} \mathrm{ht}(x)^{-2} = \mathrm{ht}(L)^{-2} &= \min_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} ||(m,n)g||^2 \\ &= \min_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \left((ma+nc)^2 + (mb+nd)^2\right) \\ &= \min_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \left(m^2(a^2+b^2) + n^2(c^2+d^2) + 2mn(ac+bd)\right). \end{aligned}$$

Thus we have that

$$\frac{\mathrm{Im}(g.i)}{\mathrm{ht}(x)^2} = \min_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \left(m^2\frac{a^2+b^2}{c^2+d^2} + n^2 + 2mn\frac{ac+bd}{c^2+d^2}\right) = 1.$$

□

Let $H > 1$ and denote

$$X_{\geq H} = \{x \in X \,:\, \mathrm{ht}(x) \geq H\}.$$

Proposition 7.2. *Let $\mathfrak{a} \subset \mathcal{O}_d$ be a proper $\mathcal{O}_d$-ideal. Then $x_{\mathfrak{a}}.A \cap X_{\geq H}$ is nonempty if and only if $\mathfrak{a}^{-1}$ is in the same ideal class as an ideal $\mathfrak{b} \subset \mathcal{O}_d$ of norm $\leq \frac{1}{2}H^{-2}d^{1/2}$.*

Proof. We first observe the following. If we identify $x \in X$ with the unimodular lattice L, we claim that $x.A \cap X_{\geq H}$ is nonempty if and only if there is some nonzero vector

$$(u,v) \in L \quad \text{with} \quad |uv| \leq \frac{1}{2}H^{-2}. \tag{7.1}$$

This observation follows from calculating the minimal norm achieved under the A-action for an element $0 \neq (u,v) \in L$. So consider the continuously on t dependent function

$$\left|\left|(u,v)\cdot a_t\right|\right|^2 = \left|\left|(u,v)\cdot\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2}\end{pmatrix}\right|\right|^2 = u^2e^t + v^2e^{-t},$$

which has derivative

$$u^2e^t - v^2e^{-t}.$$

The derivative is zero if and only if (assuming w.l.o.g. $u \neq 0$) $t = \log\left(|\frac{v}{u}|\right)$. Hence the minimum of the function $||(u,v)\cdot a_t||^2$ is $|2uv|$.

Using this, assume we have $(u,v) \in L = \mathbb{Z}^2 x$ with $|uv| \leq \frac{1}{2}H^{-2}$. Then, as above, there is some t_0 so that

$$\left|\left|(u,v)\cdot a_{t_0}\right|\right| = \sqrt{|2uv|} \leq H^{-1}$$

implying $\mathrm{ht}(L.a_{t_0})^{-1} \leq H^{-1}$ and so $\mathrm{ht}(L.a_{t_0}) \geq H$.

For the converse assume that for some t_0, $\mathrm{ht}(L.a_{t_0}) \geq H$ or equivalently $\mathrm{ht}(L.a_{t_0})^{-1} \leq H^{-1}$. Then there is $(u,v) \in L$ such that

$$\sqrt{|2uv|} \leq \left|\left|(u,v)\cdot a_{t_0}\right|\right| \leq H^{-1},$$

implying the claim.

Now let $\mathfrak{a}$ be a proper $\mathcal{O}_d$-ideal with integral basis a_1, a_2. Then

$$\det\begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2)\end{pmatrix} = \sqrt{d(\mathfrak{a})} = (\mathcal{O}_d : \mathfrak{a})\sqrt{d(\mathcal{O}_d)} = \mathrm{N}(\mathfrak{a})d^{\frac{1}{2}}.$$

In the following we want to normalize $x_{\mathfrak{a}}$ to arrive at an element with determinant ± 1. So we get

$$x_{\mathfrak{a}} = \frac{1}{\sqrt{\mathrm{N}(\mathfrak{a})d^{\frac{1}{2}}}}\begin{pmatrix} a_1 & \tau(a_1) \\ a_2 & \tau(a_2)\end{pmatrix}$$

and the lattice $L = \mathbb{Z}^2 x_{\mathfrak{a}}$ given by $x_{\mathfrak{a}}$ is

$$L = \left\{ \frac{1}{\sqrt{\mathrm{N}(\mathfrak{a})d^{\frac{1}{2}}}}(na_1 + ma_2, \tau(na_1 + ma_1)) \,:\, n,m \in \mathbb{Z} \right\}.$$

So the above condition translates to the following conclusion: $x_{\mathfrak{a}}.A$ intersects $X_{\geq H}$ if and only if $\mathfrak{a}$ contains an element λ so that

$$|\mathrm{N}(\lambda)| \leq \frac{1}{2}H^{-2}\mathrm{N}(\mathfrak{a})d^{1/2}.$$

Furthermore $\mathrm{N}(\mathfrak{a}^{-1}) = \mathrm{N}(\mathfrak{a})^{-1}$. So $x_{\mathfrak{a}}.A$ intersects $X_{\geq H}$ if and only if $N(\lambda\mathfrak{a}^{-1}) \leq \frac{1}{2}H^{-2}d^{1/2}$ for some $\lambda \in \mathfrak{a}$ so that $\lambda\mathfrak{a}^{-1} \subset \mathcal{O}_d$. □

Corollary 7.3. *The set of connected components of $\mathcal{G}_d \cap X_{\geq H}$ injects[2] into the set of proper $\mathcal{O}_d$-ideals $\mathfrak{b} \subset \mathcal{O}_d$ such that $N(\mathfrak{b}) \leq \frac{1}{2}H^{-2}d^{1/2}$.*

Proof. Let $H > 1$. In a unimodular lattice $L \in X_{\geq H}$ there is, up to sign, only one primitive non-zero vector of length $\leq H^{-1}$. More precisely, if there were two primitive vectors $x, y \in L$ of length less than H^{-1}, then

$$\mathrm{vol}(L) \leq ||x|| \cdot ||y|| \leq H^{-2} < 1$$

contradicting the unimodularity of L. The maps are given as follows. This shows that if we fix $\mathfrak{a}$ as above, then by the proof of the last proposition

[2]In fact we consider the set of connected components of $\mathcal{G}_d \cap X_{\geq H}$ up to identifying the A^+ and the A^- part.

each connected component of $\mathcal{G}_d \cap X_{\geq H}$ corresponds to a unique (up to sign) primitive element $\lambda \in \mathfrak{a}$ with $N(\lambda\mathfrak{a}^{-1}) \leq \frac{1}{2}d^{1/2}H^{-2}$. This defines the claimed injective map. □

Towards the proof of Duke's Theorem [ELMV12], we will need to show that for all $\varepsilon > 0$,

$$\mu_d(X_{\geq d^\varepsilon}) \to 0$$

as $d \to \infty$. In fact, a special case of the next proposition (in the case $H = d^\varepsilon$) shows that

$$\mu_d(X_{\geq d^\varepsilon}) \ll_\varepsilon d^\varepsilon.$$

Proposition 7.4. *For all $\varepsilon > 0$ and $H \geq 1$ we have*

$$\mu_d(X_{\geq H}) \ll_\varepsilon d^\varepsilon H^{-2}.$$

Proof. For any orbit in $\mathcal{G}_d$ the maximal height achieved is $\leq d^{\frac{1}{4}}$ as there is no ideal of norm less than 1. We show next that for $H > 1$ any connected component of $\mathcal{G}_d \cap X_{\geq H}$ has length $3\log(d)$. Indeed, such a connected component corresponds to the segment of some oriented geodesic circle whose points have imaginary part between H^2 and $d^{\frac{1}{2}}$. More precisely, we want to bound the length of the geodesic segment between two points $z_1 = (x_1, H^2)$ and $z_2 = (x_2, H^2)$ in $\mathbb{H}$, where we choose x_1 and x_2 such that the geodesic arc connecting z_1 and z_2 stays below $d^{\frac{1}{2}}$. This then shows that $|x_1 - x_2| \leq 2d^{1/2}$ by Pythagoras. Thus, using the hyperbolic distance formula

$$\begin{aligned}
d_{\mathbb{H}}(z_1, z_2) &= 2\log\left(\frac{\sqrt{(x_2-x_1)^2} + \sqrt{(x_2-x_1)^2 + 4H^4}}{2H^2}\right) \\
&\leq 2\log\left(\frac{2d^{1/2} + 2\sqrt{2}d^{1/2}}{2H^2}\right) \\
&\leq 2\log(\frac{(1+\sqrt{2})d^{\frac{1}{2}}}{H^2}) \\
&\leq 2\log(1+\sqrt{2}) + 2\log(d^{1/2}) - \ln(H^2) \\
&\leq 2\log(1+\sqrt{2}) + 2\log(d^{1/2}) \\
&\leq 3\log(d)
\end{aligned}$$

for $d \geq 3$ and hence for all d as we only consider non-square discriminants. Together with Corollary 7.3,

$$\text{length}(\mathcal{G}_d \cap X_{\geq H}) \leq 3\log(d) N_{\leq H}(d)$$

for $N_{\leq H}(d)$ being the number of proper ideals $\mathfrak{a} \subset \mathcal{O}_d$ of norm $N(\mathfrak{a}) \leq \frac{1}{2}H^{-2}d^{\frac{1}{2}}$. Recall that for any $n \in \mathbb{N}$ the number of proper ideals in $\mathcal{O}_d$ of norm equal to n can be bounded by the squaring number of divisors of n

and so by $\ll_\varepsilon n^\varepsilon$. By summing over all $1 \leq n \leq \frac{1}{2}H^{-2}d^{1/2}$ we conclude

$$N_{\leq H}(d) \ll_\varepsilon \sum_{1\leq n\leq \frac{1}{2}H^{-2}d^{1/2}} n^\varepsilon \ll_\varepsilon \frac{1}{2}H^{-2}d^{1/2}(\frac{1}{2}H^{-2}d^{1/2})^\varepsilon \ll_\varepsilon (H^{-2}d^{1/2})^{1+\varepsilon}.$$

So we see that

$$\text{length}(\mathcal{G}_d \cap X_{\geq H}) \ll_\varepsilon \log(d)(H^{-2}d^{\frac{1}{2}})^{1+\varepsilon}.$$

As $\log(d)$ is dominated by $d^{\frac{1}{2}}$ and as $H^{-2(1+\varepsilon)} \leq H^{-2}$ we get

$$\text{length}(\mathcal{G}_d \cap X_{\geq H}) \ll_\varepsilon H^{-2}d^{\frac{1}{2}(1+\varepsilon)}.$$

Moreover, a straightforward consequence of Dirichlet's Class Number formula is

$$\text{length}(\mathcal{G}_d) = |d|^{\frac{1}{2}+o(1)}.$$

This implies

$$\mu_d(X_{\geq H}) = \frac{\text{length}(\mathcal{G}_d \cap X_{\geq H})}{\text{length}(\mathcal{G}_d)} \ll_\varepsilon H^{-2}d^\varepsilon.$$

□

References

[Duk88] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. **92** (1988), no. 1, 73–90. ↑2

[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. (2) **58** (2012), no. 3-4, 249–313. ↑1, 2, 15, 16, 19

[EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259, Springer-Verlag London, Ltd., London, 2011. ↑5, 8, 9

[Kog] C. Kogler, *Number Theory and Group Theory with a view towards Homogeneous Dynamics*. In preparation, Available at `http://www.constantinkogler.com/Files/SemesterpaperCK.pdf`. ↑11, 15

[Kog18] C. Kogler, *Closed Geodesics on Compact Hyperbolic Surfaces*, 2018. Bachelor Thesis, Available at `http://www.constantinkogler.com/Files/ConstantinKoglerBachelorThesis.pdf`. ↑7

[Lin68] Y. Linnik, *Ergodic properties of algebraic fields*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 45, Springer-Verlag New York Inc., New York, 1968. ↑2

[Neu92] J. Neukirch, *Algebraische Zahlentheorie*, Springer-Verlag, Berlin, 1992. ↑9, 10, 11