



# Special curves in modular surfaces

Matteo Tamiozzo

*Abstract.* We show that geodesics in  $\mathbf{H}$  attached to a maximal split torus or a real quadratic torus in  $GL_{2,\mathbf{Q}}$  are the only irreducible algebraic curves in  $\mathbf{H}$  whose image in  $\mathbf{R}^2$  via the  $j$ -invariant is contained in an algebraic curve.

## 1 Introduction

Maximal tori in  $GL_{2,\mathbf{Q}}$  arise from embeddings of étale  $\mathbf{Q}$ -algebras  $E$  of degree 2 in  $M_2(\mathbf{Q})$ . Such an algebra is either isomorphic to  $\mathbf{Q} \times \mathbf{Q}$ , or to an imaginary quadratic field  $K$ , or to a real quadratic field  $F$ . If  $E = K$ , the unique fixed point  $\tau$  of  $K^\times$  acting on the Poincaré upper half-plane  $\mathbf{H}$  has coordinates in  $\bar{\mathbf{Q}}$ , and the classical theory of complex multiplication tells us that the  $j$ -invariant  $j(\tau)$  also belongs to  $\bar{\mathbf{Q}}$  (more precisely, to an abelian extension of  $K$ ). Schneider proved that all the  $\bar{\mathbf{Q}}$ -points in  $\mathbf{H}$  whose  $j$ -invariant is an algebraic number are obtained in this way [Sch36]; they are called *special points*.

If  $E = \mathbf{Q} \times \mathbf{Q}$  or  $E = F$ , one can instead attach to it a geodesic in  $\mathbf{H}$ , which we call a *special geodesic*. It is the unique geodesic in  $\mathbf{H}$  whose endpoints in  $\mathbf{P}^1(\mathbf{R})$  are fixed by the action of  $E^\times$ . We show the following properties of special geodesics.

- (1) They are the only geodesics in  $\mathbf{H}$  containing infinitely many special points; dually, special points are the only ones belonging to infinitely many special geodesics.
- (2) They are *weakly bialgebraic* curves in  $\mathbf{H}$ , i.e., they are irreducible algebraic curves in  $\mathbf{H} \simeq \{(x, y) \in \mathbf{R}^2 \mid y > 0\}$ , and their projection to any modular curve, seen as a surface over  $\mathbf{R}$ , is contained in an algebraic curve.
- (3) They are the only weakly bialgebraic curves in  $\mathbf{H}$ .
- (4) An irreducible curve in  $\mathbf{R}^2$  containing infinitely many images via the  $j$ -invariant of special points must contain the image of a special geodesic.

In the last part of the document, we describe the Zariski closure of the image of a special geodesic in the modular curve with full level. We give a condition implying that the image of such a geodesic is algebraic, and we examine some examples in which this property fails: for a geodesic attached to a real quadratic field, nontriviality of the class group turns out to be related with this failure.

---

Received by the editors July 30, 2021; revised November 19, 2021, accepted November 19, 2021.

Published online on Cambridge Core December 2, 2021.

The author's research is supported by the ERC Grant 804176.

AMS subject classification: 11F03, 11J89.

Keywords: Geodesics, modular curve,  $j$ -invariant, functional transcendence.

Similar results have been studied and established in great generality for subvarieties of Shimura varieties seen as *complex* algebraic varieties: we refer the reader to the survey [KUY18] and the references therein. The main observation in this note is that in the simple case of modular curves, regarding the relevant spaces as real algebraic varieties instead, one still obtains a group-theoretic characterization of (weakly) bialgebraic subvarieties—although the groups which appear are not those classically considered in the theory of Shimura varieties.

This being said, the heart of our proof of the facts listed above rests on a “base change” argument allowing to reduce ourselves to a similar problem on the product of two (complex) modular curves, for which (an analogue of) the Ax–Lindemann–Weierstrass theorem and the André–Oort conjecture are known.

The same idea can be used to characterize irreducible algebraic curves in  $\mathbf{R}^2$  whose image via the exponential map is algebraic; we treat this simpler case first at the beginning of the document. Finally, let us mention that geodesics in modular curves have recently been used for arithmetic purposes in [DV21, Ric21].

## 1.1 Notations and conventions

By a subvariety of an algebraic variety, we always mean a *closed* subvariety. We will often tacitly identify complex algebraic varieties with their complex points—but we will be careful to distinguish between real algebraic varieties and their real points. We will denote by  $\mathbf{Z}_{>0}$  (resp.  $\mathbf{R}_{>0}$ ) the set of positive integers (resp. positive real numbers).

## 2 The exponential map

### 2.1 The complex situation

Let us consider the exponential

$$\begin{aligned} \exp : \mathbf{C} &\rightarrow \mathbf{C}^\times \\ z &\mapsto e^{2\pi iz} \end{aligned}$$

and, for every  $n \geq 1$ , the map  $\exp^n : \mathbf{C}^n \rightarrow (\mathbf{C}^\times)^n$ , which equals  $\exp$  on each component. We identify the source (resp. target) with the set of  $\mathbf{C}$ -points of  $\mathbf{A}_{\mathbf{C}}^n$  (resp.  $\mathbf{G}_{m,\mathbf{C}}^n$ ). As in [KUY18, Definition 4.3], an irreducible algebraic variety  $V \subset \mathbf{A}_{\mathbf{C}}^n$  (resp.  $W \subset \mathbf{G}_{m,\mathbf{C}}^n$ ) is called *bialgebraic* if  $\exp^n(V)$  is algebraic (resp. each analytic irreducible component of the preimage of  $W$  via  $\exp^n$  is algebraic).

**Theorem 2.1.1** (Ax–Lindemann–Weierstrass [BT20, Corollary 4.1.2 and Theorem 4.1.3])

- (1) *An irreducible algebraic variety  $V \subset \mathbf{A}_{\mathbf{C}}^n$  (resp.  $W \subset \mathbf{G}_{m,\mathbf{C}}^n$ ) is bialgebraic if and only if it is a translate of a  $\mathbf{C}$ -vector subspace of  $\mathbf{C}^n$  defined over  $\mathbf{Q}$  (resp. a translate of a subtorus of  $\mathbf{G}_{m,\mathbf{C}}^n$ ).*
- (2) *Let  $V \subset \mathbf{A}_{\mathbf{C}}^n$  be an irreducible algebraic variety. The Zariski closure of  $\exp^n(V)$  is bialgebraic.*

## 2.2 The real situation

We are now interested in studying the situation described in Section 2.1 for  $n = 1$ , regarding the relevant spaces as *real* rather than *complex* varieties. Precisely, let us consider the algebraic group  $\text{Res}_{\mathbf{C}/\mathbf{R}}\mathbf{G}_{m,\mathbf{C}} = \text{Spec } \mathbf{R}[X, Y] \left[ \frac{1}{X^2+Y^2} \right]$ , which will be denoted by  $G$  in this section. We identify  $\mathbf{R}^2 = \mathbf{A}^2(\mathbf{R})$  with  $\mathbf{C}$  via the map sending  $(x, y)$  to  $x + iy$ . We consider the map

$$E : \mathbf{R}^2 \rightarrow G(\mathbf{R}) = \mathbf{R}^2 \setminus \{0\}$$

$$(x, y) \mapsto (\text{Re } \exp(x + iy), \text{Im } \exp(x + iy)).$$

Our aim is to determine irreducible algebraic subvarieties of  $\mathbf{A}_{\mathbf{R}}^2$  whose image via  $E$  is an algebraic subvariety of  $G$ .

**Example 2.2.1** For every  $t \in \mathbf{R}$ , let  $L_t = \{(x, t), x \in \mathbf{R}\}$ . This is an irreducible algebraic subvariety of  $\mathbf{R}^2$ , whose image

$$E(L_t) = \{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 = e^{-2\pi t}\}$$

is also algebraic.

A vertical line  $S_t = \{(t, y), y \in \mathbf{R}\}$  has image the half-line  $R_t = \{re^{2\pi it}, r > 0\} \subset G(\mathbf{R})$ , which is semi-algebraic but not algebraic.

**Definition 2.2.2** A subset  $\mathcal{V} \not\subseteq \mathbf{R}^2$  is called *strongly bialgebraic* if it satisfies the following conditions:

- (1) There exist algebraic subvarieties  $V \subset \mathbf{A}_{\mathbf{R}}^2, W \subset G$  such that  $\mathcal{V} = V(\mathbf{R})$  and  $E(\mathcal{V}) = W(\mathbf{R})$ .
- (2)  $\mathcal{V}$  cannot be written in the form  $\mathcal{V} = V_1(\mathbf{R}) \cup V_2(\mathbf{R})$ , where  $V_1, V_2 \subset \mathbf{A}_{\mathbf{R}}^2$  are algebraic subvarieties, and the inclusions  $V_i(\mathbf{R}) \subset \mathcal{V}$  are proper for  $i = 1, 2$ .

We call  $\mathcal{V} \not\subseteq \mathbf{R}^2$  *weakly bialgebraic* if it satisfies the following conditions:

- (1) There exist algebraic subvarieties  $V \subset \mathbf{A}_{\mathbf{R}}^2, W \not\subseteq G$  such that  $\mathcal{V} = V(\mathbf{R})$  and  $E(\mathcal{V}) \subset W(\mathbf{R})$ .
- (2)  $\mathcal{V}$  cannot be written in the form  $\mathcal{V} = V_1(\mathbf{R}) \cup V_2(\mathbf{R})$ , where  $V_1, V_2 \subset \mathbf{A}_{\mathbf{R}}^2$  are algebraic subvarieties, and the inclusions  $V_i(\mathbf{R}) \subset \mathcal{V}$  are proper for  $i = 1, 2$ .

**Remark 2.2.3** The notion of strongly (resp. weakly) bialgebraic subset we introduced mirrors the definition given in the complex setting in [KUY18, Definition 4.3] (resp. [BT20, Definition 1.2.8])—as in [KUY18], we require irreducibility in our definition. As Example 2.2.1 shows, these two notions do not coincide in our situation: every point in  $\mathbf{R}^2$  is strongly bialgebraic, and the lines  $L_t$  are strongly bialgebraic. On the other hand, the lines  $S_t$  are weakly bialgebraic but not strongly bialgebraic. We will now show that there are no other examples of weakly bialgebraic subsets  $\mathcal{V} \not\subseteq \mathbf{R}^2$ .

**Theorem 2.2.4** Let  $\mathcal{V} \subset \mathbf{R}^2$  be a weakly bialgebraic subset which is not a singleton. Then, either  $\mathcal{V} = L_t$  for some  $t \in \mathbf{R}$  or  $\mathcal{V} = S_t$  for some  $t \in \mathbf{R}$ .

**Proof** *Step 0: Preliminaries.* Take  $\mathcal{V} \subset \mathbf{R}^2$  weakly bialgebraic, and suppose that it is not a point. By assumption,  $\mathcal{V}$  is the set of common real zeros of finitely many nonconstant polynomials in  $\mathbf{R}[X, Y]$ . The sum of their squares  $S \in \mathbf{R}[X, Y]$  is a polynomial whose set of real zeros is  $\mathcal{V}$ . As  $\mathcal{V}$  is irreducible, by definition, it is the vanishing locus of an irreducible factor  $P \in \mathbf{R}[X, Y]$  of  $S$ . Furthermore, the set  $\mathcal{V}$  is infinite; hence, such a  $P$  is irreducible in  $\mathbf{C}[X, Y]$  as well. In addition, there exists a smooth point  $(x, y) \in \mathcal{V}$ , and a neighborhood of  $(x, y)$  in  $\mathcal{V}$  is diffeomorphic to the open interval  $(0, 1)$ . Finally, by assumption, there exists a nonconstant polynomial  $Q \in \mathbf{R}[X, Y]$  such that  $E(\mathcal{V}) \subset \{(x, y) \in \mathbf{R}^2 \setminus \{0\} \mid Q(x, y) = 0\}$ .

*Step 1: Base change.* For  $(x, y) \in \mathbf{R}^2$ , we have  $\overline{\exp(x + iy)} = \exp(-x + iy)$ ; hence,

$$\begin{aligned} \operatorname{Re} \exp(x + iy) &= \frac{\exp(x + iy) + \exp(-x + iy)}{2}, \\ \operatorname{Im} \exp(x + iy) &= \frac{\exp(x + iy) - \exp(-x + iy)}{2i}. \end{aligned}$$

We introduce the maps

$$\begin{aligned} f : \mathbf{C}^2 &\rightarrow \mathbf{C}^2 & g : \mathbf{C}^2 &\rightarrow \mathbf{C}^2 \\ (v, w) &\mapsto (v + iw, -v + iw) & (a, b) &\mapsto \left( \frac{a + b}{2}, \frac{a - b}{2i} \right), \end{aligned}$$

so that the restriction of  $g \circ \exp^2 \circ f$  to  $\mathbf{R}^2 \subset \mathbf{C}^2$  equals  $E$ . Let us denote by  $C_P$  (resp.  $C_Q$ ) the complex plane curve with equation  $P = 0$  (resp.  $Q = 0$ ). We know that  $Q \circ E : \mathbf{R}^2 \rightarrow \mathbf{R}$  vanishes on  $\mathcal{V}$ ; hence, the same is true for the map  $h = Q \circ g \circ \exp^2 \circ f : \mathbf{C}^2 \rightarrow \mathbf{C}$ . It follows that the vanishing locus of the holomorphic function  $h$  restricted to the complex curve  $C_P$  contains a subset diffeomorphic to  $(0, 1)$ . As  $C_P$  is irreducible, it is connected in the Euclidean topology; hence,  $h$  must vanish identically on  $C_P$ .

*Step 2: Application of Ax–Lindemann–Weierstrass.* The outcome of the previous step is that the image of  $C_P$  via the map  $g \circ \exp^2 \circ f$  is contained in  $C_Q$ . In other words,  $\exp^2(f(C_P)) \subset g^{-1}(C_Q)$ . Hence, the Zariski closure  $Z \subset \mathbf{G}_{m, \mathbf{C}}^2$  of  $\exp^2 \circ f(C_P)$  is contained in  $g^{-1}(C_Q)$ . It follows from Theorem 2.2.1 that  $Z$  is bialgebraic, and it is a translate of a one-dimensional subtorus of  $\mathbf{G}_{m, \mathbf{C}}^2$ . The curve  $f(C_P)$  is connected, because  $C_P$  is; hence,  $f(C_P)$  is contained in a translate of a one-dimensional subspace of  $\mathbf{C}^2$  defined over  $\mathbf{Q}$ .

*Step 3: The final computation.* By construction,  $\mathcal{V}$  consists of the real points of  $C_P$ , so  $f(\mathcal{V}) \subset f(C_P)$ . A line  $L \subset \mathbf{A}_{\mathbf{C}}^2$  with slope in  $\mathbf{Q} \cup \{\infty\}$  is either vertical or has equation  $Y = rX + \alpha$  with  $r \in \mathbf{Q}$  and  $\alpha = a + ib \in \mathbf{C}$ . Notice that  $f(C_P)$  cannot be contained in a vertical line: indeed, in this case, there would be  $\alpha = a + ib \in \mathbf{C}$  such that every point  $(x, y) \in \mathcal{V}$  satisfies  $x + iy = a + ib$ , contradicting the assumption that  $\mathcal{V}$  is not a point. Now, assume that  $L : Y = rX + \alpha$  is not a vertical line and  $f(C_P) \subset L$ ; then, every point  $(x, y) \in \mathcal{V}$  satisfies

$$-x + iy = r(x + iy) + (a + ib) \Rightarrow (r + 1)x = -a, (r - 1)y = -b.$$

As we are supposing  $\mathcal{V}$  not to be a point, we must have either  $r = -1$ ; hence,  $y = \frac{b}{2}$ , or  $r = 1$  and  $x = -\frac{a}{2}$ . Therefore,  $\mathcal{V}$  is one of the lines considered in Example 2.2.1. ■

### 2.2.5 Special points and bialgebraicity for the exponential

By the Gelfond–Schneider theorem, if  $\alpha \in \bar{\mathbf{Q}}$  is not rational, then  $e^{2\pi i\alpha}$  is transcendental. In other words, the only points  $(x, y) \in \mathbf{R}^2$  with algebraic coordinates and such that  $E(x, y)$  has algebraic coordinates are those with  $x \in \mathbf{Q}$  and  $y = 0$ . The image of such a point in  $G(\mathbf{R})$  is a torsion point—i.e., a root of unity—which will be called a *special point*. With this terminology, Theorem 2.2.4 has the following consequence.

**Corollary 2.2.6** *Let  $W \subsetneq G$  be an irreducible subvariety with more than one real point.*

- (1) *The set  $W(\mathbf{R})$  is the image of a strongly bialgebraic subset of  $\mathbf{R}^2$  if and only if it is a  $G(\mathbf{R})$ -translate of the unit circle  $S^1(\mathbf{R})$ .*
- (2) *If  $W$  satisfies the equivalent conditions in (1), then  $W$  contains a special point if and only if it contains infinitely many special points.*
- (3) *If  $W(\mathbf{R})$  contains infinitely many special points, then  $W(\mathbf{R}) = S^1(\mathbf{R})$ .*

**Proof** The first point is a direct consequence of Theorem 2.2.4, and the second point follows immediately. Finally, the third point follows from the fact that the Zariski closure of an infinite set of roots of unity in  $G$  is  $S^1$ . ■

**Remark 2.2.7** The third point in the above corollary is an analogue in our situation of the Manin–Mumford conjecture for complex tori. Observe that the unique irreducible subvariety of  $G$  containing infinitely many special points arises from the embedding of groups  $S^1 \subset G$ ; hence, one could regard it as a *special subvariety*. The image of the inclusion  $\mathbf{G}_m(\mathbf{R}) \subset G(\mathbf{R})$  instead coincides with the Zariski closure of the image via  $E$  of the weakly bialgebraic subvariety  $S_0$ ; however, it only contains finitely many special points.

## 3 The $j$ -invariant

### 3.1 Setup

In this section, we set  $G = GL_{2,\mathbf{Q}}$ ; if  $H$  is a subgroup of  $G(\mathbf{R})$ , we will denote by  $H^+ \subset H$  the subgroup of matrices with positive determinant. Let

$$\mathbf{H} = \{z = x + iy \in \mathbf{C} \mid y > 0\} \simeq G(\mathbf{R})/\mathbf{R}_{>0}O_2(\mathbf{R})$$

be the Poincaré upper half-plane, with hyperbolic metric  $\frac{dx^2+dy^2}{y^2}$ . We identify  $\mathbf{H}$  with a subset of the  $\mathbf{R}$ -points of  $\mathbf{A}_{\mathbf{R}}^2$ ; this allows us, in particular, to talk about algebraic subvarieties of  $\mathbf{H}$ . For each congruence subgroup  $\Gamma \subset SL_2(\mathbf{Z})$ , the quotient  $\Gamma \backslash \mathbf{H}$  is a Riemann surface, which is the analytification of an algebraic curve over  $\mathbf{C}$  whose Weil restriction to  $\mathbf{R}$  will be denoted by  $Y_\Gamma$ . Hence,  $Y_\Gamma$  is a real surface with  $Y_\Gamma(\mathbf{R}) \simeq \Gamma \backslash \mathbf{H}$ ; we will call it the *modular surface of level  $\Gamma$* . Let  $p_\Gamma : \mathbf{H} \rightarrow Y_\Gamma(\mathbf{R})$  be the projection map.

**Remark 3.1.1** As a side remark, let us notice that the varieties  $Y_\Gamma$  admit models over number fields. Indeed, each Riemann surface  $\Gamma \backslash \mathbf{H}$  has a (canonical) model

$\tilde{Y}_{\Gamma, \mathbf{Q}(\zeta_n)}$  over a suitable cyclotomic field  $\mathbf{Q}(\zeta_n)$ , which we may take to be different from  $\mathbf{Q}$ . Let  $\mathbf{Q}(\zeta_n)^+ \subset \mathbf{Q}(\zeta_n)$  be the maximal totally real subfield. Then, we claim that  $Y_{\Gamma, \mathbf{Q}(\zeta_n)^+} = \text{Res}_{\mathbf{Q}(\zeta_n)/\mathbf{Q}(\zeta_n)^+} \tilde{Y}_{\Gamma, \mathbf{Q}(\zeta_n)}$  is a model of  $Y_{\Gamma}$ . Indeed, let  $S$  be an  $\mathbf{R}$ -scheme. We will denote by  $S_{|\mathbf{Q}(\zeta_n)^+}$  (resp.  $(S \times_{\mathbf{R}} \mathbf{C})_{|\mathbf{Q}(\zeta_n)}$ ) the scheme  $S$  (resp.  $S \times_{\mathbf{R}} \mathbf{C}$ ) seen as a  $\mathbf{Q}(\zeta_n)^+$ -scheme (resp.  $\mathbf{Q}(\zeta_n)$ -scheme). We have canonical identifications

$$\begin{aligned} \text{Hom}_{\mathbf{R}}(S, Y_{\Gamma, \mathbf{Q}(\zeta_n)^+} \times_{\mathbf{Q}(\zeta_n)^+} \mathbf{R}) &= \text{Hom}_{\mathbf{Q}(\zeta_n)^+}(S_{|\mathbf{Q}(\zeta_n)^+}, Y_{\Gamma, \mathbf{Q}(\zeta_n)^+}) \\ &= \text{Hom}_{\mathbf{Q}(\zeta_n)}(S_{|\mathbf{Q}(\zeta_n)^+} \times_{\mathbf{Q}(\zeta_n)^+} \mathbf{Q}(\zeta_n), \tilde{Y}_{\Gamma, \mathbf{Q}(\zeta_n)}) \\ &= \text{Hom}_{\mathbf{Q}(\zeta_n)}((S \times_{\mathbf{R}} \mathbf{C})_{|\mathbf{Q}(\zeta_n)}, \tilde{Y}_{\Gamma, \mathbf{Q}(\zeta_n)}) \\ &= \text{Hom}_{\mathbf{C}}(S \times_{\mathbf{R}} \mathbf{C}, \tilde{Y}_{\Gamma, \mathbf{Q}(\zeta_n)} \times_{\mathbf{Q}(\zeta_n)} \mathbf{C}) \\ &= \text{Hom}_{\mathbf{R}}(S, Y_{\Gamma}). \end{aligned}$$

**Definition 3.1.2** Let  $\Gamma \subset SL_2(\mathbf{Z})$  be a congruence subgroup. A subset  $\mathcal{V} \not\subset \mathbf{H}$  is called *weakly bialgebraic* if it satisfies the following conditions:

- (1) There exist algebraic subvarieties  $V \subset \mathbf{A}_{\mathbf{R}}^2$ ,  $W \not\subset Y_{\Gamma}$  such that  $\mathcal{V} = V(\mathbf{R}) \cap \mathbf{H}$  and  $p_{\Gamma}(\mathcal{V}) \subset W(\mathbf{R})$ .
- (2)  $\mathcal{V}$  cannot be written in the form  $\mathcal{V} = (V_1(\mathbf{R}) \cap \mathbf{H}) \cup (V_2(\mathbf{R}) \cap \mathbf{H})$ , where  $V_1, V_2 \subset \mathbf{A}_{\mathbf{R}}^2$  are algebraic subvarieties, and the inclusions  $V_i(\mathbf{R}) \cap \mathbf{H} \subset \mathcal{V}$  are proper for  $i = 1, 2$ .

**Remark 3.1.3** Observe that the fact that a subset  $\mathcal{V} \subset \mathbf{H}$  is weakly bialgebraic does not depend on  $\Gamma$ : assume that there exists  $\Gamma \subset SL_2(\mathbf{Z})$  congruence subgroup and  $W \not\subset Y_{\Gamma}$  such that  $\mathcal{V} = V(\mathbf{R}) \cap \mathbf{H}$  and  $p_{\Gamma}(\mathcal{V}) \subset W(\mathbf{R})$ . Let  $\Gamma' \subset SL_2(\mathbf{Z})$  be another congruence subgroup. Then,  $\Gamma \cap \Gamma'$  has finite index in  $\Gamma$ ; hence, the maps  $p_1 : Y_{\Gamma \cap \Gamma'} \rightarrow Y_{\Gamma}$  and  $p_2 : Y_{\Gamma \cap \Gamma'} \rightarrow Y_{\Gamma'}$  are finite. We have

$$p_{\Gamma'}(\mathcal{V}) = p_2 \circ p_{\Gamma \cap \Gamma'}(\mathcal{V}) \subset (p_2 \circ p_1^{-1}(W))(\mathbf{R}),$$

and  $p_2 \circ p_1^{-1}(W) \not\subset Y_{\Gamma'}$  is algebraic.

Our aim is to describe weakly bialgebraic subsets of  $\mathbf{H}$ . In view of the above remark, it suffices to consider the case  $\Gamma = SL_2(\mathbf{Z})$ , so that  $Y_{\Gamma} = \mathbf{A}_{\mathbf{R}}^2$  and  $p_{\Gamma} : \mathbf{H} \rightarrow \mathbf{R}^2$  is the map sending  $z$  to  $(\text{Re } j(z), \text{Im } j(z))$ . We will focus on this case in what follows, with the exception of Section 3.2.2 below. We start by describing two concrete examples illustrating the general phenomena we will later study.

### 3.2 Two examples

#### 3.2.1 The positive imaginary axis

The  $j$ -invariant is injective and takes real values on the half-line  $\{(x, y) \in \mathbf{R}^2, x = 0, y \geq 1\}$ , and  $j(it)$  goes to infinity as  $t > 0$  goes to infinity. Furthermore,  $j(i) = 1, 728$  and  $j(-1/z) = j(z)$  for  $z \in \mathbf{H}$ . It follows that the image via  $j$  of the vertical half-line  $\mathcal{C}_{0, \infty} = \{(x, y) \in \mathbf{R}^2 \mid x = 0, y > 0\} \subset \mathbf{H}$  is the semi-algebraic set  $\{(x, y) \in \mathbf{R}^2 \mid y = 0, x \geq 1728\}$ . In particular,  $\mathcal{C}_{0, \infty}$  is a weakly bialgebraic curve in  $\mathbf{H}$ . Notice that  $j(\mathcal{C}_{0, \infty})$  is not algebraic; the smallest algebraic set containing it is the axis  $Y = 0$  in

$\mathbf{R}^2$ . This is the image of the union  $\mathcal{C}_{0,\infty} \cup \partial^+$ , where  $\partial^+$  is the part of the boundary of the usual fundamental domain for the  $SL_2(\mathbf{Z})$ -action with positive first coordinate. Let us end this example by pointing out that  $\mathcal{C}_{0,\infty}$  is the unique geodesic in  $\mathbf{H}$  with endpoints  $0, \infty$ , and it contains infinitely many special points. It can also be described as the orbit of any of its points via the action of the group  $\left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, t \in \mathbf{R}^\times \right\}$  of matrices of determinant 1 in  $T(\mathbf{R})$ , where  $T \subset G$  is the diagonal torus.

### 3.2.2 Bernoulli's lemniscate

We learned the example we are going to explain now from the MathOverflow post [mo220], which was the starting point of our investigations.<sup>1</sup> Let us take  $\Gamma = \Gamma(2)$ , so that  $Y_{\Gamma(2)}(\mathbf{R}) = \mathbf{C} \setminus \{0, 1\}$  and the projection map is Klein's  $\lambda$ -function. Let us consider the matrix  $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ , and the curve  $\mathcal{C}_A \subset \mathbf{H}$  with equation  $Az = \bar{z}$ ; it is the half-circle with center  $(1, 0)$  and radius  $\sqrt{2}$ . In other words, it is the only geodesic in  $\mathbf{H}$  with endpoints  $1 \pm \sqrt{2} \in \mathbf{R}$ . It can be described in a slightly different way, which will be generalized later on: we have an embedding

$$\begin{aligned} \iota : F = \mathbf{Q}(\sqrt{2}) &\rightarrow M_2(\mathbf{Q}) \\ a + b\sqrt{2} &\mapsto a\text{Id} + bA, \end{aligned}$$

inducing a map  $\iota_{\mathbf{R}} : F_{\mathbf{R}}^1 = \{x \in F \otimes_{\mathbf{Q}} \mathbf{R} \mid N(x) = 1\} \rightarrow G(\mathbf{R})^+$ . The geodesic  $\mathcal{C}_A$  is the  $F_{\mathbf{R}}^1$ -orbit of any of its points. In particular,  $3\text{Id} + 2A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \in \Gamma(2) \cap F_{\mathbf{R}}^1$  fixes  $\mathcal{C}_A$ . It follows that the image of  $\mathcal{C}_A$  in  $Y_{\Gamma(2)}$  coincides with the image of the compact set  $F_{\mathbf{R}}^1/(\pm\epsilon^{\mathbf{Z}})$ , where  $\epsilon = 3 + 2\sqrt{2}$ ; therefore, it is a compact geodesic in  $Y_{\Gamma(2)}$ . In fact,  $p_{\Gamma(2)}(\mathcal{C}_A)$  is the lemniscate with equation  $(X^2 + Y^2)((X - 1)^2 + Y^2) = \frac{1}{16}$ . In particular,  $\mathcal{C}_A$  is bialgebraic. Finally, notice that, as in the previous example, the curve  $\mathcal{C}_A$  contains infinitely many special points.

### 3.3 Special geodesics

Geodesics in  $\mathbf{H}$  are either vertical lines or half-circles with center on the real axis. They can be described as follows: given  $A \in GL_2(\mathbf{R})$  with trace zero and hyperbolic (i.e., with negative determinant), let us consider the curve  $\mathcal{C}_A \subset \mathbf{H}$  with equation  $Az = \bar{z}$ . Writing  $z = x + iy$  and  $A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ , we see that points  $x + iy \in \mathcal{C}_A$  satisfy the equation

$$\frac{az + b}{cz - a} = \bar{z} \Leftrightarrow c|z|^2 - 2a\text{Re}z - b = 0 \Leftrightarrow c(x^2 + y^2) - 2ax - b = 0.$$

Hence,  $\mathcal{C}_A$  is the unique geodesic with endpoints the fixed points of  $A$  in  $\mathbf{P}^1(\mathbf{R})$ . Every geodesic in  $\mathbf{H}$  can be obtained in this way: indeed, a nonvertical geodesic has an

<sup>1</sup>We thank Gregorio Baldi for pointing it out to us.

equation of the form  $(X - x_0)^2 + Y^2 = r$ , and it suffices to choose  $a, b, c$  such that  $\frac{a}{c} = x_0$  and  $\frac{bc+a^2}{c^2} = r$ . A vertical geodesic has equation  $X = x_0$ , so we may choose  $a = 1, b = -2x_0$ , and  $c = 0$ .

**Remark 3.3.1** An alternative way to think about geodesics in  $\mathbf{H}$  is the following: for every  $A \in GL_2(\mathbf{R})$  with trace zero and negative determinant, the map  $z \mapsto A\bar{z}$  is an anti-holomorphic involution of  $\mathbf{H}$ , and the curve  $\mathcal{C}_A$  is the set of fixed points of this involution. This is the point of view adopted by Jaffee [Jaf74], who first suggested to use this idea to study the arithmetic properties of symmetric spaces attached to certain reductive groups when they do not carry a complex structure.

### 3.3.2 Geodesics attached to real quadratic fields

Given a real quadratic field  $F = \mathbf{Q}(\sqrt{d})$  and an embedding  $\iota : F \rightarrow M_2(\mathbf{Q})$ , the element  $\iota(\sqrt{d}) \in GL_2(\mathbf{R})$  has trace zero and is hyperbolic. The geodesic attached to  $\iota(\sqrt{d})$  via the previous construction will be denoted by  $\mathcal{C}_{F,\iota}$ . This construction has the following group-theoretic interpretation (which in particular shows that  $\mathcal{C}_{F,\iota}$  only depends on  $F$  and  $\iota$ , and not on the choice of  $d$ ). Consider the torus  $T_F = \text{Res}_{F/\mathbf{Q}}\mathbf{G}_{m,F}$ ; the embedding  $\iota$  induces an embedding of groups, abusively denoted by the same symbol,  $\iota : T_F \hookrightarrow G$ , which in turn induces an map

$$T_F(\mathbf{R})/\mathbf{R}^\times \rightarrow G(\mathbf{R})/\mathbf{R}^\times.$$

There are two points in  $\mathbf{R} \subset \mathbf{P}^1(\mathbf{R})$  fixed by  $T_F(\mathbf{Q})$ , and  $\mathcal{C}_{F,\iota}$  is the unique geodesic in  $\mathbf{H}$  having these endpoints. The image of the subgroup  $T_F(\mathbf{R})^+/\mathbf{R}^\times$  via the above map acts on  $\mathbf{H}$ ; the geodesic  $\mathcal{C}_{F,\iota}$  is the  $T_F(\mathbf{R})^+/\mathbf{R}^\times$ -orbit of any of its points.

- Remark 3.3.3** (1) In particular, in the above construction of  $\mathcal{C}_{F,\iota}$  we may choose  $d$  such that  $A = \iota(\sqrt{d})$  belongs to  $M_2(\mathbf{Z})$ , and the greatest common divisor of the coefficients of  $A$  is 1.
- (2) All the embeddings of  $F$  in  $M_2(\mathbf{Q})$  are conjugate; hence, the resulting geodesics are translates of a given one via the action of  $G(\mathbf{Q})^+$ .

### 3.3.4 Geodesics attached to split tori

Another distinguished class of geodesics consists of vertical geodesics with rational first coordinate and half-circles with rational endpoints on the real axis (which give rise to modular symbols). In group-theoretic terms, those arise from embeddings of the split torus  $T = \mathbf{G}_{m,\mathbf{Q}} \times \mathbf{G}_{m,\mathbf{Q}}$  in  $G$ , as follows: letting  $\iota : T \hookrightarrow G$  be an embedding, there is a unique geodesic  $\mathcal{C}_{\mathbf{Q} \times \mathbf{Q},\iota}$  in  $\mathbf{H}$  with endpoints the points of  $\mathbf{P}^1(\mathbf{R})$  fixed by  $\iota(T(\mathbf{Q}))$ . For example, if  $\delta : T \hookrightarrow G$  is the diagonal embedding, then the group  $\delta(T(\mathbf{Q}))$  fixes the points  $0, \infty \in \mathbf{P}^1(\mathbf{R})$ , and the geodesic  $\mathcal{C}_{\mathbf{Q} \times \mathbf{Q},\delta}$  is the vertical line with equation  $X = 0$ . As above, this is also the  $T(\mathbf{R})^+/\mathbf{R}^\times$ -orbit of any of its points. Any other embedding  $\iota : T \rightarrow G$  is  $G(\mathbf{Q})^+$ -conjugate to  $\delta$ ; hence, the attached geodesic is a  $G(\mathbf{Q})^+$ -translate of  $\mathcal{C}_{\mathbf{Q} \times \mathbf{Q},\delta}$ . This construction accounts for all the geodesics in  $\mathbf{H}$  whose endpoints in  $\mathbf{P}^1(\mathbf{R})$  are rational.



**Definition 3.3.5** A geodesic in  $\mathbf{H}$  obtained from an embedding of  $T$  or  $T_F$  for some real quadratic field  $F$  via the above construction is called a *special geodesic*.

**Remark 3.3.6** Concretely, special geodesics in  $\mathbf{H}$  have equation  $Az = \bar{z}$  for some hyperbolic trace-zero matrix  $A \in M_2(\mathbf{Q})$  (which can be rescaled so as to have integral coprime coefficients). Such a geodesic is attached to a split torus (resp. to a torus coming from a real quadratic field) if the absolute value of the determinant of  $A$  is (resp. is not) a square in  $\mathbf{Q}$ . Notice that special points satisfy instead equations of the form  $Az = z$  with  $A \in M_2(\mathbf{Q})$  elliptic trace-zero matrix.

Observe that every geodesic in  $\mathbf{H}$  is an algebraic curve. The following proposition gives a first characterization of special geodesics among all geodesics in  $\mathbf{H}$ .

**Proposition 3.3.7** Let  $\mathcal{C} \subset \mathbf{H}$  be a geodesic. The following are equivalent:

- (1)  $\mathcal{C}$  is special.
- (2)  $\mathcal{C}$  contains infinitely many special points.
- (3)  $\mathcal{C}$  is defined over  $\mathbf{Q}$ .
- (4) The endpoints of  $\mathcal{C}$  in  $\mathbf{P}^1(\mathbf{R})$  are rational or quadratic conjugates.

**Proof** By construction, every special geodesic is defined over  $\mathbf{Q}$ ; furthermore, the endpoints of a geodesic defined by an equation with rational coefficients are either rational or quadratic conjugates. On the other hand, for any two rational or quadratic conjugate points of  $\mathbf{P}^1(\mathbf{R})$ , the unique geodesic having them as endpoints is defined over  $\mathbf{Q}$ . Now, a geodesic  $\mathcal{C}$  defined over  $\mathbf{Q}$  is either vertical with rational first coordinate—hence special—or a half-circle with equation  $(X - x_0)^2 + Y^2 = r$ , with  $x_0, r \in \mathbf{Q}$ ; in the latter case, we may choose  $a, b, c \in \mathbf{Q}$  such that  $\frac{a}{c} = x_0$  and  $\frac{bc+a^2}{c^2} = r$ , and set  $A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ . Then,  $\mathcal{C}$  has equation  $Az = \bar{z}$ , and by Remark 3.3.6, it is special. Hence, we have established the equivalence of (1), (3), and (4).

Finally, let us show that geodesics defined over  $\mathbf{Q}$  are precisely those containing infinitely many special points. This is clear for vertical geodesics. Now, take a geodesic  $\mathcal{C}$  with equation  $(X - x_0)^2 + Y^2 = r$ . If  $x_0, r$  are rational, then any point in  $\mathcal{C}$  with rational first coordinate is special. Conversely, assume that  $\mathcal{C}$  contains infinitely many special points. Then, there exist  $(x_1, y_1), (x_2, y_2) \in \mathbf{Q}^2$  with  $x_1 \neq x_2$  and  $y_1, y_2 > 0$ , and  $d_1, d_2 \in \mathbf{Z}_{>0}$  such that

$$(x_i - x_0)^2 + d_i y_i^2 = r, \quad i = 1, 2.$$

Taking the difference, we obtain

$$x_1^2 - x_2^2 + d_1 y_1^2 - d_2 y_2^2 - 2x_0(x_1 - x_2) = 0 \Rightarrow x_0 \in \mathbf{Q} \Rightarrow r \in \mathbf{Q},$$

so  $\mathcal{C}$  is defined over  $\mathbf{Q}$ . ■

**Remark 3.3.8** The computations in the above proof give the following more precise information.

- (1) A geodesic is special if and only if it contains at least two distinct special points—in which case it contains infinitely many. However, unlike in the classical case [KUY18, Theorem 3.5], containing one special point is not enough for a geodesic to be special—indeed, any half-circle with real irrational center passing through a special point is not defined over  $\mathbf{Q}$ .
- (2) A special point attached to an imaginary quadratic field  $K$  is contained in a special geodesic attached to a real quadratic field  $F = \mathbf{Q}(\sqrt{d})$  if and only if  $d$  is the norm of an element in  $K$ .

Let us also notice that the following “dual” version of the above proposition holds true.

**Proposition 3.3.9** *Let  $z \in \mathbf{H}$  be a point. The following assertions are equivalent:*

- (1)  $z$  is a special point.
- (2)  $z$  belongs to infinitely many special geodesics.

**Proof** Take  $z = x + iy \in \mathbf{H}$  special, so that  $x \in \mathbf{Q}$  and  $y \in \mathbf{R}_{>0}$  has rational square; let  $D = y^2 \in \mathbf{Q}$ . For every  $x_0 \in \mathbf{Q}$ , we have  $(x - x_0)^2 + y^2 = (x - x_0)^2 + D$ ; letting  $r = (x - x_0)^2 + D$ , we see that  $z$  belongs to the special geodesic with equation  $(X - x_0)^2 + Y^2 = r$ .

Conversely, assume that  $z = x + iy \in \mathbf{H}$  belongs to infinitely many special geodesics. Then, there are two couples of rational numbers  $(x_0, r), (x'_0, r')$  such that  $x_0 \neq x'_0$  and

$$(x - x_0)^2 + y^2 = r, \quad (x - x'_0)^2 + y^2 = r'.$$

Taking the difference, we deduce that  $x \in \mathbf{Q}$ ; hence,  $y^2 \in \mathbf{Q}$  and  $z$  is a special point. ■

### 3.4 The $j$ -invariant and bialgebraicity

We now wish to describe weakly bialgebraic subsets of  $\mathbf{H}$ . In view of Remark 3.1.3, it suffices to study weakly bialgebraic subsets for the map

$$J : \mathbf{H} \rightarrow \mathbf{R}^2, \\ z \mapsto (\operatorname{Re} j(z), \operatorname{Im} j(z)).$$

In order to do this, we will follow the same general strategy used in the proof of Theorem 2.2.4; in particular, we will make use of the Ax–Lindemann–Weierstrass theorem for the map

$$j^2 = j \times j : \mathbf{H}^2 \rightarrow \mathbf{C}^2,$$

due to Pila.

**Theorem 3.4.1 [Pil11]** *Let  $V \subset \mathbf{A}_{\mathbf{C}}^2$  be an irreducible algebraic variety. If there is a point  $P \in V \cap \mathbf{H}^2$  such that  $j^2$  maps an open neighborhood of  $P$  in  $V$  to an algebraic subvariety properly contained in  $\mathbf{C}^2$ , then one of the following assertions holds true:*

- (1)  $V$  is a point.

- (2)  $V \cap \mathbf{H}^2 = \{(x, y) \in \mathbf{H}^2 \mid y = c\}$  or  $V \cap \mathbf{H}^2 = \{(x, y) \in \mathbf{H}^2 \mid x = c\}$  for some  $c \in \mathbf{C}$ .
- (3)  $V \cap \mathbf{H}^2 = \{(x, y) \in \mathbf{H}^2 \mid y = Ax\}$  for some  $A \in G(\mathbf{Q})^+$ .

**Theorem 3.4.2** *Weakly bialgebraic subvarieties of  $\mathbf{H}$  are either points or special geodesics.*

**Proof** *Step 0: Preliminaries.* Take  $\mathcal{V} \subset \mathbf{H}$  weakly bialgebraic; then,  $\mathcal{V}$  is the set of zeros in  $\mathbf{H}$  of an irreducible polynomial  $P(X, Y) \in \mathbf{R}[X, Y]$ . Assume that  $\mathcal{V}$  is not a point; then, it is infinite; hence,  $P$  is irreducible in  $\mathbf{C}[X, Y]$ . There exists a smooth point  $x \in \mathcal{V}$ , and a neighborhood of  $x$  in  $\mathcal{V}$  is diffeomorphic to the open interval  $(0, 1)$ . Finally, by assumption, there exists a nonconstant polynomial  $Q \in \mathbf{R}[X, Y]$  such that  $J(\mathcal{V}) \subset \{(x, y) \in \mathbf{R}^2 \mid Q(x, y) = 0\}$ .

*Step 1: Base change.* Recall that for  $z \in \mathbf{C}$ , we have  $\overline{j(z)} = j(-\bar{z})$ . Hence, for  $(x, y) \in \mathbf{R} \times \mathbf{R}_{>0}$ , we have

$$\operatorname{Re} j(x + iy) = \frac{j(x + iy) + j(-x + iy)}{2}, \quad \operatorname{Im} j(x + iy) = \frac{j(x + iy) - j(-x + iy)}{2i}.$$

Consider the maps

$$\begin{aligned} f : \mathbf{C}^2 &\rightarrow \mathbf{C}^2 & g : \mathbf{C}^2 &\rightarrow \mathbf{C}^2 \\ (v, w) &\mapsto (v + iw, -v + iw) & (a, b) &\mapsto \left(\frac{a+b}{2}, \frac{a-b}{2i}\right); \end{aligned}$$

let  $U \subset \mathbf{C}^2$  be the set of points  $(v, w)$  such that the imaginary parts of  $v + iw, -v + iw$  are positive. In particular,  $\mathbf{R} \times \mathbf{R}_{>0}$  is contained in  $U$ , and the restriction of  $g \circ j^2 \circ f$  to  $\mathbf{R} \times \mathbf{R}_{>0}$  equals  $J$ . Let  $C_P$  (resp.  $C_Q$ ) be the complex plane curve with equation  $P = 0$  (resp.  $Q = 0$ ). The map  $h = Q \circ g \circ j^2 \circ f : U \rightarrow \mathbf{C}$  vanishes on  $\mathcal{V}$ ; because  $\mathcal{V}$  contains a subset  $I$  diffeomorphic to  $(0, 1)$ , the holomorphic map  $h$  must vanish identically on the connected component  $C$  of  $C_P \cap U$  containing  $I$ .

*Step 2: Application of Ax-Lindemann-Weierstrass.* The outcome of the previous step is that the image of  $C$  via the map  $g \circ j^2 \circ f$  is contained in  $C_Q$ . In other words,  $j^2(f(C))$  is contained in the algebraic subvariety  $g^{-1}(C_Q) \not\subseteq \mathbf{A}_{\mathbf{C}}^2$ . Therefore, we deduce that  $f(C_P)$  is a subvariety of  $\mathbf{A}_{\mathbf{C}}^2$  as in Theorem 3.4.1.

*Step 3: The final computation.* As  $f(\mathcal{V}) \subset f(C_P)$  and  $\mathcal{V}$  is not a point, we deduce that  $f(C_P) \cap \mathbf{H}^2$  has equation either  $X = c$ , or  $Y = c$ , or  $Y = AX$  for some  $A \in G(\mathbf{Q})^+$  and  $c \in \mathbf{C}$ . In the first case, we find that every  $(x, y) \in \mathcal{V}$  satisfies  $x + iy = c$ , contradicting the assumption that  $\mathcal{V}$  is not a point; similarly, the second case cannot occur. Therefore, we deduce that there exists  $A \in G(\mathbf{Q})^+$  such that every  $(x, y) \in \mathcal{V}$  satisfies

$$-x + iy = A(x + iy).$$

Let  $\tilde{A} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} A$ ; writing  $\tilde{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , the previous equation becomes

$$c(x^2 + y^2) + (d - a)x - (a + d)iy - b = 0.$$

In particular, we must have  $a = -d$ . By Remark 3.3.6, we deduce that every weakly bialgebraic set  $\mathcal{V}$  which is not a point must be a special geodesic.

Finally, special geodesics are indeed weakly bialgebraic, as explained below in Section 3.5. ■

**Remark 3.4.3** Let  $\mathcal{V} \subset \mathbf{H}$  be a weakly bialgebraic subset which is not a singleton. The above theorem tells us, in particular, that  $\mathcal{V}$  is an orbit for the action of a subgroup of  $PSL_2(\mathbf{R})$  arising from the group of real points of a split or real quadratic torus in  $G$ . Letting  $\iota : \mathbf{H} \rightarrow \mathbf{H} \times \mathbf{H}$  be the map sending  $z$  to  $(z, -\bar{z})$ , Step 3 of our argument rests on the observation that the pullback via  $\iota$  of a subvariety with equation  $Y = AX$  for  $A \in G(\mathbf{Q})^+$  turns out to be either empty or a geodesic attached to a torus as above. There is a more conceptual way to show that each connected component  $\mathcal{V}^0 \subset \mathcal{V}$  must be the orbit of a connected Lie subgroup of  $PSL_2(\mathbf{R})$ , which may be of use to generalize Theorem 3.4.2. With the same notations as in the proof, consider the commutative diagram

$$\begin{array}{ccc} \mathbf{H} & \xrightarrow{\iota} & \mathbf{H} \times \mathbf{H} \\ \downarrow J & & \downarrow g \circ j^2 \\ \mathbf{R}^2 & \longrightarrow & \mathbf{C}^2, \end{array}$$

where the bottom arrow is the natural inclusion. Steps 1 and 2 in the proof of Theorem 3.4.2 (which rely on Theorem 3.4.1) imply that  $f(C_P) \cap \mathbf{H}^2$  is a totally geodesic submanifold of the symmetric space  $\mathbf{H}^2$ . Therefore, the same is true for  $\iota^{-1}(f(C_P) \cap \mathbf{H}^2) = \mathcal{V}$ , so  $\mathcal{V}^0 \subset \mathbf{H}$  is a connected totally geodesic submanifold. It follows from the proof of [Hel78, Chapter IV, Theorem 7.2] that  $\mathcal{V}^0$  is the orbit of a connected Lie subgroup of  $PSL_2(\mathbf{R})$ —whose Lie algebra arises from a Lie triple system as in *loc. cit.*

### 3.5 Modular polynomials and special geodesics

#### 3.5.1 Modular curves in the modular surface

For  $N \geq 1$ , let  $\Phi_N(X, Y) \in \mathbf{Z}[X, Y]$  be the  $N$ th modular polynomial. Recall that  $\Phi_N(X, Y) = \Phi_N(Y, X)$  if  $N > 1$ , and  $\Phi_N$  is irreducible in  $\mathbf{C}(X)[Y]$ . It follows that  $\Phi_N$  is also irreducible in  $\mathbf{C}[X, Y]$ . Let  $\tilde{\Phi}_N(X, Y) = \Phi_N(X + iY, X - iY)$ . If  $N = 1$ , then  $\tilde{\Phi}_N(X, Y) = 2iY$ , whereas for  $N > 1$ , we have:

- (1)  $\tilde{\Phi}_N(X, Y) \in \mathbf{Z}[X, Y]$  and  $\tilde{\Phi}_N(X, -Y) = \tilde{\Phi}_N(X, Y)$ .
- (2)  $\tilde{\Phi}_N(X, Y)$  is irreducible in  $\mathbf{C}[X, Y]$ .

The first (resp. second) property follows from symmetry (resp. irreducibility) of  $\Phi_N$ . Let

$$\mathcal{Z}_N = \{\tilde{\Phi}_N = 0\} \subset \mathbf{A}_{\mathbf{R}}^2;$$

the above properties imply that  $\mathcal{Z}_N$  is a geometrically irreducible plane curve, symmetric with respect to the axis  $Y = 0$ .

Now, take  $A \in G(\mathbf{Q})$  with trace zero and negative determinant and consider the curve  $\mathcal{C}_A$  in  $\mathbf{H}$  with equation  $Az = \bar{z}$ . The curve  $\mathcal{C}_A$  is unchanged if we replace  $A$  by  $\lambda A$  for some  $\lambda \in \mathbf{Z}_{>0}$ ; hence, we may (and will) assume that  $A$  has integral coprime coefficients. Let  $N$  be the absolute value of the determinant of  $A$ ; if  $z$  is a point of  $\mathcal{C}_A$ , the elliptic curves attached to the lattices  $\mathbf{Z} \oplus z\mathbf{Z}$  and  $\mathbf{Z} \oplus \bar{z}\mathbf{Z}$  are related by an isogeny with cyclic kernel of cardinality  $N$ ; hence, we have  $\Phi_N(j(z), j(-\bar{z})) = 0$ . As  $j(-\bar{z}) = \overline{j(z)}$ , we see that  $J(\mathcal{C}_A) \subset \mathbf{R}^2$  is contained in the curve  $\mathcal{Z}_N$ . This shows that every special geodesic is weakly bialgebraic. Notice, however, that the curve  $\mathcal{Z}_N$  may contain distinct images of geodesics  $\mathcal{C}_A$ ; this is related to the existence of trace-zero matrices of determinant  $-N$  which are not  $SL_2(\mathbf{Z})$ -conjugate. We examine this phenomenon below.

### 3.5.2 The curve $\mathcal{Z}_1$

For  $N = 1$ , we obtain the curve  $Y = 0$ . If  $A \in M_2(\mathbf{Z})$  has trace zero and determinant  $-1$ , then  $J(\mathcal{C}_A)$  is contained in  $\mathcal{Z}_1$ . If we take  $A_1^1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ , then the curve  $\mathcal{C}_{A_1^1}$  is the positive imaginary axis. As observed in Example 3.2.1, its image is the closed half-line  $\{(x, y) \in \mathbf{R}^2 \mid y = 0, x \geq 1728\}$ . For  $A_2^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , the curve  $\mathcal{C}_{A_2^1}$  is the intersection of the unit circle with center at the origin and the upper half-plane. Notice that  $A_1^1$  and  $A_2^1$  are not  $SL_2(\mathbf{Z})$ -conjugate, and the images  $J(\mathcal{C}_{A_1^1})$  and  $J(\mathcal{C}_{A_2^1})$  do not coincide. One checks that every matrix  $A \in M_2(\mathbf{Z})$  with trace zero and determinant  $-1$  is  $SL_2(\mathbf{Z})$ -conjugate either to  $A_1^1$  or to  $A_2^1$ , and we have

$$\mathcal{Z}_1(\mathbf{R}) = J(\mathcal{C}_{A_1^1}) \cup J(\mathcal{C}_{A_2^1}).$$

### 3.5.3 The curve $\mathcal{Z}_N$ for $N$ square-free

Take  $N > 1$  square-free; the ring  $\mathbf{Z}[\sqrt{N}]$  is an order in the real quadratic field  $\mathbf{Q}(\sqrt{N})$ , and every matrix in  $M_2(\mathbf{Z})$  with determinant  $-N$  has coprime coefficients. Hence, for every  $SL_2(\mathbf{Z})$ -conjugacy class of matrices with integral coefficients, trace zero, and determinant  $-N$ , we obtain a curve  $J(\mathcal{C}_A) \subset \mathcal{Z}_N(\mathbf{R})$ , where  $A$  is any matrix in the given conjugacy class. The set of  $SL_2(\mathbf{Z})$ -conjugacy classes of matrices in  $M_2(\mathbf{Z})$  with trace zero and determinant  $-N$  is in bijection with the set  $Cl^+(\mathbf{Z}[\sqrt{N}])$  (not necessarily invertible) narrow ideal classes of  $\mathbf{Z}[\sqrt{N}]$  (see [Con, Remark 9]). Identifying these two sets, we have an inclusion

$$\bigcup_{A \in Cl^+(\mathbf{Z}[\sqrt{N}])} J(\mathcal{C}_A) \subset \mathcal{Z}_N(\mathbf{R}).$$

In fact, the above inclusion is an equality. Indeed, if  $(x, y) \in \mathbf{R}^2$  satisfies  $\Phi_N(x + iy, x - iy) = 0$ , writing  $x + iy = j(z)$ , we see that the elliptic curves attached to the lattices  $\mathbf{Z} \oplus z\mathbf{Z}$  and  $\mathbf{Z} \oplus (-\bar{z})\mathbf{Z}$  are related by an isogeny with cyclic kernel of

cardinality  $N$ ; hence, we must have  $-\bar{z} = \frac{az+b}{cz+d}$  for some matrix  $\tilde{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbf{Z})$  with determinant  $N$ . The matrix  $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tilde{A}$  has determinant  $-N$ , and  $Az = \bar{z}$ . Furthermore, as  $z \in \mathbf{H}$ , the trace of  $A$  must be zero, and  $(x, y)$  belongs to  $J(\mathcal{C}_A)$ .

Let us also observe that if  $A \in M_2(\mathbf{Z})$  has trace zero and determinant  $-N$ , then the same is true for  $-A$ , and  $\mathcal{C}_{-A} = \mathcal{C}_A$ . Letting  $\widetilde{Cl}^+(\mathbf{Z}[\sqrt{N}])$  be the quotient of  $Cl^+(\mathbf{Z}[\sqrt{N}])$  by the equivalence relation identifying  $A$  with  $-A$ , we obtain the following proposition.

**Proposition 3.5.4** *Let  $N > 1$  be a square-free integer, and  $\mathcal{Z}_N \subset \mathbf{A}_{\mathbf{R}}^2$  the curve with equation  $\tilde{\Phi}_N(X, Y) = 0$ . Then,*

$$\bigcup_{A \in \widetilde{Cl}^+(\mathbf{Z}[\sqrt{N}])} J(\mathcal{C}_A) = \mathcal{Z}_N(\mathbf{R}).$$

Furthermore, if  $A, B \in M_2(\mathbf{Z})$  are two matrices with trace zero, determinant  $-N$ , and distinct image in  $\widetilde{Cl}^+(\mathbf{Z}[\sqrt{N}])$ , then  $J(\mathcal{C}_A) \neq J(\mathcal{C}_B)$ .

**Proof** The first assertion follows from the discussion before the statement of the proposition. To prove the last assertion, notice that a geodesic  $Az = \bar{z}$  in  $\mathbf{H}$  is uniquely determined by its endpoints in  $\mathbf{P}^1(\mathbf{R})$ , which are the fixed points of  $A$ . It follows that two matrices with integral coefficients, trace zero, and determinant  $-N$  giving rise to the same geodesic in  $\mathbf{H}$  must be either equal or opposite. Take  $A, B \in M_2(\mathbf{Z})$  having distinct image in  $\widetilde{Cl}^+(\mathbf{Z}[\sqrt{N}])$ . Then, for every  $M \in SL_2(\mathbf{Z})$ , the geodesics  $M \cdot \mathcal{C}_A$  and  $\mathcal{C}_B$  are distinct; hence, they have at most one point of intersection. Therefore,  $J(\mathcal{C}_A) \neq J(\mathcal{C}_B)$ . ■

### 3.5.5 The curves $\mathcal{Z}_2$ and $\mathcal{Z}_3$

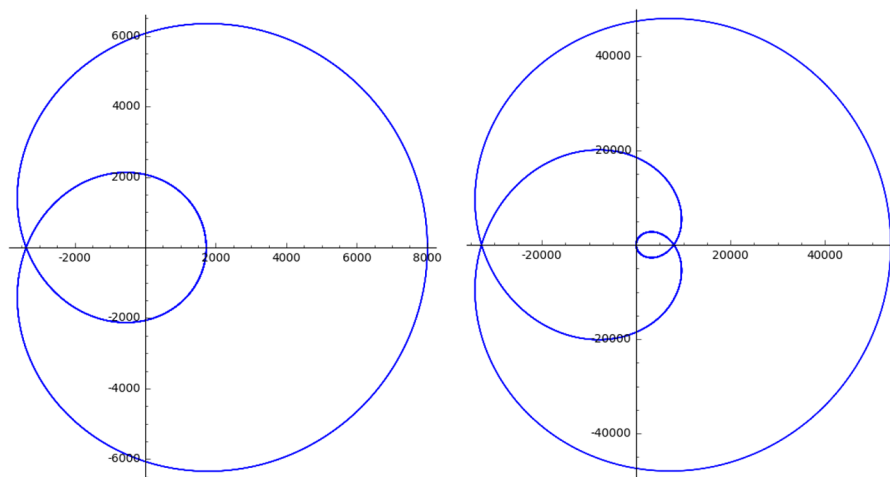
As the narrow class group of  $\mathbf{Z}[\sqrt{2}]$  is trivial, the image of any geodesic attached to a matrix with integral coefficients, trace zero, and determinant  $-2$  is the curve  $\mathcal{Z}_2$ .

The narrow class group of  $\mathbf{Z}[\sqrt{3}]$  has two elements. Two matrices with trace zero and determinant  $-3$  which are not  $SL_2(\mathbf{Z})$ -conjugate are  $A_1^3 = \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}$  and  $A_2^3 = \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}$ . However,  $A_1^3$  is  $SL_2(\mathbf{Z})$ -conjugate to  $-A_2^3$ , which gives rise to the same geodesic as  $A_2^3$ . Hence,

$$J(\mathcal{C}_{A_1^3}) = J(\mathcal{C}_{A_2^3}) = \mathcal{Z}_3(\mathbf{R}).$$

Notice, however, that there are two  $SL_2(\mathbf{Z})$ -orbits of oriented geodesics  $\mathcal{C}_A$  with  $A$  of trace zero and determinant  $-3$ .

The curve  $\mathcal{Z}_2$  (resp.  $\mathcal{Z}_3$ ) is represented below on the left (resp. right).



Generalizing the previous example, we obtain the following result.

**Corollary 3.5.6** *Let  $N > 1$  be a square-free integer congruent to 2 or 3 modulo 4. Assume that the class group of  $\mathbf{Z}[\sqrt{N}]$  is trivial. Then,  $J(\mathcal{C}_A) = \mathcal{Z}_N(\mathbf{R})$  for any  $A \in M_2(\mathbf{Z})$  with trace zero and determinant  $-N$ .*

**Proof** If  $\mathbf{Z}[\sqrt{N}]^\times$  contains an element of norm  $-1$ , then the narrow class group of  $\mathbf{Z}[\sqrt{N}]$  is trivial; hence, the statement follows from Proposition 3.5.4. Now, assume that the equation  $X^2 - NY^2 = -1$  has no integral solution; in this case,  $Cl^+(\mathbf{Z}[\sqrt{N}])$  has two elements. Consider the two matrices  $A_1^N = \begin{pmatrix} 0 & N \\ 1 & 0 \end{pmatrix}$  and  $A_2^N = \begin{pmatrix} 0 & 1 \\ N & 0 \end{pmatrix}$ .

Given  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$ , we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ N & 0 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} Nbd - ac & a^2 - Nb^2 \\ Nd^2 - c^2 & -Nbd + ac \end{pmatrix}.$$

It follows that the matrices  $A_1^N$  and  $A_2^N$  are not  $SL_2(\mathbf{Z})$ -conjugate, so they correspond to the two elements of  $Cl^+(\mathbf{Z}[\sqrt{N}])$ ; on the other hand,

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ N & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -N \\ -1 & 0 \end{pmatrix}.$$

As  $\mathcal{C}_{A_1^N} = \mathcal{C}_{-A_1^N}$ , the result follows. ■

**Remark 3.5.7** For a general  $N > 1$  square-free and congruent to 2, 3 modulo 4, the set  $\widehat{Cl}^+(\mathbf{Z}[\sqrt{N}])$  may have different cardinality from the (narrow) class group of

$\mathbf{Z}[\sqrt{N}]$ . Indeed, the matrices  $A \in M_2(\mathbf{Z})$  with trace zero and determinant  $-N$  whose image in  $Cl^+(\mathbf{Z}[\sqrt{N}])$  is fixed by the involution sending  $A$  to  $-A$  give rise to *reciprocal geodesics* studied in [Sar07] (see, in particular, pages 223 and 224 of *loc. cit.* for a description of the elements of the narrow class group corresponding to such matrices). For example, take  $N = 82$ . Then,  $Cl^+(\mathbf{Z}[\sqrt{82}]) = Cl(\mathbf{Z}[\sqrt{82}])$  is cyclic of order 4, and only two elements correspond to reciprocal geodesics. Explicitly, the matrix  $\begin{pmatrix} -1 & 27 \\ 3 & 1 \end{pmatrix}$  is not  $SL_2(\mathbf{Z})$ -conjugate to its opposite.

### 3.5.8 The curve $\mathcal{Z}_5$

We have seen above that the curve  $\mathcal{Z}_1$  is the union of two images of geodesics, each of which is not algebraic. Let us give further examples of this phenomenon.

There are two elements in  $\widetilde{Cl}^+(\mathbf{Z}[\sqrt{5}])$  (because there is one invertible and one noninvertible ideal class for the ring  $\mathbf{Z}[\sqrt{5}]$ ) corresponding to the matrices  $A_1^5 = \begin{pmatrix} 0 & 5 \\ 1 & 0 \end{pmatrix}$  and  $A_2^5 = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$ . The matrix  $A_1^5$  is not  $SL_2(\mathbf{Z})$ -conjugate to  $\pm A_2^5$ ; hence, we learn from Proposition 3.5.4 that  $\mathcal{Z}_5(\mathbf{R}) = J(\mathcal{C}_{A_1^5}) \cup J(\mathcal{C}_{A_2^5})$ ; furthermore, each  $J(\mathcal{C}_{A_i^5})$  is a compact curve (in the Euclidean topology) properly contained in  $\mathcal{Z}_5(\mathbf{R})$ . We have

$$\mathcal{C}_{A_1^5} : X^2 + Y^2 = 5, \quad \mathcal{C}_{A_2^5} : X^2 + Y^2 - X - 1 = 0.$$

In particular,  $(2, 1) \in \mathcal{C}_{A_1^5}$  and  $(0, 1) \in \mathcal{C}_{A_2^5}$ , so the curves  $J(\mathcal{C}_{A_1^5})$  and  $J(\mathcal{C}_{A_2^5})$  intersect at  $j(2 + i) = j(i)$ . Finally, it follows from Theorem 3.5.10 below that each  $J(\mathcal{C}_{A_i^5})$  is not algebraic.

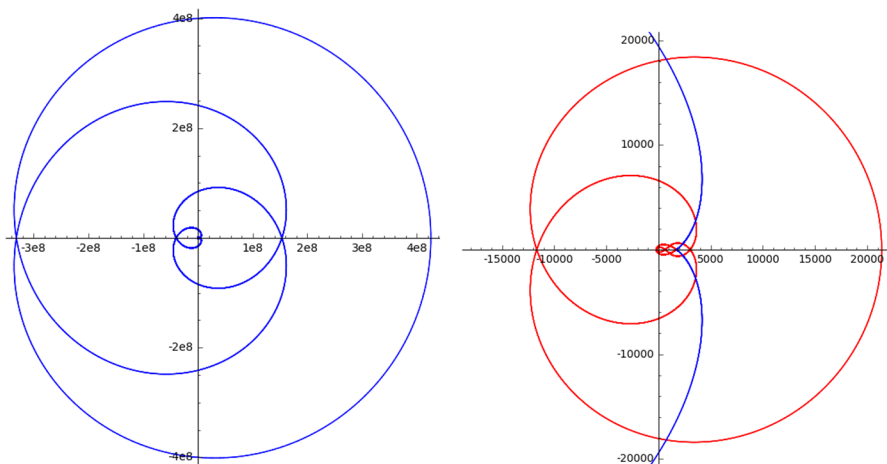
### 3.5.9 The curve $\mathcal{Z}_{10}$

Let us finally describe the curve  $\mathcal{Z}_{10}$ . The narrow class group of  $\mathbf{Z}[\sqrt{10}]$  coincides with its class group, and it has two elements. The two matrices

$$A_1^{10} = \begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix}, \quad A_2^{10} = \begin{pmatrix} 0 & 5 \\ 2 & 0 \end{pmatrix}$$

are not  $SL_2(\mathbf{Z})$ -conjugate; better,  $A_1^{10}$  is not conjugate to  $\pm A_2^{10}$ . The curve  $\mathcal{C}_{A_1^{10}}$  (resp.  $\mathcal{C}_{A_2^{10}}$ ) is the upper half-circle with equation  $X^2 + Y^2 = 10$  (resp.  $X^2 + Y^2 = \frac{5}{2}$ ). Proposition 3.5.4 tells us that  $J(\mathcal{C}_{A_1^{10}}) \cup J(\mathcal{C}_{A_2^{10}}) = \mathcal{Z}_{10}(\mathbf{R})$ , and the inclusions  $J(\mathcal{C}_{A_i^{10}}) \subset \mathcal{Z}_{10}(\mathbf{R})$  are strict. Notice that  $(3, 1) \in \mathcal{C}_{A_1^{10}}$  and  $(\frac{3}{2}, \frac{1}{2}) \in \mathcal{C}_{A_2^{10}}$ . As  $\frac{3}{2} + \frac{i}{2} = \frac{2i+1}{i+1}$ , we have  $j(3 + i) = j(i) = j(\frac{3}{2} + \frac{i}{2})$ ; hence, the curves  $J(\mathcal{C}_{A_i^{10}})$  intersect at 1, 728. The situation is represented in the picture below, where the curve  $J(\mathcal{C}_{A_1^{10}})$  (resp.  $J(\mathcal{C}_{A_2^{10}})$ ) is depicted in blue (resp. red).





As before, the following André–Oort-type statement in our situation implies that each  $J(\mathcal{C}_{A_i^{10}})$  is not algebraic.

**Theorem 3.5.10** *Let  $\mathcal{C} \subset \mathbf{A}_{\mathbf{R}}^2$  be an irreducible curve. Assume that  $\mathcal{C}(\mathbf{R})$  contains infinitely many images of special points in  $\mathbf{H}$ . Then,  $\mathcal{C}(\mathbf{R}) = \mathcal{Z}_N(\mathbf{R})$  for some  $N \geq 1$ .*

**Proof** We can write  $\mathcal{C}(\mathbf{R}) = \{(x, y) \in \mathbf{R}^2 \mid P(x, y) = 0\}$  for some geometrically irreducible polynomial  $P(X, Y) \in \mathbf{R}[X, Y]$ . Let  $\mathcal{C}_{\mathbf{C}} \subset \mathbf{A}_{\mathbf{C}}^2$  be the base change of  $\mathcal{C}$  to  $\mathbf{C}$ . Consider the map

$$f : \mathbf{C}^2 \rightarrow \mathbf{C}^2$$

$$(x, y) \mapsto (x + iy, x - iy).$$

Observe that  $(x, y) \in \mathbf{R}^2$  is the image of a special point in  $\mathbf{H}$  if and only if  $f(x, y) \in \mathbf{C}^2$  is the image of a couple of special points in  $\mathbf{H} \times \mathbf{H}$ . Therefore, the image  $\tilde{\mathcal{C}} = f(\mathcal{C}_{\mathbf{C}})$  is an irreducible algebraic curve in  $\mathbf{C}^2$  containing infinitely many special points. The curve  $\tilde{\mathcal{C}}$  cannot be vertical: indeed, in this case, there would be  $c \in \mathbf{C}$  such that  $x + iy = c$  for every  $(x, y) \in \mathcal{C}(\mathbf{R})$ , contradicting the fact that  $\mathcal{C}(\mathbf{R})$  is infinite. For the same reason,  $\tilde{\mathcal{C}}$  cannot be horizontal. Hence, by [And98], the curve  $\tilde{\mathcal{C}}$  is the vanishing locus of a modular polynomial  $\Phi_N(X, Y)$ . Therefore,

$$\mathcal{C}(\mathbf{R}) = \{(x, y) \in \mathbf{R}^2 \mid \Phi_N(f(x, y)) = 0\} = \mathcal{Z}_N(\mathbf{R}). \quad \blacksquare$$

**Acknowledgment** I started exploring the topics discussed in this document inspired by [mo220]. I am very grateful to Gregorio Baldi for bringing [mo220] to my attention, and for introducing me to functional transcendence in the real world. I am indebted to Luca Dall’Ava for several stimulating discussions and for the realization of many pictures, some of which are included in this document. I thank Leonardo Lerer for many helpful discussions and for answering several of my (naive) questions. I also wish to thank Louis Jaburi and Alice Pozzi for interesting discussions related to the contents of this note. Finally, I am grateful to the anonymous referee for useful comments and for pointing out to me the paper [Sar07].

## References

- [And98] Y. André, *Finitude des couples d'invariants modulaires singuliers sur une courbe algébrique plane non modulaire*. J. Reine Angew. Math. 505(1998), 203–208.
- [BT20] B. Bakker and J. Tsimerman, *Lectures on the Ax–Schanuel conjecture*. In: M.-H. Nicole (ed.), *Arithmetic geometry of logarithmic pairs and hyperbolicity of moduli spaces*. Hyperbolicity in Montréal. Based on three workshops, Montréal, Canada, 2018–2019, Springer, Cham, 2020, pp. 1–68.
- [Con] K. Conrad, *Ideal classes and matrix conjugation over  $\mathbb{Z}$* . <https://kconrad.math.uconn.edu/blurbs/gradnumthy/matrixconj.pdf>.
- [DV21] H. Darmon and J. Vonk, *Arithmetic intersections of modular geodesics*. J. Number Theory 230(2021), 89–111.
- [Hel78] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics, 80, Academic Press, New York–San Francisco–London, 1978, 628 pp.
- [Jaf74] H. A. Jaffee, *Real forms in Hermitian symmetric spaces and real algebraic varieties*. 1974. <https://www.math.stonybrook.edu/alumni/1974-Harris-Jaffee.pdf>.
- [KUY18] B. Klingler, E. Ullmo, and A. Yafaev, *Bi-algebraic geometry and the André–Oort conjecture*. In: T. de Fernex, B. Hassett, M. Mustață, M. Olsson, M. Popa, R. Thomas (ed.), *Algebraic geometry: Salt Lake City 2015*. Proceedings of the 2015 Summer Research Institute in Algebraic Geometry, University of Utah, Salt Lake City, UT, USA, July 13–31, 2015. Part 2, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2018, pp. 319–359.
- [mo220] MathOverflow, *Equations defining hyperbolic geodesics in  $\mathbb{C} \setminus \{0, 1\}$*  2020. <https://mathoverflow.net/questions/375352/equations-defining-hyperbolic-geodesics-in-mathbb-c-setminus-0-1> (MathOverflow post).
- [Pil11] J. Pila, *O-minimality and the André–Oort conjecture for  $\mathbb{C}^n$* . Ann. of Math. 173(2011), no. 3, 1779–1840.
- [Ric21] J. Rickards, *Computing intersections of closed geodesics on the modular curve*. J. Number Theory 225(2021), 374–408.
- [Sar07] P. Sarnak, *Reciprocal geodesics*. In: W. Duke, Y. Tschinkel (ed.), *Analytic number theory. A tribute to Gauss and Dirichlet*. Proceedings of the Gauss–Dirichlet conference, Göttingen, Germany, June 20–24, 2005, American Mathematical Society, Providence, RI, 2007, pp. 217–237.
- [Sch36] T. Schneider, *Arithmetische Untersuchungen elliptischer Integrale*. Math. Ann. 113(1936), 1–13.

*Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2AZ,  
United Kingdom*

*e-mail:* [m.tamiozzo@imperial.ac.uk](mailto:m.tamiozzo@imperial.ac.uk)