# Lectures on Elliptic Curves

Thomas Krämer

Winter 2019/20, HU Berlin

preliminary version: 1/2/2020

# Contents

# Introduction

Elliptic curves belong to the most fundamental objects in mathematics and connect many different research areas such as number theory, algebraic geometry and complex analysis. Their definition and basic properties can be stated in an elementary way: Roughly speaking, an elliptic curve is the set of solutions to a cubic equation in two variables over a field. Thus elliptic curves are very concrete and provide a good starting point to enter algebraic geometry. At the same time their arithmetic properties are closely related to the theory of modular forms and have seen spectacular applications in number theory like Andrew Wiles' proof of Fermat's last theorem. They are the object of long-standing open conjectures such as the one by Birch and Swinnerton-Dyer. Even in applied mathematics, elliptic curves over finite fields are nowadays used in cryptography.

The following notes accompany my lectures in the winter term 2019/20. The lectures will give a gentle introduction to the theory of elliptic curves with only mininum prerequisites. We start with elliptic curves over $\mathbb{C}$, which are quotients of the complex plane by a lattice arising from arclength integrals for an ellipse. As such they are objects of complex analysis: Compact Riemann surfaces. What makes their theory so rich is that at the same time they have an algebraic description as plane curves cut out by cubic polynomials. Passing to algebraic geometry, we can consider elliptic curves over arbitrary fields. These are the simplest examples of abelian varieties: Projective varieties with an algebraic group structure. Finally, we will give a glimpse of the arithmetic of elliptic curves, looking in particular at the group of points on elliptic curves over number fields. These notes will be updated on an irregular basis and are incomplete even on the few topics that we can cover in the lecture. For further reading there are many excellent textbooks such as the following:

- Cassels, J.W.S., *Lectures on Elliptic Curves*,
  LMS Student Series, Cambridge University Press (1992).

- Husemöller, D., *Elliptic Curves*,
  Graduate Texts in Math., Springer (1987).

- Silverman, J.H., *The Arithmetic Theory of Elliptic Curves*,
  Graduate Texts in Math., Springer (1986).

- —, *Advanced Topics in the Arithmetic Theory of Elliptic Curves*,
  Graduate Texts in Math., Springer (1994).

CHAPTER I

# Analytic theory of elliptic curves

## 1. Motivation: Elliptic integrals

The notion of elliptic curves emerged historically from the discussion of certain integrals that appear for instance in computing the arclength of an ellipse. These integrals are best understood in the complex setting. Recall that for any open subset $U \subseteq \mathbb{C}$, the *path integral* of a continuous function $f : U \to \mathbb{C}$ along a piecewise smooth path $\gamma : [0,1] \to U$ is defined by

$$\int_\gamma f(z)dz \;=\; \int_0^1 f(\gamma(t))\,\dot{\gamma}(t)\,dt,$$

where $\dot{\gamma}(t) = \frac{d}{dt}\mathrm{Re}(\gamma(t)) + i\frac{d}{dt}\mathrm{Im}(\gamma(t))$. The most basic example is

Example 1.1. Take $U = \mathbb{C}^*$ and put $\gamma(t) = \exp(2\pi i t)$, then $\dot{\gamma}(t) = 2\pi i \gamma(t)$ and hence

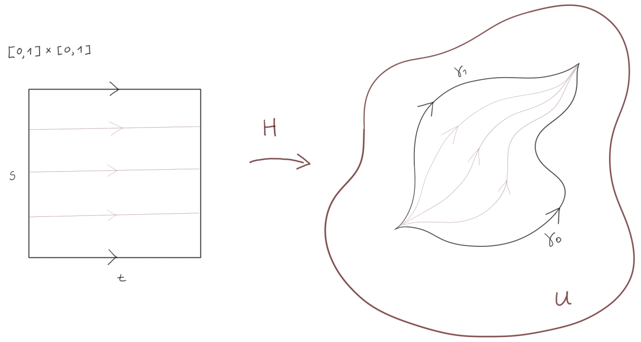
$$\int_\gamma z^n dz \;=\; \int_0^1 2\pi i \cdot e^{2\pi i(n+1)t} dt \;=\; \begin{cases} 0 & \text{if } n \neq -1, \\ 2\pi i & \text{if } n = -1. \end{cases}$$

Comparing with the corresponding integral over a constant path, one sees that in general the value of the integral depends on the chosen path and not just on its endpoints. However, the path integral of holomorphic functions is unchanged under continuous deformations of the path in the following sense:

Definition 1.2. A *homotopy* between two continuous paths $\gamma_0, \gamma_1 : [0,1] \to U$ is a continuous map

$$H :\; [0,1] \times [0,1] \;\to\; U \quad \text{with} \quad H(s,t) = \begin{cases} \gamma_0(t) & \text{if } s = 0 \text{ or } t = 0, \\ \gamma_1(t) & \text{if } s = 1 \text{ or } t = 1. \end{cases}$$

If there exists such a homotopy, we write $\gamma_0 \sim \gamma_1$ and say that the two paths $\gamma_0, \gamma_1$ are *homotopic*:



The deformation invariance of line integrals over holomorphic functions can now be made precise as follows:

Theorem 1.3 (Cauchy). *If two smooth paths $\gamma_0, \gamma_1 : [0,1] \to U$ are homotopic, then*

$$\int_{\gamma_0} f(z)dz \;=\; \int_{\gamma_1} f(z)dz \quad \text{for all holomorphic} \quad f : U \to \mathbb{C}.$$

Let us recall a few more notations from topology. The notion of homotopy $\sim$ is an equivalence relation on continuous paths in $U$ with given starting and end point, and we denote by

$$\pi_1(U,p,q) \;=\; \{\gamma : [0,1] \to U \mid \gamma(0) = p \text{ and } \gamma(1) = q\}/\sim$$

the set of *homotopy classes* of paths from $p$ to $q$. The composition of paths defines a product

$$\pi_1(U,p,q) \times \pi_1(q,r) \to \pi_1(U,r), \quad (\gamma_1,\gamma_2) \mapsto \gamma_1 \cdot \gamma_2$$

where

$$(\gamma_1 \cdot \gamma_2)(t) \;=\; \begin{cases} \gamma_1(2t) & \text{for } t \in [0,1/2], \\ \gamma_2(2t-1) & \text{for } t \in [1/2,1], \end{cases}$$

and this product is associative. Similarly, reversing the direction of paths gives an inversion map

$$\pi_1(U,p,q) \to \pi_1(U,q,p), \quad \gamma \mapsto \gamma^{-1} = (t \mapsto \gamma(1-t)).$$

For $p = q$ this makes the set of homotopy classes of closed loops at $p \in U$ a group, the *fundamental group*

$$\pi_1(U,p) = \pi_1(U,p,p).$$

We say $U$ is *simply connected* if this fundamental group is trivial. In this case any two continuous paths with the same starting point and the same end point are homotopic, hence the value of the path integral of a holomorphic function $f : U \to \mathbb{C}$ over $\gamma : [0,1] \to U$ only depends on $p = \gamma(0)$ and $q = \gamma(1)$ but not on the path itself. We can then put

$$\int_p^q f(z)dz \;=\; \int_\gamma f(z)dz$$

for any $\gamma \in \pi_1(U,p,q)$. In the non-simply connected case we have:

Corollary 1.4. *For any holomorphic function $f : U \to \mathbb{C}$ and $p \in \mathbb{C}$, the path integral defines a group homomorphism*

$$\pi_1(U,p) \to (\mathbb{C},+), \quad \gamma \mapsto \int_\gamma f(z)dz.$$

*Proof.* One can show that any continuous path is homotopic to a smooth one, so the result follows from Cauchy's theorem and from the additivity of path integrals with respect to the composition of paths. □

The image of the above homomorphism is an additive subgroup $\Lambda_f \subset \mathbb{C}$, and for $p, q \in U$ the value

$$\left( \int_p^q f(z)dz \mod \Lambda_f \right) \;\in\; \mathbb{C}/\Lambda_f$$

is well-defined modulo this subgroup.

Example 1.5. On any simply connected open $U \subseteq \mathbb{C}^* = \mathbb{C} \setminus \{0\}$ with $1 \in U$ we define a branch of the logarithm by

$$\log z = \int_1^z \frac{1}{x}\, dx.$$

If $U$ is not taken to be simply connected, the complex logarithm will in general not be well-defined globally. But on all of $U = \mathbb{C}^*$ the logarithm is well-defined modulo the subgroup $\Lambda_f = 2\pi i\mathbb{Z} \subset \mathbb{C}$ as indicated in the following diagram where $\exp : \mathbb{C} \to \mathbb{C}^*$ denotes the universal cover and $q : \mathbb{C} \to \mathbb{C}/\Lambda_f$ is the quotient map:

$$\begin{tikzcd}
\mathbb{C} \arrow[d, "\exp"'] \arrow[dr, "q"] & \\
\mathbb{C}^* \arrow[r, dotted, "\exists \log"'] & \mathbb{C}/\Lambda_f
\end{tikzcd}$$

As an exercise you may check that the multiplicativity of the exponential function translates to the fact that for all $\gamma_1, \gamma_2, \gamma_3 : [0,1] \to \mathbb{C}^*$ with $\gamma_1(t)\gamma_2(t)\gamma_3(t) = 1$ for all $t$, one has

$$\int_{\gamma_1} \frac{dz}{z} + \int_{\gamma_2} \frac{dz}{z} + \int_{\gamma_3} \frac{dz}{z} \;=\; 0$$

This is a blueprint for what we will see below for elliptic integrals.

Note that once we have the complex logarithm, we can find a closed expression for the integral over any rational function: Any $f(x) \in \mathbb{C}(x)$ has a decomposition into partial fractions

$$f(z) \;=\; \sum_{i=1}^{n} c_i \cdot (z - a_i)^{n_i} \quad \text{with} \quad a_i, c_i \in \mathbb{C}, \quad n_i \in \mathbb{Z}.$$

and for $z_0, z_1 \in \mathbb{C} \setminus \{a_i\}$ we have

$$\int_{z_0}^{z_1} (z - a_i)^{n_i}\, dz \;=\; F_i(z_1) - F_i(z_0), \quad F_i(z) \;=\; \begin{cases} \frac{(z-a_i)^{n_i+1}}{n_i+1} & \text{if } n_i \neq -1, \\ \log(z - a_i) & \text{if } n_i = -1, \end{cases}$$

where log denotes any branch of the complex logarithm on a large enough simply connected open $U \subseteq \mathbb{C}$. The above expresses the integral as a function of $z_0, z_1$ in terms of *elementary functions*, i.e. functions obtained by combining complex polynomials, exponentials and logarithms. For more complicated integrands such an expression usually does not exist, but this is no bad news: It means that there are many more interesting functions out there than the elementary ones!

Exercise 1.6. Let $a, b$ be positive real numbers with $a \leq b$. Show that for the ellipse

$$\mathbb{E} = \{(a\cos(\varphi), b\sin(\varphi)) \in \mathbb{R}^2 \mid \varphi \in \mathbb{R}\}$$

and any $\varphi_0, \varphi_1 \in [0, \pi]$, the arclength $\ell$ of the segment $\varphi_0 \leq \varphi \leq \varphi_1$ has the form

$$\ell \;=\; \frac{1}{2} \int_{x_0}^{x_1} \frac{1 - cx}{\sqrt{x(1-x)(1-cx)}} dx \quad \text{with} \quad x_i = x_i(\varphi_i) \in \mathbb{R} \quad \text{and} \quad c = 1 - \tfrac{a^2}{b^2}.$$

Such integrals usually cannot be expressed via elementary functions and have a special name:

Definition 1.7. An *elliptic integral* is a function which can be expressed in the form

$$F(v) = \int_u^v R\left(x, \sqrt{f(x)}\right)\, dx \quad \text{for some constant } u,$$

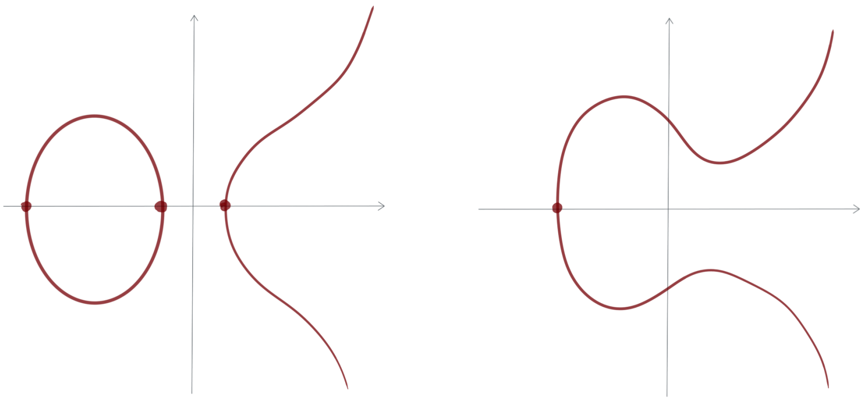
where

- $R$ is a rational function in two variables, and
- $f$ is a polynomial of degree 3 or 4 with no repeated roots.

If the above definition is read over the complex numbers, the integral will again depend on the chosen path of integration, which we always assume to avoid any poles of the integrand. In contrast to the previous examples the integrand is now a "multivalued" function as there is no distinguished sign choice for the complex square root. Hence rather than integrating along a path in the complex plane, we should integrate along a path in the zero set

$$E_0 \;=\; \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\}$$

where $y$ keeps track of the chosen square root. In the language of algebraic geometry this is the affine part of an *elliptic curve*. The set $E_0 \cap \mathbb{R}^2$ of its real points will look like this:



However, the topological and analytic properties of the set of complex points of $E_0$ will become more tangible in the framework of Riemann surfaces.

## 2. The topology of elliptic curves

Let us for simplicity assume $f(x) = x(x-1)(x-\lambda)$ for some $\lambda \in \mathbb{C} \setminus \{0,1\}$; we will see later that up to projective coordinate transformations this is no restriction of generality. On any *simply connected* open subset of $\mathbb{C} \setminus \{0,1,\lambda\}$ we can pick a branch of the logarithm and define

$$\sqrt{f(x)} \;=\; \exp\left(\frac{\log f(x)}{2}\right).$$

However, if we try to analytically continue this function along a small closed loop around any of the punctures $0, 1, \lambda$, the logarithm will change by $2\pi i$ and hence the square root will be replaced by its negative:

$$\exp\left(\frac{2\pi i + \log f(x)}{2}\right) \;=\; -\exp\left(\frac{\log f(x)}{2}\right).$$

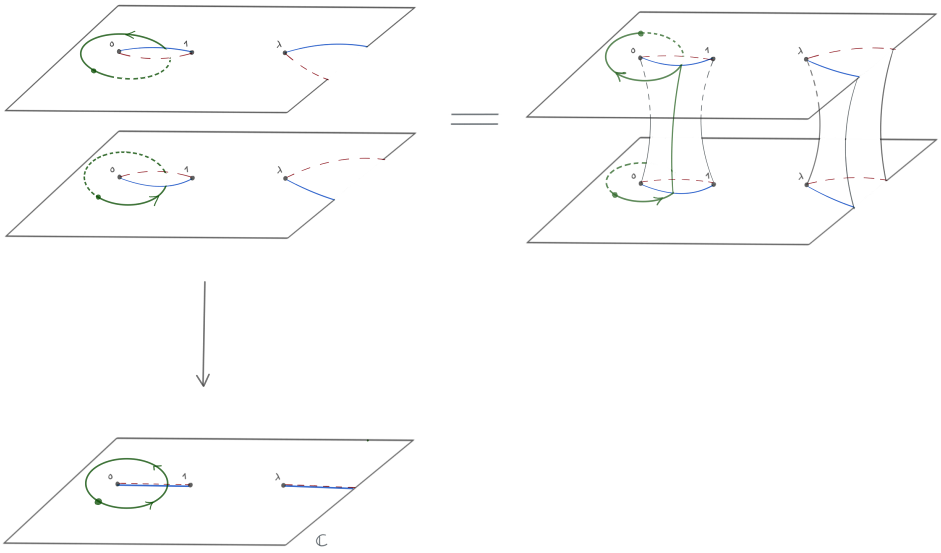
If we perform a loop around two punctures then the two signs will cancel. So if we fix a real half-line $[\lambda, \infty) \subset \mathbb{C} \setminus \{0, 1\}$ emanating from $\lambda$ in any direction, then for

$$S \;=\; [0,1] \cup [\lambda, \infty)$$

there is a holomorphic function

$$\rho : \quad U \;=\; \mathbb{C} \setminus S \;\to\; \mathbb{C} \quad \text{with} \quad \rho(x)^2 = f(x).$$

What about integrals along paths that cross the slits? We have seen above that whenever we analytically continue across one of the slits the square root is replaced by its negative. To treat both square roots equally, consider the disjoint union of two copies of the slit plane with the holomorphic function $\sqrt{f} : U \sqcup U \to \mathbb{C}$ which is $+\rho(x)$ on the first and $-\rho(x)$ on the second copy. Let us glue the two copies along their respective boundaries by inserting two copies of $S$ as shown in the following picture:



It follows from the above discussion that the resulting topological space $X_0$ carries a continuous function which on the open subset $U \sqcup U \subset X_0$ restricts to the above function $\sqrt{f}$. As a topological space $X_0$ is easier to visualize if we turn the second copy of the slit plane upside down before gluing the two copies, as shown on the right half of the above picture.

But $X_0$ is not just a topological space, we want to do complex analysis on it and compute path integrals. In order to do so, note that the projection $U \sqcup U \to U$ extends to a continuous map $p : X_0 \to \mathbb{C}$. This is a *branched double cover* with branch locus $\{0, 1, \lambda\}$ in the following sense:

Definition 2.1. By a *branched cover* of an open subset $S \subseteq \mathbb{C}$ we mean a continuous map $p : X \to S$ of topological manifolds such that every point $s \in S$ has a small open neighborhood $s \in U_s \subseteq S$ with the following property: There exists a biholomorphic map

$$\varphi : \quad U_s \;\xrightarrow{\sim}\; D \;=\; \{z \in \mathbb{C} \mid |z| < 1\} \quad \text{with} \quad \varphi(s) \;=\; 0$$

that lifts to a homeomorphism

$$\tilde{\varphi}: \quad p^{-1}(U_s) \xrightarrow{\sim} \bigsqcup_{x \in I_s} D_x$$

onto a disjoint union of copies of the unit disk $D_x = \{z \in \mathbb{C} \mid |z| < 1\}$ indexed by the set $I_s = p^{-1}(s)$ such that

$$\begin{tikzcd}
X \arrow[d, "p"'] & p^{-1}(U_s) \arrow[r, "\sim", "\tilde{\varphi}"'] \arrow[d, "p"'] & \bigsqcup_{x \in I_s} D_x \arrow[d, "\sqcup_x p_x"] \\
S & U_s \arrow[r, "\sim", "\varphi"'] & D
\end{tikzcd}$$

commutes, where the labelling is chosen such that $\tilde{\varphi}(x) = 0 \in D_x$ for each $x \in I_s$ and we assume

$$p_x: \quad D_x \longrightarrow D, \quad z \mapsto z^{e_x}$$

for some natural number $e_x \in \mathbb{N}$. We call $e_x$ the *ramification index* of $p$ at $x$. Note that these ramification indices depend only on the map $p$ but not on the specific choice of $\varphi$ or its lift $\tilde{\varphi}$. It is also clear from the above definition that any branched cover restricts to a covering map in the sense of topology on the complement of the branch locus

$$\mathrm{Br}(p) \ = \ \{s \in S \mid \exists\, x \in p^{-1}(s) \text{ with } e_x > 1\} \ \subset \ S$$

and that the latter is a discrete closed subset of $S$.

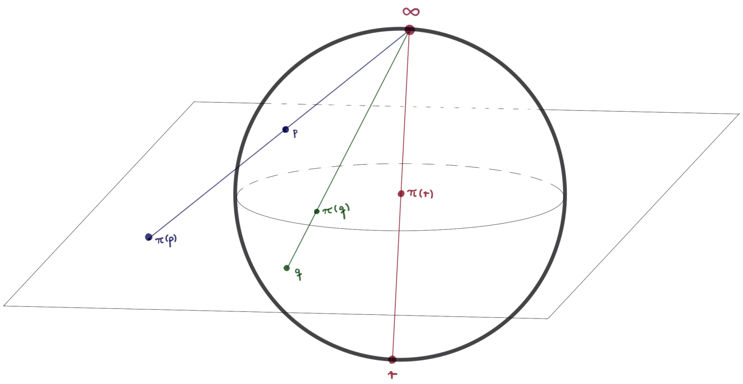
Let us now come back to the branched cover $p: X_0 \to \mathbb{C}$ obtained by glueing two copies of the slit complex plane as explained above. Comparing with the projection map from the affine elliptic curve $E_0 = \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\}$ we have:

Corollary 2.2. *There is a homeomorphism* $E_0 \xrightarrow{\sim} X_0$ *commuting with the projection to the complex plane:*

$$\begin{tikzcd}
E_0 \arrow[r, "\sim"] \arrow[dr, "(x,y)\mapsto x"'] & X_0 \arrow[d, "p"] \\
& \mathbb{C}
\end{tikzcd}$$

*Proof.* It follows from the holomorphic version of the implicit function theorem that the map $p: E_0 \to \mathbb{C}, (x,y) \mapsto x$ is also a branched cover, and as such it is determined uniquely by its restriction to the complement of any finite number of points of the target. But over $W = \mathbb{C} \setminus \{0, 1, \lambda\}$ the topological covers $E_0 \to W$ and $X_0 \to W$ are isomorphic because their monodromy coincides. □

It is often preferable to work with compact spaces. For instance, the complex plane can be compactified to a sphere by adding one point, as one may see by stereographic projection:



We denote this compactification by

$$\mathbb{P}^1 \;=\; \mathbb{C} \cup \{\infty\}$$

and call it the *Riemann sphere*. Note that the complement $\mathbb{P}^1 \setminus \{0\} = \mathbb{C}^* \cup \{\infty\}$ is also a copy of the complex plane. The Riemann sphere is obtained by glueing the two copies — which are also referred to as affine charts — along their overlap via the glueing map $\varphi : \mathbb{C}^* \to \mathbb{C}^*, z \mapsto 1/z$. By a *branched cover* of the Riemann sphere we mean a continuous map

$$p: \quad X \;\longrightarrow\; S \;=\; \mathbb{P}^1(\mathbb{C})$$

of topological manifolds which restricts over each of the two affine charts to a branched cover in the sense of definition 2.1. We will generalize this notion in the context of Riemann surfaces soon, but let us first finish our topological discussion of elliptic curves:

Lemma 2.3. *The branched double cover $p : X_0 \to \mathbb{C}$ from above extends uniquely to a branched cover*

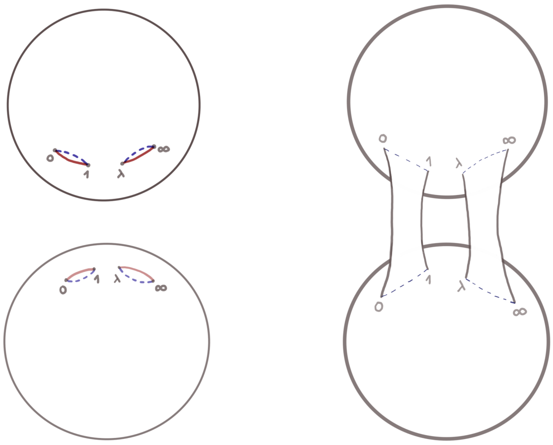
$$X = X_0 \cup \{pt\} \to \mathbb{P}^1$$

*with branch locus $\{0, 1, \lambda, \infty\}$, and we have a homeomorphism $X \simeq S^1 \times S^1$.*

*Proof.* Let $D \subset \mathbb{P}^1$ be a small disk around $\infty$. Then $D^* = D \setminus \{\infty\}$ is a pointed disk, and by the classification of branched covers of the pointed disk there exists a branched cover $X_\infty \to D$ extending the cover $p^{-1}(D^*) \to D^*$. We then get a branched cover

$$X \;=\; X_0 \cup_{p^{-1}(D^*)} X_\infty \;\to\; \mathbb{P}^1$$

by glueing. In order to show that this cover is branched at infinity, we only need to note that $p^{-1}(D^*) \to D^*$ has nontrivial monodromy. Finally, it follows from the construction of $X_0$ by glueing two copies of a slit complex plane that the compactification $X$ is obtained by glueing two copies of a slit Riemann sphere as indicated in the following picture:

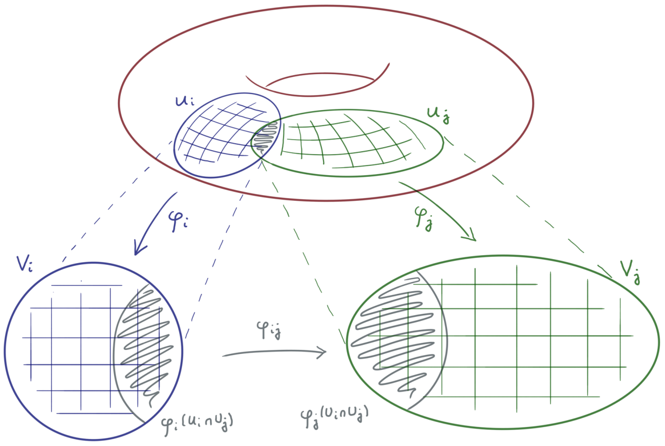


This easily implies that as a topological space $X \simeq S^1 \times S^1$ is a torus. □

## 3. Elliptic curves as complex tori

So far we have only been talking about topology, but all of the above spaces inherit from the complex plane a natural structure of Riemann surface:

Definition 3.1. A *Riemann surface* is a one-dimensional connected complex manifold, i.e. a connected Hausdorff topological space $S = \cup_{i \in I} U_i$ with an atlas of homeomorphisms $\varphi_i : U_i \xrightarrow{\sim} V_i \subseteq \mathbb{C}$ whose transition functions $\varphi_{ij} = \varphi_j \circ \varphi_i^{-1}$ are biholomorphic on the overlap of any two charts:



Example 3.2. (a) The Riemann sphere $\mathbb{P}^1(\mathbb{C})$ is a Riemann surface with two charts: As we have seen above, it is obtained by glueing to copies of the complex plane along the open subset $\mathbb{C}^* \subset \mathbb{C}$ via the gluing function $z \mapsto 1/z$.

(b) Any quotient $S = \mathbb{C}/\Lambda$ by a discrete subgroup $\Lambda \subset \mathbb{C}$ is a Riemann surface in a natural way. Notice that the discreteness is required because otherwise the quotient would not be Hausdorff. There are three possibilities: If $\Lambda = \{0\}$ we simply have $S = \mathbb{C}$. If $\Lambda = \mathbb{Z}\lambda$ for some $\lambda \in \mathbb{C} \setminus \{0\}$, the exponential map gives an isomorphism

$$\begin{aligned} S \;=\; \mathbb{C}/\mathbb{Z}\lambda &\xrightarrow{\sim} \mathbb{C}^* \\ z &\mapsto \exp(2\pi i z/\lambda). \end{aligned}$$

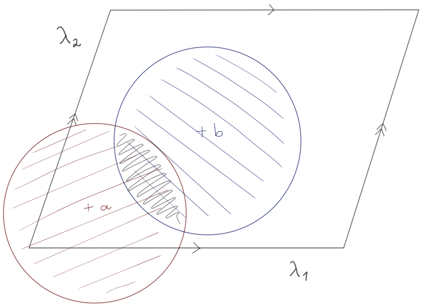
The only remaining case is that $\Lambda \subset \mathbb{C}$ is a *lattice*, by which we mean an additive subgroup $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$ generated by two $\mathbb{R}$-linearly independent $\lambda_1, \lambda_2 \in \mathbb{C}$. In this case the topological space $S = \mathbb{C}/\Lambda$ is homeomorphic to a torus, obtained by identifying the opposite sides of a fundamental parallelogram as shown below. We can construct an atlas by taking any nonempty open subset $V \subseteq \mathbb{C}$ which is small enough so that $V \cap (V + \lambda) = \varnothing$ for all $\lambda \in \Lambda \setminus \{0\}$, and consider the coordinate charts

$$V_a \;=\; V + a \quad \text{for} \quad a \;\in\; \mathbb{C}.$$

The projection $p : \mathbb{C} \to S$ restricts to homeomorphisms $p_a : V_a \xrightarrow{\sim} U_a \subseteq S$ on these charts and the transition maps between any two of the charts are given by translations

$$V_a \;\supseteq\; p_a^{-1}(U_a \cap U_b) \xrightarrow{id+\lambda_{ab}} p_b^{-1}(U_a \cap U_b) \;\subseteq\; V_b$$

where $\lambda_{ab}$ is constant:



Definition 3.3. If $S$ is a Riemann surface, then by a *holomorphic function* on an open $U \subseteq S$ we mean a function $f : U \to \mathbb{C}$ which restricts to a holomorphic function on each coordinate chart in the sense that for each such chart $\varphi_i : U_i \xrightarrow{\sim} V_i$ from $U_i \subseteq S$ to $V_i \subseteq \mathbb{C}$,

$$f \circ \varphi^{-1} : \quad \varphi^{-1}(U \cap U_i) \;\longrightarrow\; \mathbb{C}$$

is a holomorphic function. If $X$ is another Riemann surface, a map $p : X \to S$ is called a *morphism* of Riemann surfaces or a *holomorphic map* if for each coordinate chart $U_i \subset S$ the restriction $p^{-1}(U_i) \longrightarrow U_i$ is holomorphic.

Example 3.4. (a) Giving a meromorphic function on an open subset $X \subseteq \mathbb{C}$ is the same thing as giving a morphism $f : X \to \mathbb{P}^1(\mathbb{C})$ to the Riemann sphere, where we declare $f(x) = \infty$ iff $f$ has a pole at the point $x \in X$.

(b) For any lattice $\Lambda \subset \mathbb{C}$ the quotient map $\mathbb{C} \to \mathbb{C}/\Lambda$ is holomorphic. Indeed the universal cover of any Riemann surface has a unique structure of a Riemann surface making the covering map holomorphic. This extends to branched covers:

By a *branched cover* of a Riemann surface $S$ we mean a topological space $X$ together with a map $f : X \to S$ that restricts to a branched cover in the sense of the previous section over each chart of an atlas for the Riemann surface $S$:

$$\begin{tikzcd}
X \arrow[d, "f"'] & \supseteq & f^{-1}(U_i) \arrow[r, equal] \arrow[d] & X_i \arrow[d, "\substack{\text{branched}\\ \text{cover}}"] & & \\
S & \supseteq & U_i \arrow[r, "\sim", "\varphi_i"'] & V_i & \subseteq & \mathbb{C}
\end{tikzcd}$$

Exercise 3.5. Show that:

(1) If $p : X \to S$ is a branched cover as above, the topological space $X$ inherits a unique structure of a Riemann surface making $p$ holomorphic.

(2) If $\Sigma \subset S$ is a discrete subset, any topological covering map $p_0 : X_0 \to S\backslash\Sigma$ extends uniquely to a branched cover $p : X \to S$.

(3) Now let $S = \mathbb{P}^1(\mathbb{C})$ and $\Sigma = f^{-1}(0) \cup \{\infty\}$ for some $f \in \mathbb{C}[x] \setminus \{0\}$. Check that

$$p_0 : X_0 = \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x) \neq 0\} \to S \setminus \Sigma$$

is a double cover, and describe its extension $p : X \to S$ over each $s \in \Sigma$.

(4) If $f(x)$ has no multiple roots and $0 \in \Sigma$, show that there is a $g(u) \in \mathbb{C}[u]$ with

$$p^{-1}(S \setminus \{0\}) \;\simeq\; \{(u,v) \in \mathbb{C}^2 \mid v^2 = g(u)\}.$$

For $\deg(f) \in \{3,4\}$ the Riemann surfaces constructed above are the elliptic curves from the previous section. The main goal of this section is to shows that every elliptic curve over the complex numbers is isomorphic as a Riemann surface to a complex torus. The isomorphism will be obtained via certain path integrals. As in real analysis on smooth manifolds, the correct objects to integrate on a Riemann surface are not functions but differential forms:

Definition 3.6. A *holomorphic differential form* on an open subset $V \subseteq \mathbb{C}$ is a formal symbol $\omega = f(z)\,dz$ where $f : V \to \mathbb{C}$ is a holomorphic function and $z$ denotes the standard coordinate on the complex plane. If $\varphi : W \to V$ is a holomorphic map from another open subset of the complex plane, we define the pullback $\varphi^*(\omega) = f(\varphi(z))\,\frac{d}{dz}(\varphi(z))\,dz$. Note that the definition is made so that by substitution

$$\int_\gamma \varphi^*(\omega) \;=\; \int_{\varphi\circ\gamma} \omega \quad \text{for all paths} \quad \gamma : [0,1] \to W.$$

If $S$ is a Riemann surface with an atlas as above, then by a holomorphic differential form on $S$ we mean a family $\omega = (\omega_i)_{i\in I}$ of holomorphic differential forms $\omega_i$ on $V_i$ such that on the overlap of charts

$$\omega_i \;=\; \varphi_{ij}^*(\omega_j).$$

We then define the integral of such a differential form along a path $\gamma : [0,1] \to S$ by

$$\int_\gamma \omega \;=\; \sum_{\nu=1}^n \int_{\varphi_{i_\nu}\circ\gamma} \omega_i$$

for any decomposition $\gamma \sim \gamma_1 \cdots \gamma_n$ into paths $\gamma_\nu : [0,1] \to U_{i_\nu} \subseteq S$ in the charts; the compatibility condition on overlaps ensures that the outcome does not depend on the chosen decomposition. Cauchy's theorem easily implies

Corollary 3.7. *Let $S$ be a Riemann surface and $\omega$ a holomorphic differential form on it. If two smooth paths $\gamma_0, \gamma_1 : [0,1] \to S$ are homotopic, then their path integrals coincide:*

$$\int_{\gamma_0} \omega \;=\; \int_{\gamma_1} \omega.$$

*Hence for any $p \in S$ the path integral over the differential form $\omega$ gives a group homomorphism*

$$\pi_1(S,p) \to (\mathbb{C},+), \quad \gamma \mapsto \int_\gamma \omega.$$

*Proof.* Let $H : [0,1] \times [0,1] \to S$ be a homotopy with $\gamma_i = H_{\{i\}\times[0,1]}$ for $i = 0, 1$; the paths

$$\mu_i \;=\; H|_{[0,1]\times\{i\}}$$

for $i = 0, 1$ are constant, so any path integral over them vanishes and the claim is equivalent to

$$\int_\gamma \omega \;=\; 0 \quad \text{for the closed loop} \quad \gamma \;=\; \gamma_0 \cdot \mu_0 \cdot \gamma_1^{-1} \cdot \mu_1^{-1}.$$

Now $\gamma$ is contractible using the homotopy $H$, so if the image of $H$ is contained in a single coordinate chart, then we are done by Cauchy's theorem in the complex plane. In general, take a subdivision

$$[0,1] \times [0,1] \;=\; \bigcup_{i,j=1}^{N} Q_{ij} \quad \text{with} \quad Q_{ij} \;=\; [\tfrac{i-1}{N}, \tfrac{i}{N}] \times [\tfrac{j-1}{N}, \tfrac{j}{N}].$$

For $N \gg 0$ a compactness argument shows that each $H(Q_{ij}) \subset S$ will lie inside some coordinate chart $U_{ij} \subseteq S$. This reduces us to the case of a single coordinate chart, indeed the path integral is the sum

$$\int_\gamma \omega \;=\; \sum_{i,j=0}^{N} \int_{\gamma_{ij}} \omega \quad \text{for the oriented boundaries} \quad \gamma_{ij} = H|_{\partial Q_{ij}} : \quad [0,1] \to S$$

because the inner contributions from adjacent squares cancel. □

The image of the above homomorphism $\pi_1(S,s) \to \mathbb{C}$ is a subgroup $\Lambda_\omega \subset \mathbb{C}$, and for $p, q \in S$,

$$\left( \int_p^q \omega \mod \Lambda_\omega \right) \;\in\; \mathbb{C}/\Lambda_\omega$$

is well-defined modulo this subgroup. Let us now apply the above to the elliptic curve

$$X \;=\; \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\} \cup \{\infty\}$$

where $f(x) = x(x-1)(x-\lambda)$ with $\lambda \neq 0, 1$. In order to show that as a compact Riemann surface it is isomorphic to a complex torus, we will consider path integrals over the following holomorphic differential form:

Exercise 3.8. Consider the branched double cover $p : X \to \mathbb{P}^1(\mathbb{C})$. Show that the differential form

$$\omega \;=\; p^*\Big(\frac{dx}{\sqrt{f(x)}}\Big)$$

on $X \setminus p^{-1}(\{0, 1, \lambda, \infty\})$ extends to a holomorphic differential form on all of $X$.

By abuse of notation we also write $\omega = dx/\sqrt{f(x)}$ for simplicity. Thus we can consider

$$\int_\gamma \frac{dx}{\sqrt{f(x)}} \quad \text{for any path} \quad \gamma : [0,1] \to X.$$

To take a more systematic look at integrals of the above form, recall from the previous section that as a topological space $X \simeq S^1 \times S^1$ is homeomorphic to a torus. Its fundamental group

$$\pi_1(X,p) \;\simeq\; \pi_1(S^1,pt) \times \pi_1(S^1,pt) \;\simeq\; \mathbb{Z}\gamma_1 \times \mathbb{Z}\gamma_2$$

is therefore free abelian of rank two, generated by two loops $\gamma_1, \gamma_2 \in \pi_1(X,p)$. We fix these loops and denote by

$$\lambda_i \;=\; \int_{\gamma_i} \omega \;\in\; \mathbb{C}$$

their path integrals, which are also called the *fundamental periods* of the elliptic curve. By definition

$$\Lambda_\omega \;=\; \mathbb{Z}\lambda_1 + \mathbb{Z}\lambda_2 \;\subseteq\; \mathbb{C}$$

and the key step towards showing that elliptic curves are complex tori is that this is a lattice. For the proof we need to recall the notion of harmonic functions:

Exercise 3.9. A smooth function $g : U \to \mathbb{R}$ on an open subset $U \subseteq \mathbb{C}$ is called *harmonic* if

$$\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big)(g) \;=\; 0$$

where $\mathbb{R}^2 \xrightarrow{\sim} \mathbb{C}, (x,y) \mapsto z = x + iy$ denote the standard real coordinates.

(a) Show that a function is harmonic iff locally it can be written as the real part of a holomorphic function, and deduce that there is a well-defined notion of harmonic function on Riemann surfaces by looking at charts.

(b) Show that every harmonic function on a simply connected Riemann surface can be written globally as the real part of a unique holomorphic function. Can you find a counterexample in the not simply-connected case?

(c) Formulate and prove a mean value property for harmonic functions. Deduce that any harmonic function on a compact Riemann surface is constant.

We can now show that the subgroup $\Lambda_\omega \subset \mathbb{C}$ is indeed a lattice:

Theorem 3.10. *The fundamental periods $\lambda_1, \lambda_2$ are $\mathbb{R}$-linearly independent.*

*Proof.* Suppose that $\lambda_1, \lambda_2$ are $\mathbb{R}$-linearly dependent, wlog $\lambda_2 = a \cdot \lambda_1$ for some real number $a \in \mathbb{R}$. Then for any complex number $c \in \mathbb{C}^*$ with $\operatorname{Re}(c \cdot \lambda_1) = 0$ we also have $\operatorname{Re}(c \cdot \lambda_2) = 0$. But then $\operatorname{Re}(c \cdot \int_\gamma \omega) = 0$ for any closed loop $\gamma \in \pi_1(X, x_0)$, so the function

$$\tilde{g} : \quad \tilde{X} \longrightarrow \mathbb{R}, \quad x \mapsto \operatorname{Re}(c \cdot \textstyle\int_{x_0}^x \omega)$$

descends from the universal cover $p : \tilde{X} \to X$ to a well-defined function on $X$ as indicated below:

$$\begin{tikzcd}
\tilde{X} \arrow[rr, "\tilde{g}"] \arrow[dr, "p"'] & & \mathbb{R} \\
& X \arrow[ur, dotted, "\exists ! g"'] &
\end{tikzcd}$$

But $\tilde{g}$ is the real part of a holomorphic function, hence harmonic. Since $p : \tilde{X} \to X$ is a covering map, it follows that $g$ is harmonic as well. But we have seen above that any harmonic function on a compact Riemann surface is constant, so $g$ must be constant. It follows that $\tilde{g}$ is constant as well, which means that the holomorphic function

$$f: \quad \tilde{X} \longrightarrow \mathbb{C}, \quad x \mapsto \int_{x_0}^{x} \omega$$

has constant real part. Then by the Cauchy-Riemann equations $f$ must itself be constant, which is absurd because $\omega$ is not identically zero. □

Corollary 3.11. *The period map $\int \omega : X \to \mathbb{C}/\Lambda_\omega$ is an isomorphism of Riemann surfaces. In particular, for the universal cover we have a commutative diagram*

$$\begin{tikzcd}
\mathbb{C} \arrow[d, "\exists q"'] \arrow[dr, "p"] & \\
X \arrow[r, "\int \omega"'] & \mathbb{C}/\Lambda_\omega
\end{tikzcd}$$

*Proof.* The period map is easily seen to be holomorphic, and its derivative is the differential form

$$d(\textstyle\int \omega) \;=\; \omega$$

which vanishes nowhere. Using the implicit function theorem and the compactness of $X$ it follows that $\int \omega : X \to \mathbb{C}/\Lambda_\omega$ is a topological covering map (exercise), in other words

$$X \;\simeq\; \mathbb{C}/\Gamma \quad \text{for some subgroup} \quad \Gamma \;\subseteq\; \Lambda.$$

Passing to the universal cover we then get the claimed commutative diagram, except that we do not know yet that the period map is an isomorphism. But unravelling the definition of the map $q : \mathbb{C} \to X$, one sees that for any path $\gamma : [0,1] \to X$ starting at $x_0$ we have

$$\int_\gamma \omega \;=\; \tilde{\gamma}(1)$$

where $\tilde{\gamma} : [0,1] \to \mathbb{C}$ denotes the unique lift with $\tilde{\gamma}(0) = 0$ and $q \circ \tilde{\gamma} = \gamma$. If $\gamma$ runs through all elements of $\pi_1(X) = \Gamma$, then $\tilde{\gamma}(1)$ runs through $\Gamma$ while $\int_\gamma \omega$ runs through $\Lambda_\omega$ by definition of the period lattice. Hence $\Gamma = \Lambda_\omega$ and we are done. □

## 4. Complex tori as elliptic curves

In the last section we have seen that any elliptic curve over the complex numbers is isomorphic as a Riemann surface to a complex torus. We now want to show that every complex torus arises like this. For this we fix a lattice $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2 \subset \mathbb{C}$ where $\lambda_1, \lambda_2$ are any two complex numbers that are linearly independent over the reals, and consider the abstract Riemann surface $X = \mathbb{C}/\Lambda$. The idea is to find a branched double cover $p : X \to \mathbb{P}^1$ by looking at meromorphic functions on the complex plane that are periodic with respect to the lattice.

Before doing so, let us review some basic notions from complex analysis. For a meromorphic function $f$ on an open subset $U \subseteq \mathbb{C}$, its *order* at a point $a \in U$ is defined by

$$\operatorname{ord}_a(f) \;=\; \max\bigl\{n \in \mathbb{Z} \mid \; \exists \lim_{z \to a} (z-a)^{-n} f(z) \;\in\; \mathbb{C}\bigr\} \;\in\; \mathbb{Z} \cup \{+\infty\},$$

i.e.

$$\operatorname{ord}_a(f) \;=\; \begin{cases} \infty \text{ if } f \text{ is identically zero around } a, \\ \text{vanishing order of } f \text{ if } f \text{ has a zero at } a, \\ -\text{order of pole of } f \text{ if } f \text{ has a pole at } a. \end{cases}$$

The *residue* of $f$ at $a$ is defined as the coefficient $\operatorname{Res}_a(f) = c_{-1}$ in a Laurent expansion

$$f(z) \;=\; \sum_{n \gg -\infty} c_n (z-a)^n$$

on a small disc centered at $a$. By direct inspection it can also be computed as the path integral

$$\operatorname{Res}_a(f) \;=\; \frac{1}{2\pi i} \oint_{|z-a|=\epsilon} f(z)dz$$

over a small clockwise loop around $a$. In fact the RESIDUE THEOREM says that for $U \subseteq \mathbb{C}$ simply connected, any holomorphic function $f : U \setminus \{a_1, \dots, a_n\} \to \mathbb{C}$ satisfies

$$\frac{1}{2\pi i} \int_\gamma f(z)dz \;=\; \sum_{i=1}^n w_{a_i}(\gamma) \cdot \operatorname{Res}_{a_i}(f)$$

for all piecewise smooth closed loops $\gamma : [0,1] \to U \setminus \{a_1, \dots, a_n\}$. Here we denote by $w_{a_i}(\gamma) \in \mathbb{Z}$ the *winding number* of the given loop around the point $a_i$, which can be defined by

$$w_{a_i}(\gamma) \;=\; \frac{\varphi_i(1) - \varphi_i(0)}{2\pi} \;\in\; \mathbb{Z}$$

where $\varphi_i : [0,1] \to \mathbb{R}, t \mapsto \arg(\gamma(t) - a_i)$ is any continuous choice of the argument function. As special case of the residue theorem, the WINDING NUMBER FORMULA says that

$$w_a(\gamma) \;=\; \frac{1}{2\pi i} \int_\gamma \frac{dz}{z-a}$$

for any closed loop $\gamma : [0,1] \to U \setminus \{a\}$. We will apply the above results for the study of poles and zeroes of elliptic functions:

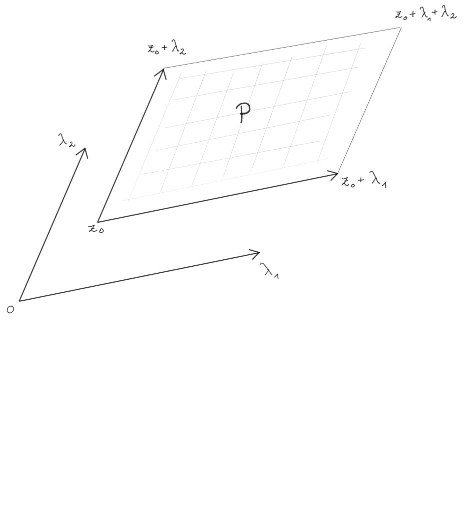
DEFINITION 4.1. An *elliptic function* with respect to the lattice $\Lambda \subset \mathbb{C}$ is a meromorphic function $f$ on the complex plane with $f(z+\lambda) \equiv f(z)$ for all $\lambda \in \Lambda$, or equivalently a morphism

$$f : \quad \mathbb{C}/\Lambda \;\longrightarrow\; \mathbb{P}^1(\mathbb{C}).$$

Note that any non-constant elliptic function must have poles, since any holomorphic function on a compact Riemann surface is constant. We will soon give a complete description of all elliptic functions for any given lattice. Let $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$ and denote by

$$P \;=\; \bigl\{z_0 + a_1\lambda_1 + a_2\lambda_2 \mid a_1, a_2 \in [0,1]\bigr\}$$

the fundamental parallelogram shifted by some fixed complex number $z_0 \in \mathbb{C}$ as in the following picture:



For a given elliptic function we can always choose $z_0$ such that the boundary $\partial P$ contains neither zeroes or poles of the function, since these form a discrete subset of the complex plane. The residue theorem then implies:

THEOREM 4.2. *Let $f$ be an elliptic function.*

(1) *If $f$ has no poles on the boundary $\partial P$ of the fundamental parallelogram, then*

$$\sum_{a\in P\backslash\partial P} \operatorname{Res}_a(f) \;=\; 0.$$

(2) *If $f$ is not constant and has neither poles nor zeroes on $\partial P$, then*

$$(i) \qquad \sum_{a\in P\backslash\partial P} \operatorname{ord}_a(f) \;=\; 0,$$

$$(ii) \qquad \sum_{a\in P\backslash\partial P} a\cdot\operatorname{ord}_a(f) \;\in\; \Lambda.$$

(3) *Non-constant elliptic functions $f:\mathbb{C}/\Lambda\to\mathbb{P}^1(\mathbb{C})$ take any value $c\in\mathbb{P}^1(\mathbb{C})$ the same number of times when counted with multiplicities.*

*Proof.* (1) Since we assumed that $f$ has no poles on $\partial P$, the residue theorem says that

$$\sum_{a\in P\backslash\partial P} \operatorname{Res}_a(f) \;=\; \frac{1}{2\pi i}\int_{\partial P} f(z)dz.$$

But for the integral on the right hand side the contributions from opposite signs of the fundamental parallelogram cancel, because $f(z)=f(z+\lambda_1)=f(z+\lambda_2)$ and the sides are oriented opposite to each other.

(2) With $f$ also the quotient $f'/f$ is an elliptic function. Its poles are precisely the zeroes and poles of the original elliptic function, and by assumption none of these lies on the boundary $\partial P$. Applying part (1) to the elliptic function $f'/f$ and using that

$$\operatorname{ord}_a(f) \;=\; \operatorname{Res}_a(f'/f),$$

we obtain that

$$\sum_{a\in P\backslash\partial P} \operatorname{ord}_a(f) \;=\; \sum_{a\in P\backslash\partial P} \operatorname{Res}_a(f'/f) \;=\; 0$$

as claimed in (i). For claim (ii) note that

$$\frac{f'(z)}{f(z)} \;=\; \sum_a \frac{\operatorname{ord}_a(f)}{z-a} + g(z)$$

where $g:U\to\mathbb{C}$ is holomorphic. Multiplying by the function $z=(z-a)+a$ we get that

$$z\cdot\frac{f'(z)}{f(z)} \;=\; \sum_a a\cdot\frac{\operatorname{ord}_a(f)}{z-a} + h(z)$$

where $h(z)=g(z)+\sum_a a\cdot\operatorname{ord}_a(f)$ is again holomorphic. So the residue theorem gives

$$\frac{1}{2\pi i}\int_{\partial P} z\cdot\frac{f'(z)}{f(z)}dz \;=\; \sum_{a\in P\backslash\partial P} a\cdot\operatorname{ord}_a(f).$$

We want to show that the integral on the left lies inside the lattice $\Lambda$. For this we write

$$\int_{\partial P} \;=\; A_1 - A_2 \quad\text{where}\quad A_\mu \;=\; \int_{z_0}^{z_0+\lambda_\mu} - \int_{z_0+\lambda_\nu}^{z_0+\lambda_\nu+\lambda_\mu}$$

with $\nu = \mu \pm 1 \in \{0, 1\}$, where the two integrals on the right are taken over the straight line segments which are part of our chosen boundary of the fundamental parallelogram. Then

$$\begin{aligned} A_\mu &= \int_0^{\lambda_\mu} \left[ (z_0 + \zeta) \cdot \frac{f'(z_0 + \zeta)}{f(z_0 + \zeta)} - (z_0 + \lambda_\nu + \zeta) \cdot \frac{f'(z_0 + \lambda_\nu + \zeta)}{f(z_0 + \lambda_j + \zeta)} \right] d\zeta \\ &= -\lambda_\nu \int_0^{\lambda_\mu} \frac{f'(z_0 + \zeta)}{f(z_0 + \zeta)} d\zeta && \text{by periodicity of } f'/f \\ &= -\lambda_\nu \int_{\gamma_\mu} \frac{dz}{z} && \text{by the substitution rule} \\ &= -\lambda_\nu \cdot 2\pi i \cdot w_0(\gamma_\mu) && \text{by the winding number formula} \end{aligned}$$

for the closed loop

$$\gamma_\mu : \quad [0, 1] \to \mathbb{C} \setminus \{0\}, \quad t \mapsto f(z_0 + t\lambda_\mu).$$

Since winding numbers are integers, we obtain $\frac{1}{2\pi i} \cdot A_\mu \in \mathbb{Z}\lambda_\mu$ for $\mu = 1, 2$. This gives

$$\frac{1}{2\pi i} \int_{\partial P} z f'(z)/f(z) dz = \frac{1}{2\pi i} (A_1 - A_2) \in \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$$

and we are done.

(3) For $c \in \mathbb{C}$, put $g(z) = f(z) - c$. Then the number of times with which the value $c$ is taken by $f$ can be computed as

$$\sum_{a, g(a) = 0} \operatorname{ord}_a(g(z)) = - \sum_{a, g(a) = \infty} \operatorname{ord}_a(g(z)) = - \sum_{a, f(a) = \infty} \operatorname{ord}_a(f(z))$$

by part (2)(i) and so we are done. □

So far we haven't seen any non-constant elliptic function, but the above tells us that the simplest configuration of poles for such a function would be to have either two simple poles at opposite lattice points with opposite residues, or a double pole with no residue at a half-lattice point. Let's try to construct an example with the latter property. A naive candidate would be the infinite series $z \mapsto \sum_{\lambda \in \Lambda} \frac{1}{(z - \lambda)^2}$ but there are convergence issues: For instance, take $\Lambda = \mathbb{Z} \oplus \mathbb{Z}i$. Subtracting the pole $1/z^2$ we are left with

$$\sum_{\substack{(m,n) \in \mathbb{Z}^2 \\ (m,n) \neq (0,0)}} \frac{1}{(z - m - in)^2}$$

but this series is *not* absolutely convergent in any neighborhood of $z = 0$:

Lemma 4.3. *Let $\Lambda \subset \mathbb{C}$ be a lattice and $s \in \mathbb{R}$. Then we have the following convergence criterion:*

$$\sum_{\lambda \in \Lambda \setminus \{0\}} |\lambda|^{-s} < \infty \iff s > 2.$$

*Proof.* We first deal with the case $\Lambda = \mathbb{Z} \oplus \mathbb{Z}i$. Here the series converges iff the integral

$$\int_{x^2 + y^2 \geq 1} \frac{dx dy}{(x^2 + y^2)^{s/2}} = \int_0^{2\pi} \int_1^\infty \frac{r dr d\varphi}{r^s} = 2\pi \int_1^\infty \frac{dr}{r^{s-1}}$$

is finite, which happens iff $s > 2$. Now consider an arbitrary lattice $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$ with $\lambda_1, \lambda_2 \in \mathbb{C}$. We will be reduced to the previous case if we can show that there exist strictly positive real numbers $c_1, c_2 > 0$ depending only on $\lambda_1, \lambda_2 \in \mathbb{C}$ such that

$$c_1 \cdot (n_1^2 + n_2^2) \;\leq\; |n_1\lambda_1 + n_2\lambda_2|^2 \;\leq\; c_2 \cdot (n_1^2 + n_2^2) \quad \text{for all} \quad n_1, n_2 \in \mathbb{Z}.$$

So we only need to show that the function

$$f: \quad \mathbb{R}^2 \setminus \{(0,0)\} \;\longrightarrow\; \mathbb{R}, \quad (x_1, x_2) \;\mapsto\; \frac{|x_1\lambda_1 + x_2\lambda_2|^2}{x_1^2 + x_2^2}$$

is bounded above and below by some strictly positive number. By homogenuity it suffices to bound the function on the unit circle. There it takes a global maximum and a global minimum by compactness. The minimum is strictly positive since $f$ is so at every point, indeed $\lambda_1, \lambda_2$ are linearly independent over $\mathbb{R}$. □

We can now make our previous naive approach work by subtracting a constant error term from each summand in the divergent series:

Lemma 4.4. *Let $\Lambda \subset \mathbb{C}$ be a lattice. Then the series $\sum_{\lambda \in \Lambda \setminus \{0\}} \big[\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\big]$ converges uniformly on any compact subset of* $\mathbb{C} \setminus \Lambda$.

*Proof.* When $z$ stays in a compact subset of the complex plane, then for $|\lambda| \to \infty$ we have

$$\Big|\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\Big| \;=\; \frac{|z||z-2\lambda|}{|\lambda|^2|z-\lambda|^2} \;\sim\; \frac{c}{|\lambda|^3}$$

and so lemma 4.3 gives uniform convergence on any compact subset of $\mathbb{C} \setminus \Lambda$. □

Definition 4.5. We define the *Weierstrass function* of the lattice $\Lambda \subset \mathbb{C}$ to be the meromorphic function

$$\wp(z) \;=\; \wp_\Lambda(z) \;=\; 1/z^2 + \sum_{\lambda \in \Lambda \setminus \{0\}} \Big[\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\Big].$$

Its basic properties are given by the following

Lemma 4.6. *The Weierstrass function is an elliptic function with poles precisely in the lattice points, where the pole order is two and the residues are zero. It is an even function in the sense that $\wp(-z) = \wp(z)$. Its derivative is the odd elliptic function*

$$\wp'(z) \;=\; -2\sum_{\lambda \in \Lambda} \frac{1}{(z-\lambda)^3}$$

*which again has poles precisely in the lattice points, with pole order three and residue zero. Moreover*

$$\wp'(z) \;=\; 0 \quad \Longleftrightarrow \quad z \;\in\; \tfrac{1}{2}\Lambda \setminus \Lambda,$$

*and all these half-lattice points are simple zeroes of the derivative $\wp'(z)$.*

*Proof.* Since we already know locally uniform convergence of the series on $\mathbb{C} \setminus \Lambda$, the main point is to show that the Weierstrass function is elliptic. Note that its derivative $\wp'$ is a sum over translates by lattice points, hence obviously periodic with respect to the lattice. Writing $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$, we obtain that for both $i = 1, 2$ the function

$$z \;\mapsto\; \wp(z) - \wp(z + \lambda_i)$$

has derivative zero and must hence be equal to a constant $c_i$. Plugging in $z = \lambda_i/2$ we obtain

$$c_i \;=\; \wp(\lambda_i/2) - \wp(-\lambda_i/2) \;=\; 0$$

since $\wp$ is obviously an even function. This shows that the Weierstrass function is elliptic. The claim about the poles, their order and residues can be read off from the defining series. Finally, the derivative of the Weierstrass function is clearly odd, so we have $\wp'(z) = -\wp'(\lambda - z)$ for all $\lambda \in \Lambda$. Taking $z = \lambda/2$ with $\lambda \in \{\lambda_1, \lambda_2, \lambda_1 + \lambda_2\}$ we get

$$\wp'\Big(\frac{\lambda_1}{2}\Big) \;=\; \wp'\Big(\frac{\lambda_2}{2}\Big) \;=\; \wp'\Big(\frac{\lambda_1+\lambda_2}{2}\Big) \;=\; 0,$$

so we have found three distinct zeroes of the derivative. But we already know that the function $\wp'$ only has a single pole modulo $\Lambda$, with pole order three. Since a non-constant elliptic function takes every value the same number of times, it follows that $\wp'$ has precisely three zeroes when counted with multiplicities. Therefore we have found all the zeroes and the multiplicities are one. □

Recall from complex analysis that the sum, difference or product of meromorphic functions ois again a meromorphic function, and similarly for the quotient of a meromorphic function by a meromorphic function which is not identically zero on any connected component of its domain. For a compact Riemann surface $X$ the field

$$\mathbb{C}(X) \;=\; \{\text{meromorphic functions } f : X \to \mathbb{P}^1(\mathbb{C})\}$$

is called its *function field*. We can now describe *all* elliptic functions as follows:

THEOREM 4.7. *Let $X = \mathbb{C}/\Lambda$ and $\wp(z) = \wp_\Lambda(z)$ as above.*

(1) *Any even elliptic function $F \in \mathbb{C}(X)$ with poles at most in $\Lambda$ can be written uniquely as*

$$F(z) \;=\; f(\wp(z))$$

*where $f(x) \in \mathbb{C}[x]$ is a polynomial of degree* $\deg(f) = \deg(g)/2$.

(2) *More generally, every even elliptic function $F(z) \in \mathbb{C}(X)$ can be written uniquely as*

$$F(z) \;=\; h(\wp(z)) \quad \textit{for a rational function} \quad h(x) \;=\; \frac{f(x)}{g(x)} \;\in\; \mathbb{C}(x).$$

(3) *For any elliptic function $F(z) \in \mathbb{C}(X)$ there are unique $h_i(x) \in \mathbb{C}(x)$ such that*

$$F(z) \;=\; h_1(\wp(z)) + h_2(\wp(z)) \cdot \wp'(z).$$

*Proof.* (1) We may assume that $F$ is not constant and hence has a pole at $z = 0$, since by assumption it is periodic with respect to the lattice and has poles at most in the lattice points. Since $F$ is an even function, it follows that it Laurent series has the form

$$F(z) \;=\; \sum_{i \geq -d} c_{2i} \cdot z^{2i} \quad \text{with} \quad c_{-2d} \;\neq\; 0 \quad \text{for} \quad d \;=\; \deg(F)/2 \;\geq\; 1.$$

So the difference $\tilde{F}(z) = \varphi(z) - c_{-2d} \cdot \wp(z)^d$ is an even elliptic function with poles at most in the lattice, and we are done by induction since $\deg(\tilde{F}) < \deg(F)$.

(2) Suppose that $a \in \mathbb{C} \setminus \Lambda$ is a non-lattice point but a pole of the even elliptic function $F(z) \in \mathbb{C}(X)$. Since the only poles of the Weierstrass function are the lattice points, it follows in particular that $\wp(a) \neq \infty$. Hence for $N \gg 0$ the even elliptic function

$$F_1(z) \;=\; (\wp(z) - \wp(a))^N \cdot F(z) \quad \text{has} \quad F_1^{-1}(\infty) \;\subseteq\; F^{-1}(\infty) \setminus (a + \Lambda).$$

If this function has still a pole which is not a lattice point, we can repeat the argument until we have $a_1, \ldots, a_n \in \mathbb{C} \setminus \Lambda$, $N_1, \ldots, N_n \in \mathbb{N}$ such that the even elliptic function

$$F_n(z) \;=\; F(z) \cdot \prod_{i=1}^{n} (\wp(z) - \wp(a_i))^{N_i}$$

has poles at most in lattice points. Then by part (1) we know $F_n(z) = f(\wp(z))$ for some $f(x) \in \mathbb{C}[x]$. Dividing by

$$g(x) \;=\; \prod_{i=1}^{n} (x - \wp(a_i))^{N_i} \;\in\; \mathbb{C}[x]$$

we obtain the desired representation of $F(z)$ as a rational function in $\wp(z)$.

(3) This follows from (2) by writing $f$ as the sum of an even and an odd elliptic function

$$f(z) \;=\; \frac{f(z)+f(-z)}{2} + \frac{f(z)-f(-z)}{2}$$

and using that any odd elliptic function is an even elliptic function times $\wp'(z)$. □

Corollary 4.8. *We have* $(\wp'(z))^2 = f(\wp(z))$ *for the cubic polynomial* $f(x)$ *given by*

$$f(x) \;=\; 4(x-e_1)(x-e_2)(x-e_3) \quad \textit{where} \quad \begin{cases} e_1 \;=\; \wp(\frac{\lambda_1}{2}), \\ e_2 \;=\; \wp(\frac{\lambda_2}{2}), \\ e_3 \;=\; \wp(\frac{\lambda_1+\lambda_2}{2}). \end{cases}$$

*Hence* $X = \mathbb{C}/\Lambda$ *is isomorphic to the compact Riemann surface associated to the elliptic curve*

$$E \;=\; \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\} \cup \{\infty\}.$$

*Proof.* Since $(\wp'(z))^2$ is an even elliptic function with poles only in the lattice points, the first part of theorem 4.7 shows that there exists a cubic $f(x) \in \mathbb{C}[x]$ with $(\wp'(z))^2 = f(\wp(z))$. To verify the given explicit form of this cubic, note that the elliptic function

$$h(z) \;=\; (\wp'(z))^2 - 4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$$

can have poles at most in the lattice points $z \in \Lambda$. Its pole order there can be read off from the Laurent expansion around the origin. By inserting $\wp(z) = z^{-2} + \cdots$ and $\wp'(z) = -2z^{-3} + \cdots$ we find that the poles of order six of the two summands cancel and so

$$\mathrm{ord}_0(h) \;\geq\; -4$$

since $h$ is an even function. As it has no poles outside the lattice it follows that $h$ is either constant or takes any value at most four times with multiplicities. On the other hand

$$h(\tfrac{\lambda_1}{2}) \;=\; h(\tfrac{\lambda_2}{2}) \;=\; h(\tfrac{\lambda_1+\lambda_2}{2}) \;=\; 0$$

and the order of vanishing at each of these three zeroes is even because $h$ is an even elliptic function. Therefore the total multiplicity of the value zero is at least six and so $h$ must be identically zero as required. For the final statement, we have a well-defined holomorphic map

$$\varphi_0: \quad X_0 \;=\; X \setminus \{0\} \;\longrightarrow\; E_0 \;=\; \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\}, \quad z \;\mapsto\; (\wp(z), \wp'(z))$$

where $0 \in X$ denotes the image of $\Lambda \subset \mathbb{C}$ under the map $\mathbb{C} \twoheadrightarrow X = \mathbb{C}/\Lambda$. The composite of this morphism of Riemann surfaces with the projection $(x,y) \mapsto x$ extends to the morphism

$$\wp: \quad X \;\twoheadrightarrow\; \mathbb{P}^1(\mathbb{C}).$$

Since the Weierstrass function takes every value precisely twice and $pr_2 : E_0 \to \mathbb{C}$ is a branched double cover, it follows that the morphism $\varphi_0$ is bijective and hence an isomorphism of branched double covers. By the unique extension properties of branched covers it follows that it extends to an isomorphism $\varphi : X \xrightarrow{\sim} E$. □

Note that the argument by which we computed the cubic polynomial $f(x)$ is basically the algorithm that we used in the proof of theorem 4.7. We can make it more explicit by keeping track of further terms in the Laurent expansions. Recall that

$$\wp(z) \;=\; z^{-2} + g(z) \quad \text{with} \quad g(z) \;=\; \sum_{\lambda \in \Lambda \setminus \{0\}} \Big[ \frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \Big].$$

By induction

$$g^{(n)}(z) \;=\; (-1)^n (n+1)! \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{(z-\lambda)^{n+2}}$$

for all $n \in \mathbb{N}$. Hence the nonvanishing Taylor coefficients of the even function $g(z)$ are

$$\frac{g^{(2n)}(0)}{(2n)!} \;=\; (2n+1) G_{2n+2} \quad \text{with} \quad G_{2n+2} \;=\; \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^{2n+2}}.$$

The series on the right are called *Eisenstein series* and play an important role in the theory of modular forms. From the above we get

$$\wp(z) \;=\; \frac{1}{z^2} + \sum_{n \geq 1} (2n+1) G_{2n+2} z^{2n}$$

Corollary 4.9. *The polynomial $f(x) \in \mathbb{C}[x]$ from the previous corollary has the form*

$$f(x) \;=\; 4x^3 - g_2 x - g_3 \quad \text{where} \quad \begin{cases} g_2 \;=\; 60 G_4, \\ g_3 \;=\; 140 G_6. \end{cases}$$

*Proof.* Consider the above Taylor expansion and its derivative. Take the cube of the first and the square of the second:

$$\begin{aligned} \wp(z) \;&=\; z^{-2} + 3G_4 z^2 + 5G_6 z^4 + \cdots \\ \wp'(z) \;&=\; -2z^{-3} + 6G_4 z + 20 G_6 z^3 + \cdots \\ (\wp(z))^3 \;&=\; z^{-6} + 9G_4 z^{-2} + 15 G_6 + \cdots \\ (\wp'(z))^2 \;&=\; 4z^{-6} - 24 G_4 z^{-2} - 80 G_6 + \cdots \end{aligned}$$

It follows that $(\wp'(z))^2 - 4(\wp(z))^3 + 60 G_4 \wp(z) = -140 G_6 + \cdots$. The left hand side is an elliptic function, but it has no poles since the right hand side doesn't. Thus it must be constant, equal to $-140 G_6$. Now use the uniqueness in theorem 4.7. □

## 5. Geometric form of the group law

We have seen in corollary 3.11 that elliptic curves over the complex numbers are complex tori, and conversely for any complex torus $X = \mathbb{C}/\Lambda$ corollary 4.8 gives the isomorphism

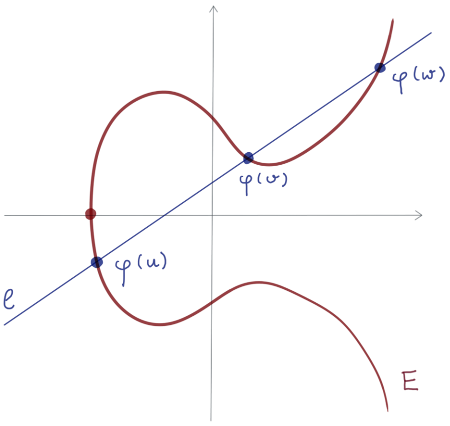
$$\varphi : \quad X \;\xrightarrow{\sim}\; E \;=\; \{(x,y) \in \mathbb{C}^2 \mid y^2 = f(x)\} \cup \{\infty\}, \quad z \;\mapsto\; (\wp(z), \wp'(z))$$

onto an elliptic curve. Now any complex torus $X$ has a natural group structure as a quotient of the additive group of complex numbers $(\mathbb{C}, +)$. On the corresponding elliptic curve $E$ this group structure has the following geometric interpretation:

Theorem 5.1. *In the above setting, let $u, v, w \in X = \mathbb{C}/\Lambda$ be pairwise distinct, then the following are equivalent:*

(1) *We have $u + v + w = 0$ on the complex torus $X$.*

(2) *The three points $\varphi(u), \varphi(v), \varphi(w) \in E$ are collinear as shown below.*



$u + v + w = 0$

*Proof.* Let $u, v, w \in X$ be pairwise distinct points with $u + v + w = 0$. By symmetry we may assume that $u, v \neq 0$. Now recall from the residue theorem 4.2 that for every non-constant elliptic function without poles or zeroes on the boundary of a chosen fundamental parallelogram, the sum of its zeroes and poles inside this parallelogram lies in $\Lambda$ when counted with multiplicities. We apply this to the function

$$f(z) \;=\; \det \begin{pmatrix} 1 & \wp(z) & \wp'(z) \\ 1 & \wp(u) & \wp'(u) \\ 1 & \wp(v) & \wp'(v) \end{pmatrix}$$

Expanding the determinant we see that this is an elliptic function of order three with zeroes at $z = u$ and at $z = v$. Since it can have poles at most in the lattice points, it follows that that its unique third zero must be $z = w$ unless $w = 0$.

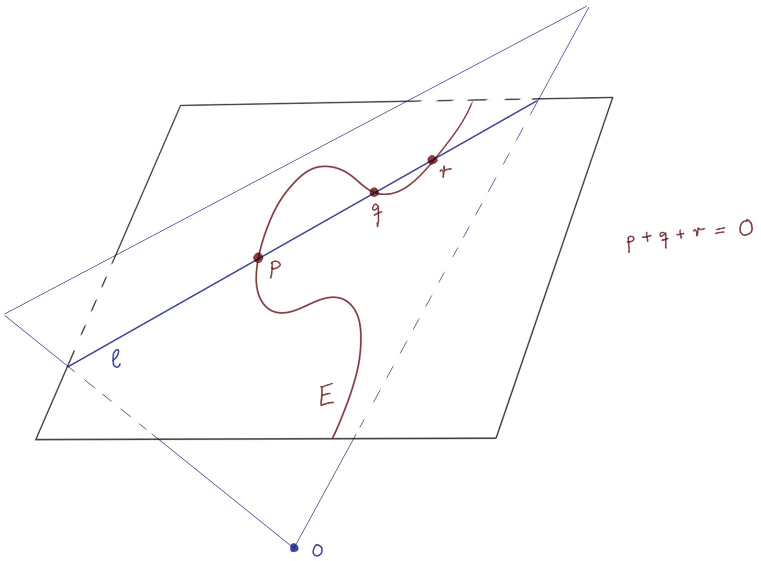
Let us see what this means geometrically. If $u, v, w \neq 0$ are all different from the origin $0 \in X$, then $\wp$ and $\wp'$ take a finite value at these points and $f(w) = 0$ means that the vectors

$$\begin{pmatrix} 1 \\ \wp(u) \\ \wp'(u) \end{pmatrix}, \begin{pmatrix} 1 \\ \wp(v) \\ \wp'(v) \end{pmatrix}, \begin{pmatrix} 1 \\ \wp(w) \\ \wp'(w) \end{pmatrix} \;\in\; \mathbb{C}^3$$

are linearly dependent, i.e. they lie inside a common plane. Intersecting with the affine plane of all vectors whose first coordinate is one, we see that this happens iff the three points

$$\varphi(u),\; \varphi(v),\; \varphi(u+v) \;\in\; \mathbb{C}^2$$

are collinear, i.e. they lie on a common affine line as in the following picture:

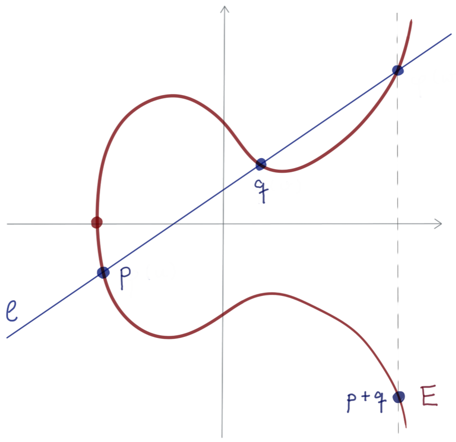


The remaining case where $w = 0$ and $u = -v$ can be understood as a limiting case of the previous one. The lines through the point at infinity $\infty \in E$ should be taken to be the lines parallel to the $y$-axis in the complex plane, each intersects the affine part $E_0 \subset \mathbb{C}^2$ in precisely two points

$$(x, \pm y) \;=\; (\wp(u), \pm\wp(u))$$

which is in accordance with the fact that $\wp(u) = \wp(-u)$ while $\wp'(u) = -\wp'(-u)$. □

Of course the assumption that $u, v, w$ are pairwise distinct was only made for simplicity, the statement holds more generally: If two of the points come together, the line through them should be understood as a tangent line at that point. Note also that the neutral element for the group $E$ is the point at infinity. The lines through infinity are parallels to the $y$-axis, so the negative of a point $(x, y)$ on the affine part of the elliptic curve is the point $(x, -y)$. Thus to compute the sum of two points $p, q \in E$ we take the line $\ell$ through these points and put $E \cap \ell = \{p, q, r\}$, then we obtain the sum $p + q$ by reflecting $r$ along the $x$-axis:



The group law on elliptic curves was known already to Euler as a relation for elliptic integrals, in the form that there exists an algebraic function $w = w(u, v)$ such that

$$\int_u^\infty \frac{dx}{\sqrt{f(x)}} + \int_v^\infty \frac{dx}{\sqrt{f(x)}} + \int_w^\infty \frac{dx}{\sqrt{f(x)}} \;=\; 0.$$

More precisely:

Corollary 5.2. *Let $\gamma_1, \gamma_2, \gamma_3 : [0,1] \to E$ be traced by a family of lines $\ell(t)$ in the sense that*

$$E_0 \cap \ell(t) \;=\; \{\gamma_1(t), \gamma_2(t), \gamma_3(t)\}$$

*for all $t \in [0,1]$, then*

$$\int_{\gamma_1} \frac{dx}{y} + \int_{\gamma_2} \frac{dx}{y} + \int_{\gamma_3} \frac{dx}{y} \;=\; 0.$$

At this point it may be convenient to recall the analogy between the complex logarithm and elliptic integrals:

$$\begin{tikzcd}
\widetilde{\mathbb{C}^*} \arrow[r, equal] \arrow[d, "p=\exp"'] & \mathbb{C} \arrow[d] \\
\mathbb{C}^* \arrow[r, "\log=\int \frac{dz}{z}"'] & \mathbb{C}/2\pi i\mathbb{Z}
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
\widetilde{X} \arrow[r, equal] \arrow[d, "p"'] & \mathbb{C} \arrow[d] \\
X \arrow[r, "\int \frac{dx}{\sqrt{f(x)}}"'] & \mathbb{C}/\Lambda
\end{tikzcd}$$

The above corollary is the precise counterpart of the additivity of the logarithm that we have seen in example 1.5, and both really express the fact that the universal covering map is a group homomorphism.

## 6. Abel's theorem

The essential point for the above was that by the residue theorem 4.2 the poles and zeroes of an elliptic function sum up to a lattice point when counted with multiplicities. With a bit more work we can show that this condition is not only necessary but also sufficient for the existence of an elliptic function with given poles and zeroes. To formulate the result we introduce the following notion:

Definition 6.1. By a *divisor* on the compact Riemann surface $X = \mathbb{C}/\Lambda$ we mean a finite formal sum

$$D \;=\; \sum_{p \in X} n_p\,[p]$$

where $n_p \in \mathbb{Z}$ are almost all zero. In what follows we denote by $\mathrm{Div}(X)$ the group of all such divisors. We say that $D$ is a *principal divisor* if there is a meromorphic function $f \in \mathbb{C}(X) \setminus \{0\}$ such that $n_p = \mathrm{ord}_p(f)$ for all $p \in X$, in which case we write

$$D \;=\; \mathrm{div}(f).$$

Since $\mathrm{div}(fg) = \mathrm{div}(f) + \mathrm{div}(g)$, the map sending a meromorphic function to its principal divisor is a group homomorphism $\mathrm{div} : \mathbb{C}(X)^\times \to \mathrm{Div}(X)$. So principal divisors form a subgroup $\mathrm{PDiv}(X) \subset \mathrm{Div}(X)$. We will see that this subgroup can be characterized easily using the group structure on the complex torus:

We know from the residue theorem 4.2 that any elliptic function has the same number of zeroes and poles on $X$ when counted with multiplicities. So any principal divisor has the form

$$\sum_{i=1}^{n}[a_i] - \sum_{i=1}^{n}[b_i] \quad \text{with} \quad a_i, b_i \in \mathbb{C},$$

where we allow repetitions to account for multiplicities. Moreover we have seen as an application of the residue theorem that any such principal divisor satisfies the further condition

$$\sum_{i=1}^{n}(a_i - b_i) \;\in\; \Lambda.$$

It turns out that this necessary condition is also sufficient:

THEOREM 6.2 (Abel). *For any complex numbers $a_i, b_i$ with $\sum_{i=1}^{n}(a_i - b_i) \in \Lambda$ there exists an elliptic function $f \in \mathbb{C}(X)$ whose divisor of poles and zeroes is given by*

$$\operatorname{div}(f) \;=\; \sum_{i=1}^{n}[a_i] - \sum_{i=1}^{n}[b_i].$$

*Proof.* Since principal divisors form a subgroup, it suffices show that any divisor of the form

$$[a] + [b] - [c] - [0] \quad \text{with} \quad a + b - c \;\in\; \Lambda$$

is principal. In fact we only care about the points on the complex torus $X = \mathbb{C}/\Lambda$, so we can even assume that $a + b = c$. Fixing a point $c \in \mathbb{C}$, we want to show that for any $a, b \in \mathbb{C}$ with $a + b = c$ there exists an elliptic function $f \in \mathbb{C}(X)$ with divisor

$$\operatorname{div}(f) \;=\; [a] + [b] - [c] - [0].$$

Recall that by theorem 4.7 any elliptic function is a rational function in $\wp(z)$, $\wp'(z)$, so in principle we know where we to look for our functions. Let us first get rid of the case $c \in \Lambda$. In this case we want elliptic functions with poles only in the lattice points, where the pole order should be two. We may assume that $b = -a \neq 0$ and then

$$f(z) \;=\; \wp(z) - \wp(a) \quad \text{has} \quad \operatorname{div}(f) \;=\; [a] + [-a] - 2 \cdot [0]$$

since the Weierstrass function is an even elliptic function of degree two with poles only in the lattice points. So in what follows we will assume $c \notin \Lambda$. We want elliptic functions with poles only in the two points $[0]$ and $[c]$, where the pole order should be one since these two points are distinct. One example of such an elliptic function is

$$f_0(z) \;=\; \frac{\wp'(z) - \wp'(-c)}{\wp(z) - \wp(-c)}$$

because clearly the set of its poles modulo the lattice is contained in $\{\pm c, 0\}$ and we have

$$\operatorname{ord}_{z_0}(\wp(z)-\wp(-c)) \;=\; \begin{cases} -2, \\ +1, \\ +1, \\ +2, \end{cases} \qquad \operatorname{ord}_{z_0}(\wp'(z)-\wp'(-c)) \;=\; \begin{cases} -3 & \text{if } z_0 = 0, \\ \phantom{+}0 & \text{if } z_0 = +c \notin \tfrac{1}{2}\Lambda, \\ +1 & \text{if } z_0 = -c \notin \tfrac{1}{2}\Lambda, \\ +1 & \text{if } z_0 = \pm c \in \tfrac{1}{2}\Lambda, \end{cases}$$

which implies

$$\operatorname{div}(f_0) \;=\; [a_0] + [b_0] - [c] - [0] \quad \text{for some} \quad a_0, b_0 \in \mathbb{C}, \quad a_0 + b_0 = c.$$

To pass from this example to the general case we use interpolation: For $\lambda \in \mathbb{C}$ we consider the function

$$f_\lambda(z) \;=\; \lambda + (1-\lambda) f_0(z) \;\in\; \mathbb{C}(X).$$

If $\lambda \neq 1$, then this is an elliptic function with poles only in the two points $[0]$, $[c]$, and these poles are simple. Thus

$$\operatorname{div}(f_\lambda) \;=\; [a_\lambda] + [b_\lambda] - [c] - [0] \quad \text{for some} \quad a_\lambda, b_\lambda \in \mathbb{C}, \quad a_\lambda + b_\lambda = c.$$

Notice that modulo the lattice the two points $[a_\lambda], [b_\lambda] \in X = \mathbb{C}/\Lambda$ are determined uniquely up to permutation. The unordered pair of the two points is a well-defined element of the set

$$S \;=\; S(c) \;=\; \Big\{(p,q) \in X^2 \,\Big|\, p = [a], q = [b], a+b = c\Big\}\Big/ (p,q) \sim (q,p).$$

This gives a continuous map

$$\begin{aligned} \varphi: \quad \mathbb{P}^1(\mathbb{C}) \;&\longrightarrow\; S \\ \lambda \;&\mapsto\; f_\lambda^{-1}(0) \;=\; \big\{[a_\lambda], [b_\lambda]\big\} \end{aligned}$$

where the set bracket on the right refers to a multiset in case that $[a_\lambda] = [b_\lambda]$; this is clear on $\mathbb{P}^1(\mathbb{C}) \setminus \{1, \infty\}$, and at $\lambda = 1$ and $\lambda = \infty$ one checks that the map extends continuously with

$$\begin{aligned} \lim_{\lambda \to 1} \varphi(\lambda) \;&=\; f_0^{-1}(\infty), \\ \lim_{\lambda \to \infty} \varphi(\lambda) \;&=\; f_0^{-1}(1). \end{aligned}$$

For the proof of the theorem we have to show that this map $\varphi : \mathbb{P}^1(\mathbb{C}) \to S$ is surjective. Note that $\varphi(\mathbb{P}^1(\mathbb{C})) \subseteq S$ is a closed subset, being compact as the image of a compact set under a continuous map. So we will be done if we can show that the image $\varphi(\mathbb{P}^1(\mathbb{C})) \subseteq S$ contains an open dense subset of $S$. For this it will clearly be enough to show that the image contains an open dense subset of $U = S \setminus \Delta$, where

$$\Delta \;=\; \{(p,p) \in X^2 \mid p \;=\; [a],\; 2a \;=\; c\,\} \;\subset\; S$$

denotes the diagonal. For this note that the preimage $\varphi^{-1}(U) \subseteq \mathbb{P}^1(\mathbb{C})$ is nonempty since $\varphi(1) \in U$ by our assumption that $[c] \neq [0]$. It will therefore be enough to show that the restriction

$$\varphi^{-1}(U) \;\longrightarrow\; U \;=\; S \setminus \Delta$$

is an open map. For this note that for any point $(p,q) \in U$ we can find a coordinate neighborhood $p \in V \subset X$ which is small enough to have $[a] \neq [c-a]$ for all $[a] \in V$ and then

$$V \;\hookrightarrow\; U, \quad [a] \;\mapsto\; \{[a], [c-a]\}$$

will be a coordinate chart on the set of unordered pairs of distinct points on $X$ with sum $[c]$. If $W \subset U$ denotes the image of this coordinate chart, it will suffice to show that the restriction

$$\varphi^{-1}(W) \;\longrightarrow\; W \;\simeq\; V$$

is an open map. But locally the representation of zeroes of holomorphic functions by a path integral shows that near any given point $\lambda_0 \in \varphi^{-1}(W)$ this map is given by

$$\lambda \;\mapsto\; a_\lambda \;=\; \oint_{|z - a_{\lambda_0}| = \varepsilon} \frac{z \cdot f_\lambda'(z)}{f_\lambda(z)} dz$$

for $\varepsilon > 0$ small enough. The integral on the right depends holomorphically on $\lambda$ and so the claim follows from the fact that holomorphic maps are open. □

Abel's theorem can be reformulated conveniently as follows. The first of the two necessary conditions says that the principal divisors are contained in the kernel of the homomorphism

$$\deg : \quad \mathrm{Div}(X) \twoheadrightarrow \mathbb{Z}, \quad D = \sum_{p\in X} n_p\,[p] \mapsto \deg(D) = \sum_{p\in X} n_p$$

sending a divisor to the sum of its multiplicities. We denote by $\mathrm{Div}(X)^0 = \ker(\deg)$ this kernel and by

$$\mathrm{Pic}^0(X) \subset \mathrm{Pic}(X) = \mathrm{Div}(X)/\mathrm{PDiv}(X)$$

its image in the *Picard group* of divisors modulo principal divisors as indicated in the diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
 & \mathrm{PDiv}(X) \arrow[r, equal] \arrow[d] & \mathrm{PDiv}(X) \arrow[d] & & \\
0 \arrow[r] & \mathrm{Div}^0(X) \arrow[r] \arrow[d] & \mathrm{Div}(X) \arrow[r] \arrow[d] & \mathbb{Z} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathrm{Pic}^0(X) \arrow[r] \arrow[d] & \mathrm{Pic}(X) \arrow[r] \arrow[d] & \mathbb{Z} \arrow[r] & 0 \\
 & 0 & 0 & &
\end{tikzcd}$$

where both rows and columns are exact. Abel's theorem then says:

Corollary 6.3. *For $X = \mathbb{C}/\Lambda$, we have an isomorphism* $\mathrm{Pic}^0(X) \xrightarrow{\sim} X$.

*Proof.* Since $\mathrm{Div}(X)$ is the free abelian group on the points of $X$, we may define a homomorphism by

$$\varphi : \quad \mathrm{Div}(X) \longrightarrow X, \quad \sum_p n_p\,[p] \mapsto \Big[\sum_p n_p \cdot p\Big].$$

This homomorphism remains surjective when restricted to the subgroup $\mathrm{Div}^0(X)$ since $[a] = \varphi([a]-[0])$ for any $a \in \mathbb{C}$. Now Abel's theorem says that the kernel of the surjective group homomorphism $\varphi : \mathrm{Div}^0(X) \twoheadrightarrow X$ is precisely $\mathrm{PDiv}(X)$. □

## 7. The $j$-invariant

In this section we want to classify all complex tori up to isomorphism. We begin with the following

Lemma 7.1. *Let $\Lambda, \Lambda' \subset \mathbb{C}$ be lattices. Then any morphism $f : \mathbb{C}/\Lambda' \to \mathbb{C}/\Lambda$ is induced by an affine-linear map*

$$\tilde{f} : \quad \mathbb{C} \longrightarrow \mathbb{C}, \quad z \mapsto \alpha z + \gamma \quad \textit{for some } \alpha, \gamma \in \mathbb{C} \textit{ with } \alpha\Lambda' \subseteq \Lambda.$$

*In particular,*

$$\mathbb{C}/\Lambda' \simeq \mathbb{C}/\Lambda \iff \exists \alpha \in \mathbb{C}^* \textit{ with } \Lambda' = \alpha\Lambda.$$

*Proof.* Pick any $\gamma \in \mathbb{C}$ with $f(0) = [\gamma] \in \mathbb{C}/\Lambda$. Since the complex plane is simply connected, the unique lifting property for the covering map $\mathbb{C} \to \mathbb{C}/\Lambda$ gives a unique continuous map $\tilde{f}$ with $\tilde{f}(0) = \gamma$ such that the following diagram commutes:

$$\begin{tikzcd}
\mathbb{C} \arrow[r, dotted, "\exists ! \tilde{f}"] \arrow[d] & \mathbb{C} \arrow[d] \\
\mathbb{C}/\Lambda' \arrow[r, "f"] & \mathbb{C}/\Lambda
\end{tikzcd}$$

Since $f$ is morphism of Riemann surfaces, the function $\tilde{f} : \mathbb{C} \to \mathbb{C}$ is holomorphic everywhere by our definition of the Riemann surface structure on complex tori. By the unique lifting property we also know that the function $z \mapsto \tilde{f}(z+\lambda) - \tilde{f}(z)$ is constant for any $\lambda \in \Lambda'$. So for the derivative of our lifted map we obtain the periodicity

$$\tilde{f}'(z+\lambda) \;=\; \tilde{f}'(z) \quad \text{for all} \quad \lambda \;\in\; \Lambda',$$

hence $\tilde{f}'$ is a bounded holomorphic function on the complex plane and therefore constant by Liouville's theorem. Hence the result follows. □

Motivated by the above, we say that two lattices $\Lambda, \Lambda' \subset \mathbb{C}$ are *equivalent* and write $\Lambda \sim \Lambda'$ if there exists $\alpha \in \mathbb{C}^*$ with $\Lambda' = \alpha\Lambda$. This clearly defines an equivalence relation on the set of all lattices, and any equivalence class contains a lattice of the form

$$\Lambda_\tau \;=\; \mathbb{Z} \oplus \mathbb{Z}\tau \quad \text{with} \quad \tau \;\in\; \mathbb{H} \;=\; \{z \in \mathbb{C} \mid \mathrm{Im}(z) \;>\; 0\},$$

since we may rescale it to have the first basis vector to be one and then change the sign of the second basis vector to have positive imaginary part. Two lattices of this form may still be equivalent but we can say precisely when this happens:

LEMMA 7.2. *(a) The upper half plane $\mathbb{H}$ is endowed with a natural action of the group*

$$\Gamma \;=\; Sl_2(\mathbb{Z}) \;=\; \Bigl\{ \bigl(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\bigr) \in Sl_2(\mathbb{R}) \;\Big|\; a,b,c,d \in \mathbb{Z} \Bigr\}$$

*via biholomorphic automorphisms*

$$M \cdot \tau \;=\; \frac{a\tau + b}{c\tau + d} \quad \textit{for} \quad M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; \Gamma \quad \textit{and} \quad \tau \;\in\; \mathbb{H}.$$

*(b) For $\tau, \tau' \in \mathbb{H}$ we have the equivalence $\Lambda_\tau \sim \Lambda_{\tau'}$ iff $\tau = M \cdot \tau'$ for some $M \in \Gamma$.*

*Proof.* (a) That $\Gamma \subset Sl_2(\mathbb{R})$ is a subgroup follows from Cramer's formula, which here reads

$$M^{-1} \;=\; \frac{1}{\det(M)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{for} \quad M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; Sl_2(\mathbb{R}).$$

For $\tau \in \mathbb{H}$ and $\det(M) = 1$ we get

$$\begin{aligned}
\mathrm{Im}(M \cdot \tau) \;&=\; \mathrm{Im}\biggl(\frac{a\tau + b}{c\tau + d}\biggr) \\
&=\; \frac{\mathrm{Im}\bigl((a\tau + b)(c\overline{\tau} + d)\bigr)}{|c\tau + d|^2} \\
&=\; (ad - bc) \cdot \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2} \\
&=\; \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2} \;>\; 0,
\end{aligned}$$

hence $M \cdot \tau \in \mathbb{H}$. One easily checks that this gives a group action $\Gamma \to Aut(\mathbb{H})$.

(b) By definition $\Lambda_{\tau'} \sim \Lambda_\tau$ iff there exists $\alpha \in \mathbb{C}^*$ with $\Lambda_{\tau'} = \alpha \cdot \Lambda_\tau$. Now one computes

$$\begin{aligned}
\Lambda_{\tau'} \subseteq \alpha \cdot \Lambda_\tau \quad &\Longrightarrow \quad \mathbb{Z} \oplus \mathbb{Z}\tau' \;\subseteq\; \alpha(\mathbb{Z} \oplus \mathbb{Z}\tau) \\
&\Longrightarrow \quad \exists a,b,c,d \in \mathbb{Z} : \begin{cases} \tau' = \alpha(a\tau + b) \\ 1 \;\; = \alpha(c\tau + d) \end{cases} \\
&\Longrightarrow \quad \exists a,b,c,d \in \mathbb{Z} : \; \tau' \;=\; \frac{\tau'}{1} \;=\; \frac{a\tau + b}{c\tau + d}.
\end{aligned}$$

Moreover, if equality holds on the left hand side, then by symmetry it follows that the integer matrix

$$M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; \mathrm{Mat}_{2\times 2}(\mathbb{Z}) \cap Gl_2(\mathbb{R})$$

is invertible. The determinant of any invertible matrix with integer entries is $\pm 1$, and in our case

$$\mathrm{Im}(M\tau) \;=\; \det(M) \cdot \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2} \;>\; 0$$

implies that $\det(M) > 0$ as required. □

Thus the isomorphism classes of complex tori are in bijection with points of the quotient $\mathbb{H}/\Gamma$ by the action of the *modular group* $\Gamma = Sl_2(\mathbb{Z})$. Note that this action factors over

$$PSl_2(\mathbb{Z}) \;=\; Sl_2(\mathbb{Z})/\langle \pm 1 \rangle,$$

so we need to consider matrices only up to a sign. The quotient $\mathbb{H}/\Gamma$ inherits from the upper half plane a unique structure of a Riemann surface such that the quotient map $\mathbb{H} \to \mathbb{H}/\Gamma$ is holomorphic, and we will discuss it in two steps:

(1) Describe $\mathbb{H}/\Gamma$ as a topological space by a glueing of boundary points for a suitable closed fundamental domain $\mathscr{F} \subset \mathbb{H}$.

(2) Describe $\mathbb{H}/\Gamma$ as a Riemann surface by finding $\Gamma$-invariant holomorphic functions on the upper half plane.

Let us begin with the first step:

Proposition 7.3. *Put* $\mathscr{F} = \{\tau \in \mathbb{H} \mid |\tau| \geq 1, |\mathrm{Re}(\tau)| \leq 1/2\} \subset \mathbb{H}$.

(1) *Any point of* $\mathbb{H}$ *can be moved to a point of* $\mathscr{F}$ *by an iterated application of the matrices*

$$S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right), T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma = Sl_2(\mathbb{Z})$$

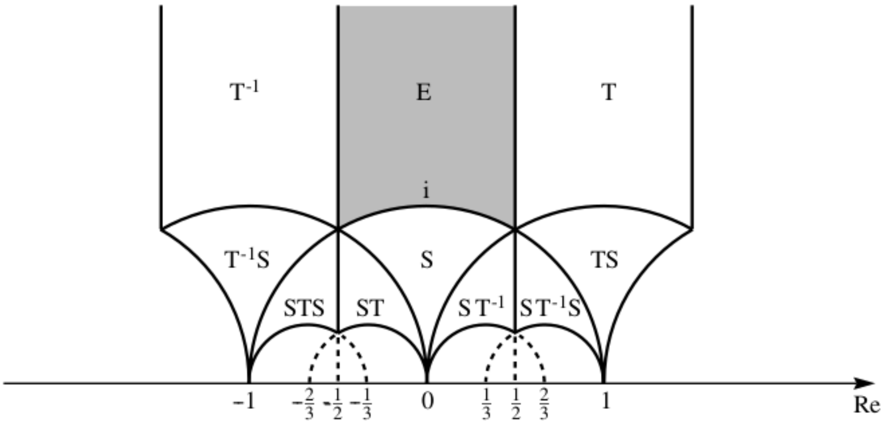
*as shown in the figure below. In fact* $\Gamma$ *is generated by these matrices.*

(2) *For* $M \in \Gamma \setminus \{\pm E\}$ *the intersection* $\mathscr{F} \cap M\mathscr{F}$ *is nonempty precisely in the following cases:*

| $M$ | $\mathscr{F} \cap M\mathscr{F}$ |
|---|---|
| $\pm T$ | $\{\tau \in \mathbb{H} \mid \lvert\tau\rvert \geq 1, \mathrm{Re}(\tau) = +1/2\}$ |
| $\pm T^{-1}$ | $\{\tau \in \mathbb{H} \mid \lvert\tau\rvert \geq 1, \mathrm{Re}(\tau) = -1/2\}$ |
| $\pm S$ | $\{\tau \in \mathbb{H} \mid \lvert\tau\rvert = 1, \mathrm{Re}(\tau) \leq 1/2\}$ |
| $\pm A_1$ | $\{\tau = \exp(2\pi i/3)\}$ |
| $\pm A_2$ | $\{\tau = \exp(2\pi i/6)\}$ |

*where*

$$A_\nu \;=\; \left(\begin{smallmatrix} 0 & -1 \\ 1 & -\epsilon_\nu \end{smallmatrix}\right), \left(\begin{smallmatrix} \epsilon_\nu & -1 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} \epsilon_\nu & 0 \\ 1 & \epsilon_\nu \end{smallmatrix}\right) \quad \textit{for the sign} \quad \epsilon_\nu \;=\; (-1)^\nu.$$



*Proof.* (1a) Let $z \in \mathbb{H}$. We have seen at the beginning of the proof of lemma 7.2 that

$$\mathrm{Im}(Mz) \;=\; \frac{\mathrm{Im}(z)}{|cz+d|^2} \quad \text{for} \quad M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; \Gamma$$

and for fixed $z \in \mathbb{H}$ this goes to zero when at least one of the entries of $(c,d) \in \mathbb{Z}^2$ goes to infinity. Hence for any subset $\Gamma' \subset \Gamma$ it follows that

$$\sup_{M \in \Gamma'} \mathrm{Im}(Mz) \;<\; \infty.$$

Taking $\Gamma' = \langle S, T \rangle$, we can therefore assume after an iterated application of $S$, $T$ that

$$\mathrm{Im}(z) \;\geq\; \mathrm{Im}(Mz) \quad \text{for all} \quad M \;\in\; \langle S, T \rangle.$$

This maximality condition is unchanged if we replace $z$ by $T^{\pm 1}(z) = z \pm 1$, so we can assume $|\mathrm{Re}(z)| \leq 1/2$ which is already one of the two inequalities defining our fundamental domain. For the remaining inequality, the maximality condition in particular says $\mathrm{Im}(z) \geq \mathrm{Im}(Sz) = \mathrm{Im}(z)/|z|^2$ and hence $|z| \geq 1$. Then $z \in \mathscr{F}$.

(1b) Let us now show that $\Gamma = \langle S, T \rangle$ is generated by the two matrices $S, T$ from above. Given $M \in \Gamma$ we pick any point $z \in M\mathscr{F}$. By part (1a) there exists a matrix $N \in \langle S, T \rangle$ such that $N^{-1}z \in \mathscr{F}$. Then

$$\mathscr{F} \cap N^{-1}M\mathscr{F} \;\neq\; \varnothing$$

and so part (2), which we will verify by an independent computation below, says that

$$\pm M \;\in\; \{NT, NT^{-1}, NS, NA_1, NA_2\}.$$

It then only remains to note that all the matrices on the right are in $\langle S, T \rangle$.

(2) If $z \in \mathscr{F}$ then we have $|cz+d| \geq 1$ for all $(c,d) \in \mathbb{Z}^2 \setminus \{(0,0)\}$. Suppose that also

$$M \cdot z \;\in\; \mathscr{F} \quad \text{for some matrix} \quad M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; \Gamma \;=\; Sl_2(\mathbb{Z}),$$

then similarly

$$\frac{1}{|cz+d|} \;=\; \left| \frac{ad-bc}{cz+d} \right| \;=\; \left| -c \cdot \frac{az+b}{cz+d} + a \cdot \frac{cz+d}{cz+d} \right| \;=\; \bigl| -c \cdot Mz + a \bigr| \;\geq\; 1$$

where in the last step we have applied the previous argument to $Mz$ and the integer vector $(-c, a) \neq (0,0)$. In this case, writing $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$ we obtain that

$$(cx+d)^2 + c^2y^2 \;=\; |cz+d|^2 \;=\; 1.$$

But $y \geq \sqrt{3}/2$ for $z \in \mathscr{F}$ and so we get $c, d \in \{0, \pm 1\}$. The same argument applied to the inverse

$$M^{-1} \;=\; \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

shows that $a, c \in \{0, \pm 1\}$. For $c = 0$ we must have $a = d$ and hence $M = \pm T^b$ for some $b \in \mathbb{Z}$, and then the condition $\mathscr{F} \cap M\mathscr{F} \neq \varnothing$ forces $b \in \{0, \pm 1\}$. So it only remains to discuss the case $c \neq 0$. Replacing the matrix $M$ by its negative we can assume $c = 1$. So

$$M \;=\; \begin{pmatrix} a & b \\ 1 & d \end{pmatrix} \;\in\; \Gamma \quad \text{with} \quad a, d \in \{0, \pm 1\} \quad \text{and} \quad b = ad - 1.$$

The case $a = d = 0$ leads to $M = S$, while $ad \in \{0, +1\}$ leads to $M \in \{A_0, A_1\}$. We leave it to the reader to verify that in the remaining case $ad = -1$ the corresponding matrix $M$ satisfies $\mathscr{F} \cap M\mathscr{F} = \varnothing$ and hence does not enter our list. □

Let us now discuss $\mathbb{H}/\Gamma$ as a Riemann surface. We want to find holomorphic functions on it, or equivalently holomorphic functions $f : \mathbb{H} \to \mathbb{C}$ that are invariant under the action of the modular group. Since we have seen that the quotient $\mathbb{H}/\Gamma$ parametrizes elliptic curves, a natural guess for coordinate functions on it are the coefficients of the equation defining the elliptic curve $E_\tau \simeq \mathbb{C}/\Lambda_\tau$ when $\tau \in \mathbb{H}$ varies: The Eisenstein series

$$G_k(\tau) \;=\; \sum_{\substack{(m,n) \in \mathbb{Z}^2 \\ (m,n) \neq (0,0)}} \frac{1}{(m\tau + n)^k}$$

are locally uniformly convergent as a function of $\tau \in \mathbb{H}$ for $k \geq 3$, so we should look at the functions

$$\begin{aligned} g_2(\tau) &= 60 G_4(\tau), \\ g_3(\tau) &= 140 G_6(\tau). \end{aligned}$$

The above approach is a bit too naive since different cubic equations may give rise to isomorphic elliptic curves, and indeed the Eisenstein series are not invariant under the modular group. However, they transform in a very specific way:

Proposition 7.4. *Let $k \geq 3$. Then for any $\tau \in \mathbb{H}$ and $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in Sl_2(\mathbb{Z})$ we have*

$$G_k(M\tau) \;=\; (c\tau + d)^k G_k(\tau).$$

*Moreover we have*

$$\lim_{\mathrm{Im}(\tau) \to \infty} G_k(\tau) \;=\; 2\zeta(k) \;=\; \sum_{n=1}^{\infty} \frac{2}{n^k} \quad \textit{for } k \geq 4 \textit{ even.}$$

*Proof.* This is a straightforward computation. For the first claim, for $m, n \in \mathbb{Z}$ one verifies

$$m \cdot M\tau + n \;=\; \frac{1}{c\tau + d} \cdot (m'\tau + n') \quad \text{where} \quad \begin{pmatrix} m' \\ n' \end{pmatrix} \;=\; M^t \cdot \begin{pmatrix} m \\ n \end{pmatrix}$$

Here $M^t = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right) \in Sl_2(\mathbb{Z})$, and $M^t : \mathbb{Z}^2 \xrightarrow{\sim} \mathbb{Z}^2$ is an isomorphism. For the second claim it suffices by periodicity to discuss the limit for $\tau$ in the vertical strip defined by the condition $|\mathrm{Re}(\tau)| \leq 1/2$. One can show that the Eisenstein converges uniformly on

$$\{\tau \in \mathbb{H} \mid |\mathrm{Re}(\tau)| \leq 1/2, \mathrm{Im}(\tau) \geq 1\},$$

so we may interchange the limit and the summation, and in the resulting series the only surviving terms for $\mathrm{Im}(\tau) \to \infty$ are those where $m = 0$. □

In particular, applying the above to the matrix $M = T$ from proposition 7.3 we get the periodicity

$$G_k(\tau + 1) \; = \; G_k(\tau)$$

for all $\tau \in \mathbb{H}$, thus we can regard the Eisenstein series as holomorphic functions on the Riemann surface $\mathbb{H}/\mathbb{Z} \subset \mathbb{C}/\mathbb{Z}$ which is obtained from the upper half plane by identifying points which differ by an integer. The exponential function induces an isomorphism

$$\exp(2\pi i(-)): \quad \mathbb{H}/\mathbb{Z} \xrightarrow{\sim} D^* \; = \; \{q \in \mathbb{C} \mid 0 < |q| < 1\}$$

from this Riemann surface onto the punctured open unit disk. We are interested in functions that behave nicely when $\mathrm{Im}(\tau) \to \infty$, so we put $D = \{q \in \mathbb{C} \mid |q| < 1\}$ and make the

Definition 7.5. A *meromorphic modular form of weight* $k$ is a meromorphic function $f : \mathbb{H} \longrightarrow \mathbb{P}^1(\mathbb{C})$ such that

- $f(M\tau) = (c\tau + d)^k f(\tau)$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in Sl_2(\mathbb{Z})$, and
- $f \circ \exp^{-1} : D^* \longrightarrow \mathbb{P}^1(\mathbb{C})$ extends to a meromorphic function on $D$.

A *(holomorphic) modular form* is a modular form which is holomorphic on $\mathbb{H}$ and for which the above meromorphic extension has no pole at the origin $0 \in D$.

Note that since $-E \in Sl_2(\mathbb{Z})$, there are no meromorphic modular forms of odd weight other than the zero function. On the other hand, by proposition 7.4 the Eisenstein series $G_k$ is a modular form of weight $k$ for $k \geq 4$. From a given set of modular forms we may construct many new ones:

- If $f, g$ are modular forms of weight $k$, then so is $af + bg$ for all $a, b \in \mathbb{C}$.
- If $f, g$ are modular forms of weights $k, l$ respectively, then
  - the product $fg$ is a modular form of weight $k + l$,
  - the quotien $f/g$ is a meromorphic modular form of weight $k - l$.

Note that in the last case we can lower the weight, but usually the resulting modular form will be meromorphic, with poles in the points where $g$ vanishes. To get a holomorphic function $f : \mathbb{H}/\Gamma \to \mathbb{C}$ we would like to construct a meromorphic modular form of weight zero which has a pole at most in $\tau = i\infty$, i.e. in $q = 0$. We can do so as follows:

Proposition 7.6. *The function* $\Delta(\tau) = g_2(\tau)^3 - 27g_3(\tau)^2$ *is a holomorphic modular form of weight* 12 *which has no zeroes on the upper half plane. Hence the quotient*

$$J(\tau) \; = \; g_2(\tau)^3/\Delta(\tau)$$

*is a meromorphic modular form of weight zero whose only pole is at* $\tau = +\infty$.

*Proof.* Recall from corollary 4.8 that the elliptic curve parametrized by $\tau \in \mathbb{H}$ has the form

$$E_\tau \; = \; \{(x, y) \in \mathbb{C}^2 \mid y^2 \; = \; f(x)\} \cup \{\infty\}$$

where

$$f(x) \; = \; 4(x - e_1)(x - e_2)(x - e_3) \quad \text{with} \quad \begin{cases} e_1 \; = \; \wp(\frac{1}{2}), \\ e_2 \; = \; \wp(\frac{\tau}{2}), \\ e_3 \; = \; \wp(\frac{1+\tau}{2}). \end{cases}$$

Here $\wp$ denotes the Weierstrass function for the lattice $\Lambda_\tau = \mathbb{Z} \oplus \mathbb{Z}\tau$. The argument that we used in the proof of that corollary also shows that the three zeroes $e_1, e_2, e_3$

of the cubic $f(x)$ are pairwise distinct: Indeed, if $e_i = e_j$ for some $i \neq j$, then the function

$$z \;\mapsto\; \wp(z) - e_i \;=\; \wp(z) - e_j$$

is an even elliptic function of degree two with a double zero at two different points of $\mathbb{C}/\Lambda_\tau$ contradicting theorem 4.2. Thus all three roots of the polynomial $f(x)$ are distinct, which can be expressed by saying that the *discriminant* of this polynomial does not vanish:

$$\Delta_f \;:=\; (e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2 \;\neq\; 0.$$

Now recall from algebra that this discriminant is a symmetric function in $e_1, e_2, e_3$ and as such it can be written in terms of the elementary symmetric functions, which are essentially the coefficients of $f(x) \in \mathbb{C}[x]$. Explicitly, in our case corollary 4.9 says that

$$f(x) \;=\; 4(x - e_1)(x - e_2)(x - e_3) \;=\; 4x^3 - g_2 x - g_3$$

so the elementary symmetric functions are

$$\begin{aligned} e_1 + e_2 + e_3 &= 0 \\ e_1e_2 + e_1e_3 + e_2e_3 &= -g_2/4 \\ e_1e_2e_3 &= g_3/4 \end{aligned}$$

and one therefore computes

$$\Delta_f \;=\; (e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2 \;=\; \cdots \;=\; \Delta(\tau).$$

Thus the notion of discriminants in algebra naturally leads to the consideration of the function

$$\Delta(\tau) \;=\; g_2(\tau)^3 - 27 g_3(\tau)^2$$

and shows that this function is nonzero for all $\tau \in \mathbb{H}$. By the remarks preceding this proof it is also a holomorphic modular form of weight six and hence $J(\tau)$ is a meromorphic modular form of weight zero. □

Remark 7.7. From an arithmetic point of view it is better to replace $J(\tau)$ by the function

$$j(\tau) \;=\; 12^3 \cdot J(\tau) \;=\; \frac{1728\, g_2(\tau)^3}{\Delta(\tau)},$$

since the Laurent expansion

$$j(\tau) \;=\; \frac{1}{q} + 744 + 196884q + 21493760q^2 + \cdots$$

in $q = \exp(2\pi i\tau)$ can then be shown to have only integers as coefficients.

Let us now come back to the discussion of the Riemann surface $\mathbb{H}/\Gamma$ that parametrizes isomorphism classes of complex tori. Recall from the description of the identifications on the boundary of the fundamental domain in proposition 7.3 that for $R > 1$ the only identifications between points $\tau$ with $\mathrm{Im}(\tau) > R$ are translations by integers, so

$$U \;=\; \{[\tau] \in \mathbb{C}/\mathbb{Z} \mid \mathrm{Im}(\tau) > R\} \;\hookrightarrow\; \mathbb{H}/\Gamma$$

embeds as an open subset in the quotient. On this open subset we consider the isomorphism

$$\exp(2\pi i(-)) : \quad U \xrightarrow{\sim} D \setminus \{0\} \quad \text{where} \quad D \;=\; \{z \in \mathbb{C} \mid |z| < \exp(-2\pi R)\}$$

Glueing in the entire disk $D$ along this isomorphism we obtain a compact Riemann surface

$$\mathbb{H}/\Gamma \cup \{i\infty\} \;=\; \lim_{\to} \left[ \begin{tikzcd} & D \\ U \arrow[ur] \arrow[dr] & \\ & X \end{tikzcd} \right]$$

where $i\infty$ denotes the point corresponding to the origin $0 \in D$. We can now complete our classification of complex tori as follows:

Theorem 7.8. *We have* $\lim_{\tau\to i\infty} \Delta(\tau) = 0$, $\lim_{\tau\to i\infty} j(\tau) = \infty$, *and the map* $j$ *extends to an isomorphism*

$$j: \quad \mathbb{H}/\Gamma \cup \{i\infty\} \;\xrightarrow{\sim}\; \mathbb{P}^1(\mathbb{C}).$$

*Proof.* The vanishing of the discriminant at $\tau = i\infty$ follows by inserting from proposition 7.4

$$\begin{aligned} g_2(i\infty) &= 120\,\zeta(4) = \tfrac{4}{3}\,\pi^4 \\ g_3(i\infty) &= 280\,\zeta(6) = \tfrac{8}{27}\,\pi^6 \end{aligned}$$

which gives

$$\lim_{\tau\to i\infty} \Delta(\tau) \;=\; \bigl(g_2(i\infty)\bigr)^3 - 27\bigl(g_3(i\infty)\bigr)^2 \;=\; \bigl[(\tfrac{4}{3})^3 - 27\cdot(\tfrac{8}{27})^2\bigr]\,\pi^{12} \;=\; 0.$$

Since $g_2(i\infty) \neq 0$, it then also follows that $\lim_{\tau\to i\infty} j(\tau) = \infty$. Thus we obtain a morphism

$$j: \quad \mathbb{H}/\Gamma \cup \{i\infty\} \;\longrightarrow\; \mathbb{P}^1(\mathbb{C})$$

which is clearly not constant. The image of this morphism is an open subset of the target because holomorphic maps are open, and it is also closed because it is the image of a compact space under a continuous map. Thus $j$ is surjective.

To show that $j$ is an isomorphism we only need to show that it is injective. We already know that the only pole of $j$ is at $\tau = i\infty$. Moreover this is a simple pole because the discriminant $\Delta(\tau)$ has a simple zero at $\tau = i\infty$, as one may deduce from the residue theorem in the form given in theorem 8.2 below. It remains to show that for any $c \in \mathbb{C}$ the function $j$ takes the value $c$ only once, or equivalently that the function $\tau \mapsto j(\tau) - c$ has a unique zero. Since the latter function is still a meromorphic modular form of weight zero with a unique pole at $\tau = i\infty$ and pole order one, this last statement will again follow from theorem 8.2. □

## 8. Appendix: The valence formula

Let us now take a closer look on meromorphic modular forms. Like for elliptic functions, the main point will be to understand their poles and zeroes inside the fundamental domain. A meromorphic modular form of weight zero can be viewed as a meromorphic function on the compact Riemann surface $X = \mathbb{H}/\Gamma \cup \{i\infty\}$, and we can guess how to count its zeroes and poles properly:

Exercise 8.1. Show that each $\tau \in \mathbb{H}$ has an open neighborhood $\tau \in U \subseteq \mathbb{H}$ such that for all $M \in PSl_2(\mathbb{Z})$,

$$U \cap MU \;=\; \begin{cases} U & \text{if } M\tau = \tau, \\ \varnothing & \text{if } M\tau \neq \tau. \end{cases}$$

Deduce that locally on this neighborhood the quotient map $q : \mathbb{H} \to \mathbb{H}/\Gamma$ is given by

$$q : U \twoheadrightarrow U/Stab(\tau) \quad \text{where} \quad Stab(\tau) \;=\; \{M \in PSl_2(\mathbb{Z}) \mid M\tau = \tau\}$$

and that for any meromorphic function $f : X \to \mathbb{P}^1(\mathbb{C})$ the order of vanishing or pole at $x = [\tau] \in X$ is

$$\operatorname{ord}_x(f) \;=\; \frac{\operatorname{ord}_\tau(f)}{e_\tau} \quad \text{where} \quad e_\tau \;=\; |Stab(\tau)| \;=\; \begin{cases} 3 & \text{if } x = [e^{\frac{\pi i}{3}}], \\ 2 & \text{if } x = [\,i\,], \\ 1 & \text{else.} \end{cases}$$

Similarly, recall that every meromorphic function $f : X \to \mathbb{P}^1(\mathbb{C})$ has a Laurent expansion

$$f(\tau) \;=\; \sum_{n \gg 0} c_n(f) q^n \quad \text{with} \quad c_n(f) \in \mathbb{C},\ q = \exp(2\pi i \tau)$$

for $\operatorname{Im}(\tau) \gg 0$ big enough, and then we have $\operatorname{ord}_{i\infty}(f) = \min\{n \in \mathbb{Z} \mid c_n(f) \neq 0\}$.

Rather than using the above observations, we will take them as the definition of the order of vanishing for a meromorphic modular form $f : \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$ of arbitrary weight $k \in \mathbb{Z}$. Note that for $k \neq 0$ such a modular form cannot be regarded as a function on the compact Riemann surface $\mathbb{H}/\Gamma \cup \{\infty\}$ since it is not invariant under the modular group. Nevertheless its order of vanishing or poles at $\tau \in \mathbb{H}$ only depends on the image $x = [\tau] \in \mathbb{H}/\Gamma$ since $c\tau + d \neq 0, \infty$ for all nonzero vectors $(c, d) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$, and it admits a Laurent series expansion around $i\infty$ because it has the periodicity property $f(\tau + 1) = f(\tau)$ for all $\tau \in \mathbb{H}$. Motivated by the above discussion, we therefore *define* the order of vanishing or zeroes of a meromorphic modular form $f : \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$ at a point $x \in \mathbb{H}/\Gamma \cup \{i\infty\}$ by the formula

$$\operatorname{ord}_x(f) \;=\; \begin{cases} \operatorname{ord}_\tau(f)/e_\tau & \text{if } x = [\tau] \text{ with } \tau \in \mathbb{H}, \\ \min\{n \mid c_n \neq 0\} & \text{if } x = i\infty. \end{cases}$$

With these notations we have the following valence formula:

Theorem 8.2. *If $f : \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$ is a meromorphic modular form of weight $k$ which is not identically zero, then $\operatorname{ord}_x(f) = 0$ for almost all $x \in X = \mathbb{H}/\Gamma \cup \{i\infty\}$, and we have*

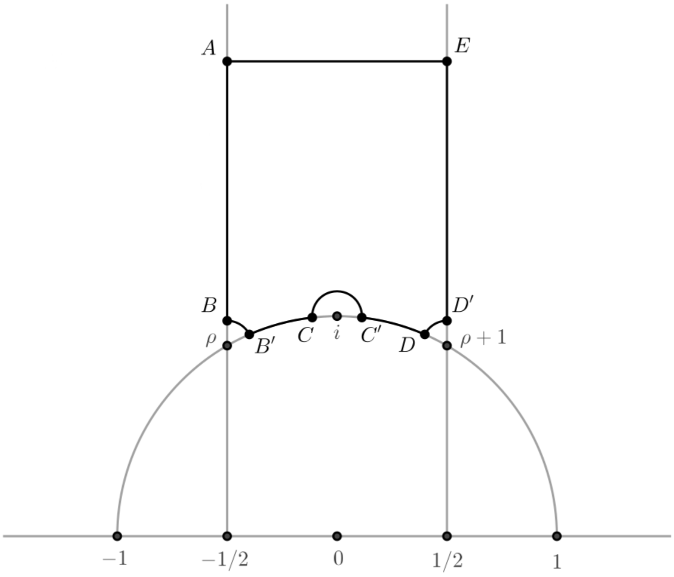
$$\sum_{x \in X} \operatorname{ord}_x(f) \;=\; \frac{k}{12}.$$

*Proof.* Let $\mathscr{F} \subset \mathbb{H}$ be the fundamental domain from proposition 7.3. Since $f$ is a meromorphic function around the point $i\infty$, it can at most have finitely many poles and zeroes in the image of the vertical half-strip $\{\tau \in \mathscr{F} \mid \operatorname{Im}(\tau) > c\}$ for fixed $c \gg 0$. Since $\{\tau \in \mathscr{F} \mid \operatorname{Im}(\tau) \leq c\}$ is a compact set, we hence see that $f$ can have at most finitely many zeroes and poles in the domain $\mathscr{F} \subset \mathbb{H}$.

For the proof of the claimed formula, let us first assume $f$ has no poles or zeroes on the boundary $\partial F$ except perhaps at the three fixed points $\rho = \exp(\pi i/3)$, $\rho + 1$ and $i$. Consider the path

$$\gamma : \quad \Big(A \to B \to B' \to C \to C' \to D \to D' \to E \to A\Big)$$

as shown below:



Around each of the three fixed points we have drawn a circle segment of very small radius $\varepsilon > 0$, while $\mathrm{Im}(A), \mathrm{Im}(E) = N \gg 0$ are supposed to be big. By the residue theorem then

$$\sum_{\tau \in \mathscr{F} \setminus \partial \mathscr{F}} \mathrm{ord}_\tau(f) \;=\; \frac{1}{2\pi i} \int_\gamma \frac{f'(z)}{f(z)} dz.$$

We now compute the integral on the right hand side step by step.

*Step 1.* The contributions from the straight line segments $AB$ and $D'E$ cancel because $f(z+1) = f(z)$ and the two segments have opposite orientations.

*Step 2.* Now take the line $EA$. Near $z = i\infty$ write $f(z)$ and $g(z) = f'(z)/f(z)$ as Fourier series

$$\begin{aligned} f(z) \;&=\; \sum_{n \geq n_0} c_n(f) \cdot e^{2\pi i n z} \\ g(z) \;&=\; \sum_{n \gg 0} c_n(g) \cdot e^{2\pi i n z} \end{aligned}$$

where $n_0 = \mathrm{ord}_{i\infty}(f)$. From the identity $f'(z) = f(z)g(z)$ we then get the Fourier coefficients

$$c_n(g) \;=\; \begin{cases} 0 & \text{for } n < n_0, \\ 2\pi i n_0 & \text{for } n = n_0, \\ \cdots & \text{for } n > n_0. \end{cases}$$

It follows that

$$\begin{aligned} \frac{1}{2\pi i} \int_E^A g(z) dz \;&=\; 2\pi i \mathrm{ord}_{i\infty}(f) + \sum_{n > 0} c_n(g) \cdot \int_E^A e^{2\pi i n z} dz \\ &=\; 2\pi i \mathrm{ord}_{i\infty}(f) \end{aligned}$$

when $N \gg 0$ is chosen big enough so that the Fourier series converges on $\mathrm{Im}(z) > N$.

*Step 3.* Next we discuss the circle segments around the points with a nontrivial stabilizer in the modular group. Write the logarithmic derivative $g(z) = f'(z)/f(z)$ locally near $z = \rho$ as

$$g(z) \;=\; \mathrm{ord}_\rho(f) \cdot (z - \rho)^{-1} + h(z)$$

where $h(z)$ is holomorphic. Pick any branch log of the complex logarithm, then we obtain

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_B^{B'} g(z)dz &= \operatorname{ord}_\rho(f)\cdot\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_B^{B'}\frac{dz}{z-\rho}\\ &= \operatorname{ord}_\rho(f)\cdot\lim_{\varepsilon\to 0}\frac{\log(B'-\rho)-\log(B-\rho)}{2\pi} \;=\; -\frac{\operatorname{ord}_\rho(f)}{6}\end{aligned}$$

and similarly

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_C^{C'} g(z)dz &= -\frac{\operatorname{ord}_i(f)}{2}\\ \lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_D^{D'} g(z)dz &= -\frac{\operatorname{ord}_{\rho+1}(f)}{6} \;=\; -\frac{\operatorname{ord}_\rho(f)}{6}\end{aligned}$$

*Step 4.* It remains to discuss the two circle segments $B'C$ and $C'D$ which are related by the transformation

$$S:\quad \mathbb{H}\to\mathbb{H},\quad z\mapsto -1/z.$$

Modularity gives

$$\begin{aligned} f(Sz) &= z^k f(z)\\ \Longrightarrow\quad f'(Sz) &= z^k f'(z)+kz^{k-1}f(z)\\ \Longrightarrow\quad g(Sz) &= z^2 g(z)+kz\end{aligned}$$

If $\alpha:[0,1]\to\mathbb{C}$ denotes any parametrization of the circle segment $B'C$, then we get

$$\begin{aligned}\int_{B'}^{C} g(z)dz &= \int_0^1 g(\alpha(t))\alpha'(t)dt,\\ \int_{C'}^{D} g(z)dz &= -\int_0^1 g(S\alpha(t))(S\alpha)'(t)dt\\ &= -\int_0^1 g(\alpha(t))\alpha'(t)-k\int_0^1\frac{\alpha'(t)}{\alpha(t)}dt\end{aligned}$$

and so

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_{B'C\cup C'D} g(z)dz &= -\lim_{\varepsilon\to 0}\frac{k}{2\pi i}\int_0^1\frac{\alpha'(t)}{\alpha(t)}t\\ &= -\lim_{\varepsilon\to 0}\frac{k}{2\pi i}\cdot\big[\log C-\log B'\big]\\ &= -\frac{k}{2\pi i}\cdot\big[\log(i)-\log(\rho)\big] \;=\; \frac{k}{12}.\end{aligned}$$

*Step 5.* In conclusion, taking the sum of all above contributions we obtain as claimed that

$$\sum_{z\in\mathscr{F}\backslash\partial F}\operatorname{ord}_z(f) \;=\; \int_\gamma g(z)dz \;=\; \frac{k}{12}-\frac{\operatorname{ord}_\rho(f)}{3}-\frac{\operatorname{ord}_i(f)}{2}-\operatorname{ord}_i(f),$$

provided that there are no other poles or zeroes of $f$ on the boundary $\partial\mathscr{F}$. If there are such poles or zeroes, one should modify the fundamental domain by pairs of corresponding small half-circles avoiding these zeroes and poles. We leave it to the reader to draw a picture and adapt the argument in steps 1 and 4 to this case. □

COROLLARY 8.3. *Any non-constant holomorphic modular form has weight $k \geq 4$ and has at least one zero in $\mathbb{H} \cup \{i\infty\}$. Moreover,*

- *$G_4(\tau)$ has a simple zero at $\tau = \exp(\pi i/3)$ and no other zeroes modulo $\Gamma$.*
- *$G_6(\tau)$ has a simple zero at $\tau = i$ and no other zeroes modulo $\Gamma$.*
- *$\Delta(\tau)$ has a simple zero at $\tau = i\infty$ and no other zeroes modulo $\Gamma$.*

*Proof.* Let $f$ be a non-constant holomorphic modular form of weight $k$. If $f$ had no zeroes on $\mathbb{H} \cup \{i\infty\}$, theorem 8.2 would imply $k = 0$. This would mean that $f$ descends to a non-constant holomorphic function on the compact Riemann surface $X = \mathbb{H}/\Gamma \cup \{i\infty\}$ which is impossible by the maximum principle. So $f$ must have strictly positive order of vanishing in at least one point of $X$. But by definition

$$\mathrm{ord}_x(f) \;\in\; \tfrac{1}{e_x} \cdot \mathbb{N}_0 \quad \text{for all} \quad x \;\in\; X, \quad e_x \;=\; \begin{cases} 3 & \text{if } x = [\exp(\pi i/3)], \\ 2 & \text{if } x = [i], \\ 1 & \text{otherwise.} \end{cases}$$

Hence it follows that $k \geq 4$ by theorem 8.2. The remaining statement similarly follows from

$$G_4(\exp(\pi i/3)) \;=\; G_6(i) \;=\; \Delta(i\infty) \;=\; 0,$$

which is left as an exercise to the reader. □

CHAPTER II

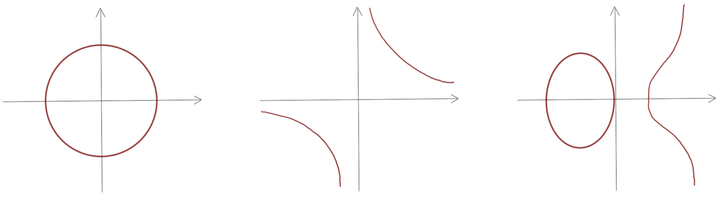
# Geometry of elliptic curves

## 1. Affine and projective varieties

In algebraic geometry we study solutions to systems of polynomial equations in several variables over commutative rings. For instance, if $f \in k[x_1, \dots, x_n]$ is a polynomial of degree $d > 0$ over a field $k$, we consider the vanishing locus

$$V(f)(k) \;=\; \big\{\, a = (a_1, \dots, a_n) \in k^n \mid f(a) = 0 \,\big\}.$$

Intuitively we regard this vanishing locus as a hypersurface of dimension $n-1$ in the $n$-dimensional affine space. For $n = 2$ we get curves in the affine plane, like in the following pictures which show points with coordinates in $k = \mathbb{R}$:



$x_1^2 + x_2^2 - 1 = 0$ $\quad$ $x_1 x_2 - 1 = 0$ $\quad$ $x_2^2 - x_1(x_1^2 - 1) = 0$

However, in general we should also consider solutions with coordinates in extension fields $K \supseteq k$. For instance, the polynomial $f(x_1, x_2) = x_1^2 + x_2^2 + 1$ has $V(f)(\mathbb{R}) = \varnothing$ but in this example $V(f)(\mathbb{C})$ is the union of two complex lines. Back to the general case, we consider for any field extension $K/k$ the affine $n$-space $\mathbb{A}^n(K) = K^n$ and put

$$V(f)(K) \;=\; \big\{\, a = (a_1, \dots, a_n) \in K^n \mid f(a) = 0 \,\big\} \;\subset\; \mathbb{A}^n(K).$$

We have natural inclusions

$$\begin{tikzcd}
V(f)(k) \arrow[r] \arrow[d, phantom, "\cap"] & V(f)(K) \arrow[d, phantom, "\cap"] \\
\mathbb{A}^n(k) \arrow[r] & \mathbb{A}^n(K)
\end{tikzcd}$$

and $V(f)(K) = V(f)(L) \cap \mathbb{A}^2(K)$ for towers of field extensions $k \subset K \subset L$.

Definition 1.1. By the *affine hypersurface* cut out by $f \in k[x_1, \dots, x_n]$ we mean the collection of subsets

$$V(f)(K) \;\subset\; \mathbb{A}^n(K)$$

where $K$ runs over all extension fields of $k$. We also briefly write $V(f) \subset \mathbb{A}^n$ and say that this is an affine hypersurface of degree $d = \deg(f)$. There are some special cases:

- For $d = 1, 2, 3, \dots$ we call $V(f) \subset \mathbb{A}^n$ a *hyperplane, quadric, cubic*, etc.
- For $n = 2$ we call $V(f) = C_f \subset \mathbb{A}^2$ an *affine plane curve* (a *line* if $d = 1$).

It follows from Hilbert's Nullstellensatz that the hypersurface $V(f) \subset \mathbb{A}^n$ is determined uniquely by the subset $V(f)(\overline{k}) \subset \mathbb{A}^n(\overline{k})$ of points in the affine space over the algebraic closure $\overline{k} \supseteq k$. Indeed, given two polynomials $f, g \in k[x_1, \dots, x_n]$ with no multiple factors over the algebraic closure of the base field, one can show that

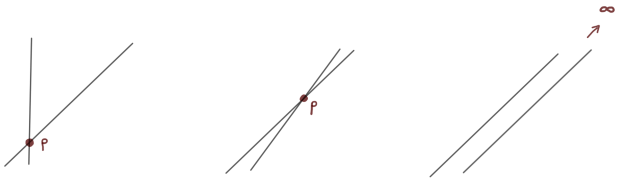
$$V(f)(\overline{k}) = V(g)(\overline{k}) \text{ inside } \mathbb{A}^n(\overline{k}) \quad \Longleftrightarrow \quad g = cf \text{ for some } c \in k \setminus \{0\},$$

so the above definition includes a lot more data than needed. But this is what makes the story interesting from an arithmetic viewpoint: For instance, Fermat's last theorem says that the set of rational points $V(f)(\mathbb{Q})$ on the hypersurface $V(f) \subset \mathbb{A}^3$ cut out by

$$f(x_1, x_2, x_3) \;=\; x_1^e + x_2^e + x_3^e \quad \text{for} \quad e \;\geq\; 3$$

only contains the obvious solutions where one of the three coordinates is zero. The interplay between geometric features over $\overline{k}$ and arithmetic features over $k$ is one of the driving forces behind arithmetic geometry. To understand the arithmetic of elliptic curves we should first discuss their basic geometric properties, for which we now recall a few more notions from algebraic geometry.

Often it is not satisfactory to work in affine space. For instance, we all know that any two distinct lines in the affine plane intersect in a point — except if they are parallel:



The above real picture suggests that any two parallel lines should meet in a point which is at "infinite distance", so it seems reasonable to "compactify" the affine plane by adding a point at infinity for each direction of lines. We can do this as follows. The construction works for affine $n$-space for any $n \in \mathbb{N}$. Consider the embedding

$$\mathbb{A}^n(K) \;\hookrightarrow\; \mathbb{A}^{n+1}(K), \quad (a_1, \dots, a_n) \;\mapsto\; (1, a_1, \dots, a_n).$$

On $\mathbb{A}^{n+1}(K) \setminus \{0\}$ we look at the equivalence relation $\sim$ given by the collinearity of points,

$$(a_0, \dots, a_n) \;\sim\; (b_0, \dots, b_n) \quad \Longleftrightarrow \quad \exists\, c \in K^* : a_i = cb_i \text{ for all } i \in \{0, \dots, n\}.$$

We define the points of the *projective $n$-space* over $K$ to be the set of equivalence classes

$$\mathbb{P}^n(K) \;=\; \big(\mathbb{A}^{n+1}(K) \setminus \{0\}\big)/\sim$$

and denote by

$$[a_0 : \cdots : a_n] \;=\; K^* \cdot (a_0, \dots, a_n) \;\in\; \mathbb{P}^n(K)$$

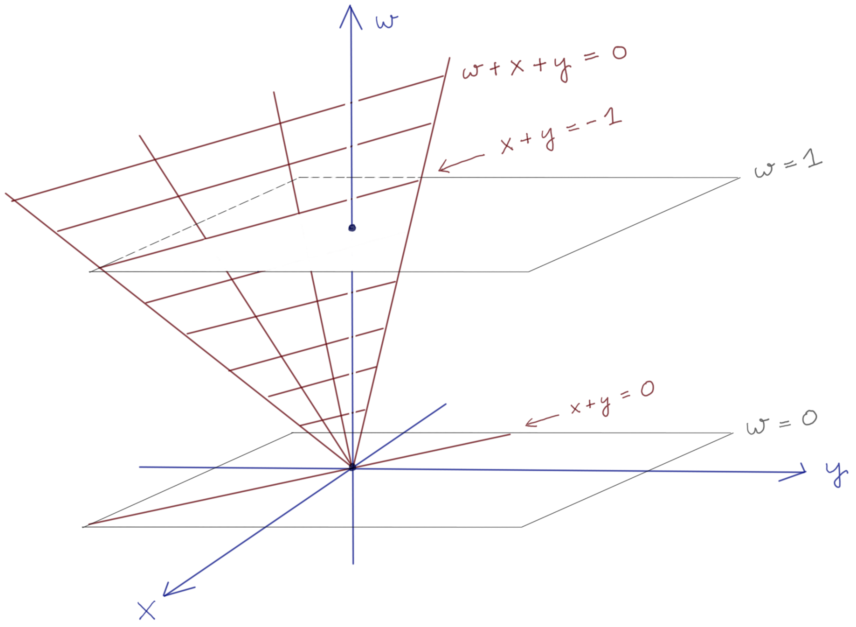
the equivalence class of a point $(a_0, \dots, a_n) \in \mathbb{A}^{n+1}(K) \setminus \{0\}$. We have a natural embedding

$$\mathbb{A}^n(K) \;\hookrightarrow\; \mathbb{P}^n(K), \quad (a_1, \dots, a_n) \;\mapsto\; [1 : a_1 : \cdots : a_n]$$

as shown in the picture below in the case $n = 2$, where we denote the coordinates by $[w : x : y] = [x_0 : x_1 : x_2]$ for simplicity. Note that the projective plane contains for each direction of affine lines in the affine plane $w = 1$ an extra point: The affine line

$$\{(x, y) \mid ax + by = c\} \;\subset\; \mathbb{A}^2(K)$$

with $a, b, c \in k$ corresponds to the "point at infinity" $[-b : a : 0] \in \mathbb{P}^2(k)$. The picture shows the case $(a, b, c) = (1, 1, 1)$:



A similar description holds in higher dimensions. If we replace affine space by projective space, the notion of an affine hypersurface is replaced by

Definition 1.2. Let $f \in k[x_0, \dots, x_n]$ be homogenous of degree $d > 0$. We define the *projective hypersurface* cut out by the polynomial $f$ to be the collection of subsets

$$V(f)(K) \;=\; \left\{ a = [a_0 : \cdots : a_n] \in \mathbb{P}^n(K) \mid f(a) \;=\; 0 \right\} \;\subset\; \mathbb{P}^n(K)$$

where $K$ runs over all extension fields of $k$. We also briefly write $V(f) \subset \mathbb{P}^n$ and say that this is a projective hypersurface of degree $d$. Like in the affine case there are special cases:

- For $d = 1, 2, 3, \dots$ we call $V(f) \subset \mathbb{P}^n$ a *hyperplane, quadric, cubic* etc.
- For $n = 2$ we call $V(f) = C_f \subset \mathbb{P}^2$ a *plane curve* (a *line* if $d = 1$).

This "compactifies" our previous notion of affine hypersurfaces:

Remark 1.3. To pass between projective and affine hypersurfaces, recall the bijection

$$\mathbb{A}^n(K) \xrightarrow{\sim} \mathbb{P}^n(K) \setminus V(x_0)(K), \quad (a_1, \dots, a_n) \;\mapsto\; [1 : a_1 : \cdots : a_n].$$

In fact

$$\mathbb{P}^n \;=\; \bigcup_{i=0}^{n} U_i \quad \text{is covered by the affine charts} \quad U_i \;=\; \mathbb{P}^n \setminus V(x_i) \;\simeq\; \mathbb{A}^n$$

generalizing the description of the Riemann sphere in example I.3.2. Now for any projective hypersurface $V(f) \subset \mathbb{P}^n$ cut out by a homogenous $f \in k[x_0, \dots, x_n]$, we get the affine hypersurface

$$\begin{tikzcd}
V(f) \cap U_0 \arrow[r, "\sim"] \arrow[d, phantom, "\cap"] & V(f^\flat) \arrow[d, phantom, "\cap"] \\
U_0 \arrow[r, "\sim"] & \mathbb{A}^n
\end{tikzcd}$$

cut out by

$$f^\flat(x_1, \dots, x_n) \;=\; f(1, x_1, \dots, x_n) \;\in\; k[x_1, \dots, x_n].$$

We call $f^\flat$ the *dehomogenization* of the polynomial $f$. A similar description holds in the other affine charts. Conversely, given an affine hypersurface $V(g) \subset \mathbb{A}^n$ cut out by a polynomial $g \in k[x_1, \dots, x_n]$ of degree $d = \deg(g)$, we obtain the projective hypersurface

$$V(g^\sharp) \;\subset\; \mathbb{P}^n \quad \text{where} \quad g^\sharp(x_0, \dots, x_n) \;=\; x_0^d \cdot g(\tfrac{x_1}{x_0}, \dots, \tfrac{x_n}{x_0}) \;\in\; k[x_0, \dots, x_n].$$

We call $g^\sharp$ the *homogenization* of $g$. These constructions are mutually inverse in the sense that

$$\begin{aligned}
(g^\sharp)^\flat &\;=\; g, \\
(f^\flat)^\sharp &\;=\; x_0^{-m} \cdot f \quad \text{for} \quad m = \deg(f) - \deg(f^\flat) \geq 0.
\end{aligned}$$

We get a bijection

$$\begin{aligned}
\big\{\text{hypersurfaces in } \mathbb{A}^n\big\} &\;\xrightarrow{\sim}\; \big\{\text{hypersurfaces in } \mathbb{P}^n \text{ not containing } V(x_0)\big\} \\
V(f) &\;\mapsto\; V(f^\sharp)
\end{aligned}$$

Thinking of projective space as a compactification of affine space, we will also refer to $V(x_0) = \mathbb{P}^n \setminus U_0$ as the *hyperplane at infinity* (the *line at infinity* if $n = 2$).

Remark 1.4. It is often convenient to change coordinates on the ambient affine or projective spaces. Starting from the action of $\mathrm{Gl}_{n+1}(k)$ on $\mathbb{A}^{n+1}(k) \setminus \{0\}$, we get an action of

$$\mathrm{PGl}_{n+1}(k) \;=\; \mathrm{Gl}_{n+1}(k)/k^*$$

on $\mathbb{P}^n(k)$. One easily checks that this action is faithful and transitive. We refer to the bijective self-maps of projective space given by the elements of $\mathrm{PGl}_{n+1}(k)$ as *projective linear transformations.* On the standard affine charts from above they correspond to fractional linear transformations.

For the rest of this section we specialize to the case $n = 2$, where an affine or projective hypersurface is also called an (affine or projective) *plane curve.* In this case we write

$$[w : x : y] \;=\; [x_0 : x_1 : x_2].$$

The correspondence between affine and projective lines other than the line at infinity then reads

$$V(ax + by + c) \subset \mathbb{A}^2 \quad \longleftrightarrow \quad V(ax + by + cw) \subset \mathbb{P}^2$$

where $(a, b, c)$ ranges over all triples with $(a, b) \neq (0, 0)$. Note that any pair of projective lines meets in a unique point, which lies on the line at infinity $V(w)$ iff the two lines are parallel. This basic observation will be generalized to an intersection theory for plane curves of higher degree in corollary 3.5.

## 2. Smoothness and tangent lines

We want to define algebraically the notion of smoothness and the tangent to a plane curve at a given point. After a projective linear coordinate transformation we can assume that the point lies in the affine standard chart $\mathbb{A}^2 \subset \mathbb{P}^2$. We therefore consider affine curves $C = C_f \subset \mathbb{A}^2$ where $f \in k[x,y]$. Since we really want to discuss geometric properties of the curve, we usually assume the polynomial $f$ has no multiple factors over the algebraic closure of the base field, but this is not essential and most of what follows holds more generally.

Definition 2.1. The *multiplicity* of the curve $C_f \subset \mathbb{A}^2$ at $p = (x_0, y_0) \in \mathbb{A}^2(\overline{k})$ is defined by

$$m_p(C_f) \;=\; \min\{\, d \in \mathbb{N}_0 \mid f_d \not\equiv 0\,\}$$

where

$$f(x,y) \;=\; \sum_{d\geq 0} f_d(x,y) \quad \text{with} \quad f_d(x,y) \;=\; \sum_{\nu=0}^{d} c_{d,\nu} \cdot (x-x_0)^{\nu}(y-y_0)^{d-\nu}$$

denotes the expansion in homogenous terms around $p$. Notice that $p \in C_f(\overline{k})$ iff $m_p(C_f) > 0$. We say that

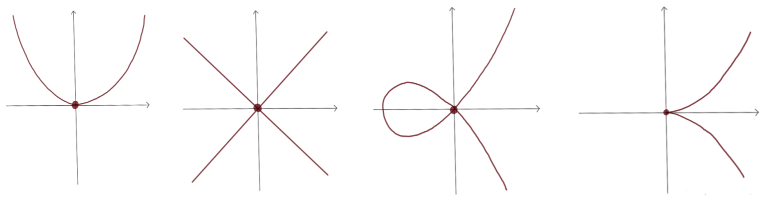
- $p$ is a *smooth point* of $C_f$ if $m_p(C_f) = 1$,
- $p$ is a *singular point* of $C_f$ if $m_p(C_f) > 1$.

We say that $C_f \subset \mathbb{A}^2$ is a *smooth curve* if all points $p \in C_f(\overline{k})$ are smooth. We say that a projective curve is smooth if it is so on each affine chart.

Example 2.2. The following table lists the multiplicities at $p = (0,0) \in \mathbb{A}^2(k)$ for some affine curves:

| $f(x,y)$ | $m_p(C_f)$ |
|---|---|
| $y - x^2$ | 1 |
| $y^2 - x^2$ | 2 |
| $y^2 - x^2(x-1)$ | 2 |
| $y^2 - x^3$ | 2 |

The following picture illustrates the situation for $k = \mathbb{R}$:



$y = x^2$ $\qquad$ $y^2 = x^2$ $\qquad$ $y^2 = x^2(x-1)$ $\qquad$ $y^2 = x^3$

The first example is a smooth point. For $\mathrm{char}(k) \neq 2$, the second and third are both examples of a *node*, also known as an *ordinary double point*: A singularity $p$ where the Hessian matrix

$$\begin{pmatrix} \frac{\partial^2}{\partial x^2} f & \frac{\partial}{\partial x}\frac{\partial}{\partial y} f \\ \frac{\partial}{\partial y}\frac{\partial}{\partial x} f & \frac{\partial^2}{\partial y^2} f \end{pmatrix} (p)$$

has maximal rank. The last example is a type of a singularity known as a *cusp*.

Definition 2.3. Let $p = (x_0, y_0) \in C_f(\overline{k})$ and $m = m_p(C_f) > 0$. Using the local expansion

$$f(x,y) \;=\; \sum_{d \geq m} f_d(x,y) \quad \text{with} \quad f_d(x,y) \;=\; \sum_{\alpha=0}^{d} c_{d,\nu} \cdot (x-x_0)^{\nu}(y-y_0)^{d-\nu}$$

we define the *tangent cone* to the curve $C_f$ at the point $p$ as the vanishing locus of the lowest degree terms

$$T_pC_f \;=\; C_{f_m} \;\subset\; \mathbb{A}^2$$

Writing

$$f_m(x,y) \;=\; \prod_{\nu=1}^{m} \ell_\nu(x,y) \quad \text{with homogenous linear forms} \quad \ell_\nu \;\in\; \overline{k}[x,y].$$

we see that

$$T_pC_f \;=\; \bigcup_{\nu=1}^{m} C_{\ell_\nu}$$

is a union of affine lines through the given point. In the special case of a smooth point this tangent cone is reduced to a single line and is called the *tangent line* to $C_f$ at $p$. We also call the projective closure $\mathbb{P}T_pC_f = C_{f_m} \subset \mathbb{P}^2$ the projective tangent cone resp. line to the curve at the given point $p$.

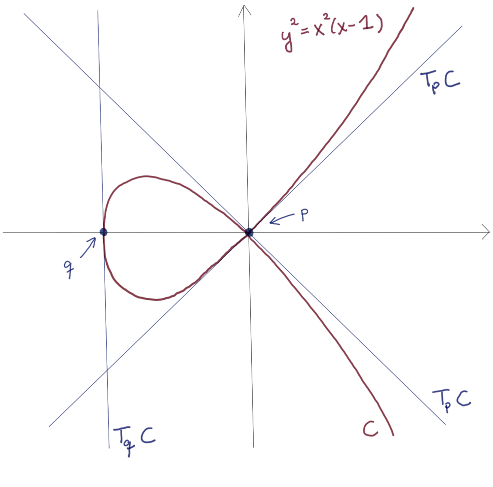
Example 2.4. For $f(x,y) = y^2 - x^2(x+1)$, the tangent cone to $C_f$ at the node $p = (0,0)$ is

$$T_pC_f \;=\; C_{x+y} \cup C_{x-y} \;\subset\; \mathbb{A}^2$$

which is a union of two distinct lines if $\operatorname{char}(k) \neq 2$. On the other hand $q = (-1,0)$ is a smooth point with

$$T_qC_f \;=\; C_{x+1} \;\subset\; \mathbb{A}^2$$

as shown below:



Exercise 2.5. Let $f \in k[w,x,y]$ be a homogenous polynomial. Show that the projective tangent line to the corresponding curve at a smooth point $p \in C_f(k)$ is the orthocomplement of the gradient:

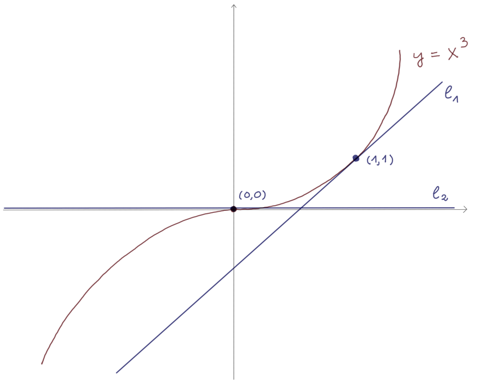
$$\mathbb{P}T_pC_f \;=\; \left\{ p \in \mathbb{P}^2(k) \mid \tfrac{\partial f}{\partial w}(p) + \tfrac{\partial f}{\partial x}(p) + \tfrac{\partial f}{\partial y}(p) \;=\; 0 \right\}$$

Formulate the analogous description in the affine setup.

There are different kinds of tangents. For instance, let $f(x,y) = y - x^3$ and denote by $C = C_f \subset \mathbb{A}^2$ the corresponding affine curve. Then the lines parametrized by

$$\begin{aligned} \ell_1: \quad t &\mapsto (x_1(t), y_1(t)) \;=\; (t+1, 3t+1), \\ \ell_2: \quad t &\mapsto (x_2(t), y_2(t)) \;=\; (t, 0) \end{aligned}$$

are both tangent to the curve at the point parametrized by $t = 0$ as shown below:



Being tangent means that the function $t \mapsto f(x_i(t), y_i(t))$ should vanish to order at least two at $t = 0$. Now we see that the second line is a better tangent than the first one:

- $f(x_1(t), y_1(t)) = -t^2(t+3)$ has a *double* zero at $t = 0$,
- $f(x_2(t), y_2(t)) = -t^3$ has a *triple* zero at $t = 0$.

This leads to the following definition, formulated in projective coordinates:

Definition 2.6. Let $C_f \subset \mathbb{P}^2$ for a homogenous polynomial $f \in k[w,x,y]$, which as usual we assume to have no multiple factors over the algebraic closure of the base field. Let $p = [w_0 : x_0 : y_0] \in \ell(k)$ for a line $\ell \not\subset C_f$. Parametrize part of the line by

$$\varphi: \quad \mathbb{A}^1 \longrightarrow \mathbb{P}^2, \quad t \mapsto [(w_0 + tw_1) : (x_0 + tx_1) : (y_0 + ty_1)]$$

where $q = [w_1 : x_1 : y_1] \in \ell(k) \setminus \{p\}$ denotes any other point on the line. Then the vanishing order

$$i_p(\ell, C_f) \;=\; \operatorname{ord}_{t=0}(\varphi(t))$$

is called the *order of contact* or the *order of tangency* of the line $\ell$ and the curve $C_f$ at the point $p$. Note that $i_p(\ell, C_f) > 0$ iff $p \in (\ell \cap C_f)(k)$. The order of contact is related to the multiplicity of the curve as follows:

Remark 2.7. If $k$ is infinite, then the multiplicity of $C_f \subset \mathbb{P}^2$ at $p \in \mathbb{P}^2(k)$ is given by

$$m_p(C_f) \;=\; \min\bigl\{\, i_p(\ell, C_f) \mid \ell \subset \mathbb{P}^2 \text{ is a line over } k \text{ with } p \in \ell(k) \,\bigr\}.$$

*Proof.* By a projective linear coordinate transformation over the field $k$ we may assume $p = [1:0:0]$. For any given line $\ell \subset \mathbb{P}^2$ over $k$ we may also assume that it contains a point

$$q \;=\; [0 : x_1 : y_1] \;\in\; \ell(k)$$

with $x_1, y_1 \in k$. Then part of the line is parametrized by $t \mapsto \varphi(t) = [1 : tx_1 : ty_2]$, and writing

$$f^\flat(x,y) \;=\; f(1,x,y) \;=\; \sum_{d \geq m} f_d(x,y)$$

with each $f_d(x,y) \in k[x,y]$ homogenous of degree $d$ and with $m = m_p(C_f)$, we obtain that

$$f(\varphi(t)) \;=\; f^\flat(tx_1, ty_1) \;=\; \sum_{d\geq m} f_m(x_1,y_1)\cdot t^m.$$

Hence

$$m_p(C_f) \;\leq\; i_p(\ell, C_f).$$

To see that equality holds for some line $\ell \subset \mathbb{P}^2$ over $k$, note that by our definition of the multiplicity of a curve at a point, the polynomial $f_m \in k[x,y]$ is not the zero polynomial. Since $k$ is an infinite field, we can find a point $(x_1,y_1) \in \mathbb{A}^2(k)\setminus\{(0,0)\}$ where $f_m(x_1,y_1) \neq 0$, and then the line passing through the origin and through this point will do the job. □

In what follows we are mostly interested in the case where $p \in C_f(k)$ is a smooth point. A line $\ell \subset \mathbb{P}^2$ with $i_p(\ell, C_f) = 1$ is said to be *transversal* to the curve $C_p$ at the point $p$. Passing to higher order of contact, the tangent line to the curve at $p$ is the unique line $\ell \subset \mathbb{P}^2$ with $i_p(\ell, C_f) \geq 2$. Points where the tangent has higher order of contact than expected deserve a special name:

Definition 2.8. A point $p \in C_f(k)$ is called a *flex point* of the curve $C_f \subset \mathbb{P}^2$ or simply a *flex* if

- it is a smooth point of $C_f$, and
- the tangent line $\ell = \mathbb{P}T_pC_f$ has $i_p(\ell, C_f) \geq 3$.

It is easy to check that curves of degree at most two have no flex points. As a harder exercise, you may try to verify that over an algebraically closed field $k = \overline{k}$ any smooth plane curve of degree $d \geq 3$ has a flex point.

## 3. Intersection theory for plane curves

In the last section we have discussed intersections of curves with a line, let us now take a look at intersections between curves of arbitrary degrees. For our purpose it will be enough to consider the case where the two curves to be intersected do not have any curve in common, which algebraically means that the two defining equations should not have any common factor. Recall that for $f, g \in k[x,y]$, having no common factor over $k$ is the same as having no common factor over the algebraic closure $\overline{k}$. To take care of points at infinity, we consider as in remark 1.3 the homogenization

$$\begin{aligned} f^\sharp(w,x,y) &= w^{\deg(f)}\cdot f(x/w, y/w),\\ g^\sharp(w,x,y) &= w^{\deg(g)}\cdot g(x/w, y/w).\end{aligned}$$

We say that the two affine plane curves $C_f, C_g \subset \mathbb{A}^2$ *intersect at infinity* if the corresponding projective curves over the algebraic closure of the base field have a common point at infinity, i.e.

$$C_{f^\sharp}(\overline{k}) \cap C_{g^\sharp}(\overline{k}) \cap C_w(\overline{k}) \;\neq\; \varnothing.$$

This is easy to read off: Let $f^*, g^* \in k[x,y]$ be the sum of the terms in $f, g$ of highest degree, i.e.

$$\begin{aligned} f^*(x,y) &= f^\sharp(0,x,y),\\ g^*(x,y) &= g^\sharp(0,x,y).\end{aligned}$$

Then $C_f, C_g \subset \mathbb{A}^2$ intersect at infinity iff $f^*, g^* \in k[x,y]$ have a common factor.

PROPOSITION 3.1. *If $f, g \in k[x,y]$ are polynomials of degree $m, n \geq 0$ without common factors, then*

$$|(C_f \cap C_g)(\overline{k})| \;\leq\; \dim_k k[x,y]/(f,g) \;\leq\; mn.$$

*Moreover, in the second inequality "=" holds unless $C_f$ and $C_g$ intersect at infinity.*

*Proof.* Put $R = k[x,y]$. Since the claims are invariant under passing to the algebraic closure of the base field, we may assume $k = \overline{k}$.

*Step 1.* To show that $d = |(C_f \cap C_g)(k)| \leq \dim_k R/(f,g)$, we use a partition of unity argument similar to the one in the Chinese remainder theorem. We label the points in $(C_f \cap C_g)(k)$ as $p_i = (x_i, y_i)$ for $1 \leq i \leq d$, without repetitions, and we put

$$h_i(x,y) \;=\; \prod_{x_j \neq x_i} (x - x_j) \cdot \prod_{y_j \neq y_i} (y - y_j) \;\in\; R.$$

Then

$$h_i(p_j) \begin{cases} = 0 & \text{for } j \neq i, \\ \neq 0 & \text{for } j = i. \end{cases}$$

Hence if we have

$$\sum_{i=1}^{d} c_i h_i \;\in\; (f,g) \quad \text{for certain } c_i \in k,$$

then evaluating at $p_i$ we get $c_i = 0$ for all $i$. So the images of $h_1, \dots, h_d$ in $R/(f,g)$ are linearly independent over $k$ and therefore $d \leq \dim_k R/(f,g)$.

*Step 2.* Next we show that $\dim_k R/(f,g) \leq mn$. For this we consider for $d \in \mathbb{N}_0$ the subspace

$$R_d \;=\; \Big\{ \sum_{i+j \leq d} c_{ij} x^i y^j \mid c_{ij} \in k \Big\}$$

of polynomials of total degree at most $d$. It suffices to show $\dim_k R_d/(f,g) \leq mn$ for all $d$. We have a diagram

$$\begin{tikzcd}
R_{d-m} \times R_{d-n} \arrow[r, "\alpha_d"] \arrow[d, phantom, "\cap"] & R_d \arrow[r, "\pi_d"] \arrow[d, phantom, "\cap"] & R_d/(f,g) \arrow[r] \arrow[d, phantom, "\cap"] & 0 \\
R \times R \arrow[r, "\alpha = (f,g)"] & R \arrow[r, "\pi"] & R/(f,g) \arrow[r] & 0
\end{tikzcd}$$

where $\pi$ denotes the quotient map, $\alpha(u,v) = uf + vg$, and the top row arises by restriction of the bottom row to the respective subspaces. Note that $\pi_d$ is surjective and $\mathrm{im}(\alpha_d) \subset \ker(\pi_d)$ by exactness of the bottom row. In general the top row will not not be exact, but we can easily compute $\ker(\alpha_d)$: If $(u,v) \in \ker(\alpha_d)$, then by definition $uf = -vg$, and since $R$ is a unique factorization domain where $f$ and $g$ have no common factor, it follows that $g$ divides $u$ and $f$ divides $v$. Therefore we obtain

$$\ker(\alpha_d) \;=\; \big\{ (wg, -wf) \mid w \in R_{d-m-n} \big\} \;\simeq\; R_{d-m-n}.$$

Then

$$\begin{aligned}
\dim_k R/(f,g) \;&=\; \dim_k \mathrm{im}(\pi_d) \\
&\leq\; \dim_k R_d - \dim_k \mathrm{im}(\alpha_d) \\
&=\; \dim_k R_d - \big(\dim_k R_{d-m} + \dim_k R_{d-n} - \dim_k \ker(\alpha_d)\big) \\
&=\; \dim_k R_d - \dim_k R_{d-m} - \dim_k R_{d-n} + \dim_k R_{d-m-n} \\
&=\; \tbinom{d+2}{2} - \tbinom{d-m+2}{2} - \tbinom{d-n+2}{2} + \tbinom{d-m-n}{2} \\
&=\; mn,
\end{aligned}$$

where the last step is a computation with binomial coefficients left to the reader as an exercise. Note that the inequality comes from the inclusion $\operatorname{im}(\alpha_d) \subseteq \ker(\alpha_d)$ and equality holds iff $\operatorname{im}(\alpha_d) = \ker(\pi_d)$.

*Step 3.* It remains to show that if $C_f$ and $C_g$ do not intersect at infinity, then the inclusion $\operatorname{im}(\alpha_d) \subseteq \ker(\pi_d)$ is an equality. Since $\ker(\pi_d) \subset \ker(\pi) = \operatorname{im}(\alpha)$ by the exactness of the bottom row of the diagram from the previous step, we can write any $h \in \ker(\pi_d)$ as

$$h \;=\; uf + vg \quad \text{for certain} \quad u, v \;\in\; R.$$

If $\deg(u) > d - m$, then the terms of highest degree in the above identity must cancel, so we get

$$u^* f^* + v^* g^* \;=\; 0.$$

But $C_f$ and $C_g$ do not intersect at infinity, so $f^*$ and $g^*$ have no common factor, and it follows that

$$\begin{aligned} u^* &= wg^* \\ v^* &= -wf^* \end{aligned}$$

for some $w \in R$. Then

$$h \;=\; (u - wg)f + (v + wf)g \quad \text{where} \quad \begin{cases} \deg(u - wg) < \deg(u), \\ \deg(v + wf) < \deg(v), \end{cases}$$

and by induction we reduce to $(u, v) \in R_{d-m} \times R_{d-n}$ so that $h \in \operatorname{im}(\alpha_d)$. □

We now want to count the points in $(C_f \cap C_g)(\overline{k})$ with the right multiplicities so that also the first inequality in the above proposition becomes an equality. The idea is to split up $\dim_k k[x,y]/(f,g)$ into local contributions:

Definition 3.2. (a) The *local ring of the affine plane at* $p = (x_0, y_0) \in \mathbb{A}^2(\overline{k})$ is defined by

$$\mathscr{O}_{\mathbb{A}^2,p} \;=\; \Big\{ \tfrac{u}{v} \in \overline{k}[x,y] \,\Big|\, u, v \in \overline{k}[x,y],\, v(p) \neq 0 \Big\},$$

i.e. it is the localization of $\overline{k}[x,y]$ at the maximal ideal $(x - x_0, y - y_0)$. We denote by

$$\mathfrak{m}_p \;=\; \Big\{ \tfrac{u}{v} \in \overline{k}[x,y] \,\Big|\, u(p) = 0 \neq v(p) \Big\} \;\trianglelefteq\; \mathscr{O}_{\mathbb{A}^2,p}$$

the unique maximal ideal of this local ring. Its residue field is $\mathscr{O}_{\mathbb{A}^2,p}/\mathfrak{m}_p = k$.

(b) If $f, g \in k[x,y]$ have no common factor, then for $p \in \mathbb{A}^2(\overline{k})$ we define the multiplicity

$$i_p(f,g) \;=\; \dim_{\overline{k}} \mathscr{O}_{\mathbb{A}^2,p}/(f,g)_p \quad \text{for the ideal} \quad (f,g)_p \;\trianglelefteq\; \mathscr{O}_{\mathbb{A}^2,p}.$$

If $f, g$ have no multiple factors over the algebraic closure of the base field, then up to multiplication by a nonzero constant they are determined uniquely by the curves $C_f$, $C_g$. In this case we will also use the notation $i_p(C_f, C_g) = i_p(f,g)$ and call this number the *intersection multiplicity* of the curves at the given point.

Example 3.3. (a) For $p \in \mathbb{A}^2(\overline{k})$, we have $i_p(f,g) > 0$ iff $p \in (C_f \cap C_g)(\overline{k})$. We say that

- $C_f$ and $C_g$ meet *transversely* at $p$ if $i_p(f,g) = 1$.
- $C_f$ and $C_g$ are *tangent* at $p$ if $i_p(f,g) > 1$.

If $C_f \subset \mathbb{P}^2$ is a line, the definitions easily imply that $i_p(C_f, C_g)$ coincides with the order of tangency that we defined in the previous section. For instance, at $p = (0,0)$ the curves cut out by $f(x,y) = y$ and $g(x,y) = y - x^n$ have the intersection multiplicity

$$i_p(f,g) \;=\; n$$

which is a different way of saying that $p$ is a flex of the curve $C_g$ for $n > 1$.

(b) For $f(x,y) = y$ and $g(x,y) = y^2 - x(x^2-1)$, the Chinese remainder theorem implies

$$\begin{aligned} k[x,y]/(f,g) \;&\simeq\; k[x]/(x(x^2-1)) \\ &\simeq\; k[x]/(x) \oplus k[x]/(x-1) \oplus k[x]/(x+1) \end{aligned}$$

and $i_p(f,g) = 1$ at each of the three points $p = (0,0), (1,0), (-1,0) \in (C_f \cap C_g)(\overline{k})$.

The above example indicates how to split up the global intersection number into local contributions in general:

Theorem 3.4. *For $f, g \in k[x,y]$ without common factors, we have a natural isomorphism*

$$\varphi: \quad \overline{k}[x,y]/(f,g) \;\xrightarrow{\sim}\; \prod_p \mathscr{O}_{\mathbb{A}^2,p}/(f,g)_p.$$

*of $k[x,y]$-algebras, where the product on the right ranges over all $p \in (C_f \cap C_g)(\overline{k})$.*

*Proof.* We may assume $k = \overline{k}$. The map in question is the product of the localization maps

$$\varphi_p: \quad k[x,y]/(f,g) \;\twoheadrightarrow\; \mathscr{O}_{\mathbb{A}^2,p}/(f,g)_p.$$

One easily sees that each of these localization maps is surjective. To see that their product $\varphi$ is also surjective, we use a partition of unity argument like in the Chinese remainder theorem: Pick for each point $p \in (C_f \cap C_g)(k)$ a polynomial $h_p \in k[x,y]$ with

$$h_p(q) \;=\; \begin{cases} 1 & \text{for } q = p, \\ 0 & \text{for } q \in (C_f \cap C_g)(k) \setminus \{p\}. \end{cases}$$

Let $N \in \mathbb{N}$ with $N > \dim_k \mathscr{O}_{\mathbb{A}^2,q}/(f,g)_q$ for all $q \in (C_f \cap C_g)(k) \setminus \{p\}$. Looking at terms of smallest degrees in the expansion of polynomials around these points, one deduces

$$h_p^N \;\in\; (f,g)_q \quad \text{for all} \quad q \;\in\; (C_f \cap C_g)(k) \setminus \{p\}.$$

On the other hand, by the surjectivity of the localization map $\varphi_p$ we may find a polynomial $F \in k[x,y]$ such that $\varphi_p(F) = 1/h_p^N \in \mathscr{O}_{\mathbb{A}^2,p}/(f,g)_p$, and it then follows that

$$\varphi(F) \;=\; (0,\dots,0,1,0,\dots,0) \;\in\; \prod_p \mathscr{O}_{\mathbb{A}^2,q}/(f,g)_q$$

is the standard basis vector in component of the product labelled by $p$.

So far we have not used that the base field is algebraically closed. However, this assumption is required for the injectivity of $\varphi$, which we will deduce from Hilbert's Nullstellensatz: Given $b \in \ker(\varphi)$, we need to show that $b \in (f,g)$. For this we consider the ideal

$$\mathfrak{a} \;=\; \{a \in k[x,y] \mid ab \in (f,g)\} \;\trianglelefteq\; k[x,y].$$

If $\mathfrak{a} \neq (1)$ is not the unit ideal, then Hilbert's Nullstellensatz over $k = \overline{k}$ says that there exists a point $p \in \mathbb{A}^2(k)$ with

$$a(p) \;=\; 0 \quad \text{for all} \quad a \;\in\; \mathfrak{a}.$$

In particular $f(p) = g(p) = 0$ and so $p \in (C_f \cap C_g)(k)$. Our assumption $b \in \ker(\varphi)$ implies that

$$\varphi_p(b) \;=\; 0 \;\in\; \mathscr{O}_{\mathbb{A}^2,p}/(f,g)_p.$$

By definition of the localization map $\varphi_p$ we then find polynomials $a, u, v \in k[x,y]$ with $a(p) \neq 0$ and

$$b \;=\; \frac{uf + vg}{a} \quad \text{in} \quad \mathscr{O}_{\mathbb{A}^2,p}.$$

But then $ab \in (f,g) \trianglelefteq k[x,y]$ and so $a \in \mathfrak{a}$ which gives the contradiction $h(p) = 0$. □

Geometrically this means the following. Recall that with our naive set-theoretic notion of a plane curve one has $C_f = C_{f^2}$ for any homogenous $f \in k[w,x,y]$. By the *degree* $d = \deg(C)$ of a plane curve $C \subset \mathbb{P}^2$ we mean the smallest degree of a homogenous polynomial $f[w,x,y]$ such that we have $C = C_f$. Similarly, we say that two plane curves have *no common component* if they can be cut out by two polynomials without common factors. We then have

Corollary 3.5 (Bézout's theorem). *If $C_1, C_2 \subset \mathbb{P}^2$ are curves with no common component, then*

$$\sum_{p \in (C_1 \cap C_2)(\overline{k})} i_p(C_1, C_2) \;=\; \deg(C_1)\deg(C_2).$$

*Proof.* Without loss of generality $k = \overline{k}$. Applying a projective linear coordinate transformation we may furthermore assume that $C_1$ and $C_2$ do not intersect at infinity. Then the claim follows from proposition 3.1 and theorem 3.4. □

Recall that to pass back from projective to affine curves, we have associated in remark 1.3 to any homogenous polynomial $f(w,x,y) \in k[w,x,y]$ the inhomogenous polynomial

$$f^\flat(x,y) \;=\; f(1,x,y) \in k[x,y].$$

The degree of the latter may be smaller than the one of the former, but we have:

Corollary 3.6. *Let $f, h \in k[w,x,y]$ be homogenous. If the curves $C_f, C_h \subset \mathbb{P}^2$ have no common component and do not intersect at infinity, then the following are equivalent for $g \in k[w,x,y]$ homogenous:*

(1) *$g^\flat \in (f^\flat, h^\flat)_p \trianglelefteq \mathscr{O}_{\mathbb{P}^2,p}$ for all $p \in (C_f \cap C_h)(\overline{k})$.*

(2) *There exist homogenous $a, b \in k[w,x,y]$ such that*

$$g \;=\; af + bh \quad \textit{and} \quad \begin{cases} \deg(a) \;=\; \deg(g) - \deg(f), \\ \deg(b) \;=\; \deg(g) - \deg(h). \end{cases}$$

*Proof.* Condition (2) obviously implies (1). For the converse, if (1) holds, then the isomorphism

$$k[x,y]/(f^\flat, h^\flat) \;\xrightarrow{\sim}\; \prod_p \mathscr{O}_{\mathbb{A}^2,p}/(f^\flat, h^\flat)_p$$

from theorem 3.4 implies that $g^\flat \in (f^\flat, h^\flat) \trianglelefteq k[x,y]$. Homogenizing again we find that

$$w^r g(w,x,y) \;\in\; (f,h) \;\trianglelefteq\; k[w,x,y] \quad \text{for some } r \in \mathbb{N}_0.$$

Expanding polynomials in terms of the variable $w$ and using that $C_f$ and $C_h$ do not intersect at infinity, one then reduces to the case where $r = 0$. □

Corollary 3.7. *Let $f, g, h \in k[w,x,y]$ be homogenous polynomials with no common factor, and assume all points of the intersection $(C_f \cap C_h)(\overline{k})$ are smooth on $C_f$. If $C_f \cdot (C_g - C_h) \geq 0$, then there is a curve $C \subset \mathbb{P}^2$ of degree $\deg(g) - \deg(h)$ such that*

$$C_f \cdot (C_g - C_h) \;=\; C_f \cdot C.$$

*Proof.* Up to a projective linear change of coordinates we may assume that the curves $C_f, C_h \subset \mathbb{P}^2$ do not intersect at infinity. Then one may check as an exercise that if $p \in (C_f \cap C_h)(\overline{k})$ is a smooth point on the curve $C_f$ and $i_p(f,g) \geq i_p(f,h)$, then

$$g^\flat \;\in\; (f^\flat, h^\flat)_p \;\trianglelefteq\; \mathscr{O}_{\mathbb{A}^2,p}.$$

Then corollary 3.6 says that there exist homogenous polynomials $a, b \in k[w,x,y]$ such that

$$g \;=\; af + bh \quad \text{and} \quad \begin{cases} \deg(a) \;=\; \deg(g) - \deg(f), \\ \deg(b) \;=\; \deg(g) - \deg(h). \end{cases}$$

Then

$$\begin{aligned} C_f \cdot C_h \;&=\; C_f \cdot C_{af+bh} \\ &=\; C_f \cdot C_{bh} \\ &=\; C_f \cdot C_b + C_h \end{aligned}$$

and hence $C = C_b$ does the job. □

EXAMPLE 3.8. Smoothness is essential for the above. The curve $C_f \subset \mathbb{P}^2$ cut out by the cubic

$$f(w,x,y) = y^2 w - x^3$$

has a cusp at $p = [1:0:0]$. If we denote by $q = [0:0:1]$ the point at infinity, one computes

$$\begin{aligned} C_f \cdot C_g \;&=\; 4[p] + 2[q] \qquad && \text{for } g(w,x,y) = x^2 \\ C_f \cdot C_h \;&=\; 3[p] && \text{for } h(w,x,y) = y \end{aligned}$$

and hence

$$C_f \cdot (C_g - C_h) \;=\; [p] + 2[q] \;\geq\; 0.$$

Here $\deg(g) - \deg(h) = 1$, but the only line passing through $p, q$ is $C = C_x$ and we have

$$C_f \cdot C \;=\; 2[p] + [q] \;\neq\; C_f \cdot (C_g - C_h).$$

Thus corollary 3.7 may fail in the presence of singular points. In the next section we will apply it in the following setup:

COROLLARY 3.9. *Let $E \subset \mathbb{P}^2$ be a smooth cubic and $C_1, C_2 \subset \mathbb{P}^2$ two other cubics, not necessarily smooth. If*

$$\begin{aligned} E \cdot C_1 \;&=\; [p_1] + \cdots + [p_8] + [p] \\ E \cdot C_2 \;&=\; [p_1] + \cdots + [p_8] + [q] \end{aligned}$$

*for some $p_1, \dots, p_8, p, q \in E(\overline{k})$ (not necessarily distinct), then $p = q$.*

*Proof.* Working over the algebraic closure $\overline{k}$, pick a line $\ell \subset \mathbb{P}^2$ through $p$ not tangent to $E$. Then

$$E \cdot \ell \;=\; [p] + [r] + [s]$$

with $p, r, s \in E(\overline{k})$ pairwise distinct. So

$$\begin{aligned} E \cdot (\ell \cup C_2) \;&=\; [p] + [r] + [s] + [p_1] + \cdots + [p_8] + [q] \\ &=\; [q] + [r] + [s] + E \cdot C_1 \end{aligned}$$

and by corollary 3.7 there exists a line $\ell'$ with $E \cdot \ell' = [q] + [r] + [s]$. Since both lines $\ell$ and $\ell'$ pass through the two distinct points $r \neq s$, we must have $\ell' = \ell$ and therefore

$$[q] + [r] + [s] \;=\; E \cdot \ell' \;=\; E \cdot \ell \;=\; [p] + [r] + [s],$$

which implies $p = q$ as claimed. □

## 4. The group law on elliptic curves

The intersection theory developed above will allow to extend the group law from theorem I.5.1 to elliptic curves over arbitrary fields $k$. We begin with

DEFINITION 4.1. An *elliptic curve* over a field $k$ is a pair $(E, o)$ where $E \subset \mathbb{P}^2$ is a smooth cubic over $k$ and $o \in E(k)$ is a chosen point.

The basic example is motivated by the differential equation of the Weierstrass function over the complex numbers:

LEMMA 4.2. *For any cubic polynomial $g(x) \in k[x]$ without multiple roots in $\overline{k}$ we have the elliptic curve $(E, o)$, where*

- *$E = C_f \subset \mathbb{P}^2$ is the cubic defined by $f(1, x, y) = y^2 - g(x)$,*
- *$o = [0 : 0 : 1] \in E(k)$ is the point at infinity, which is a flex point.*

*Proof.* The main point is to check the smoothness of $E$. On the chart $\mathbb{A}^2 \subset \mathbb{P}^2$ with coordinates $(x, y) = [1 : x : y]$, this follows from the fact that the system of equations

$$f(1, x_0, y_0) \;=\; \tfrac{\partial}{\partial x} f(1, x_0, y_0) \;=\; \tfrac{\partial}{\partial y} f(1, x_0, y_0) \;=\; 0$$

has no solution $(x_0, y_0) \in \mathbb{A}^2(\overline{k})$ since otherwise we would have $g(x_0) = g'(x_0) = 0$ contradicting our assumption that the polynomial $g(x)$ has no double root in the algebraic closure of the base field. It remains to check that the cubic $E \subset \mathbb{P}^2$ has no singular point at infinity, i.e. of the form $[0 : x_0 : y_0]$. In homogenous coordinates $E$ is the zero locus of

$$\begin{aligned} f(w, x, y) \;&=\; y^2 w - w^3 g(x/w) \\ &=\; y^2 w - a x^3 - b w x^2 - c w^2 x - d w^3 \end{aligned}$$

for some $a \in k^*$ and $b, c, d \in k$, hence putting $w = 0$ one sees that the only point at infinity is

$$0 \;=\; [0 : 0 : 1] \;\in\; E(k).$$

This is the origin in the chart with affine coordinates $(w, x) = [w : x : 1]$ where $E$ is given by

$$f(w, x, 1) \;=\; w - a x^3 - b w x^2 - c w^2 x - d w^3 \;=\; 0,$$

and it is a smooth point because $\frac{\partial}{\partial w} f(0, 0, 1) \neq 0$. To check that $o$ is a flex point, note that the tangent line to $E$ at this point is parametrized in the above chart by

$$t \;\mapsto\; (w(t), x(t)) \;=\; (0, t),$$

and $f(w(t), x(t), 1) = -a x^3$ has a triple zero at $t = 0$ as required. □

In the above example the point $o \in E(k)$ is a flex. In the general definition of an elliptic curve we did not assume this, but we will see in section 6 that any elliptic curve can be transformed in the above form by changing the embedding $E \subset \mathbb{P}^2$ suitably. But let us first discuss the group structure on elliptic curves, which does not require the normal form from the above lemma.

REMARK 4.3. Let $E \subset \mathbb{P}^2$ be a smooth cubic and $\ell \subset \mathbb{P}^2$ a line over $k$. Then precisely one of the following cases occurs:

(1) $(\ell \cap E)(\overline{k}) = \{p_1, p_2, p_3\}$, with $\ell$ meeting $E$ transversely at $p_1, p_2, p_3$.

(2) $(\ell \cap E)(\overline{k}) = \{p_1, p_2\}$, with $\ell$ transverse to $E$ at $p_1$ but tangent at $p_2$.

(3) $(\ell \cap E)(\overline{k}) = \{p_1\}$ consists of a single point, which is a flex point to $E$.

Counting with multiplicities, we always have $\ell \cdot E = [p_1] + [p_2] + [p_3]$ where $p_3 = p_2$ in the second and $p_3 = p_2 = p_1$ in the third case. Note that if two of the three points are defined over a given field $K \supseteq k$ then so is the third. We thus obtain a map $\cdot : E(K) \times E(K) \to E(K)$ by imposing $p_i \cdot p_j = p_k$ for $\{i,j,k\} = \{1,2,3\}$ in each of the above three situations.

THEOREM 4.4. *Let $(E,o)$ be an elliptic curve over $k$. Then for any field $K/k$ the composition law*

$$+ : \quad E(K) \times E(K) \longrightarrow E(K), \quad p+q \;=\; o \cdot (p \cdot q)$$

*makes $E(K)$ into an abelian group whose neutral element is the point $o \in E(K)$.*

*Proof.* It is obvious from the definitions that $+$ is commutative and $o + p = p$ for all $p \in E(K)$. Furthermore, if we define the inverse by $-p = p \cdot (o \cdot o)$, then one has

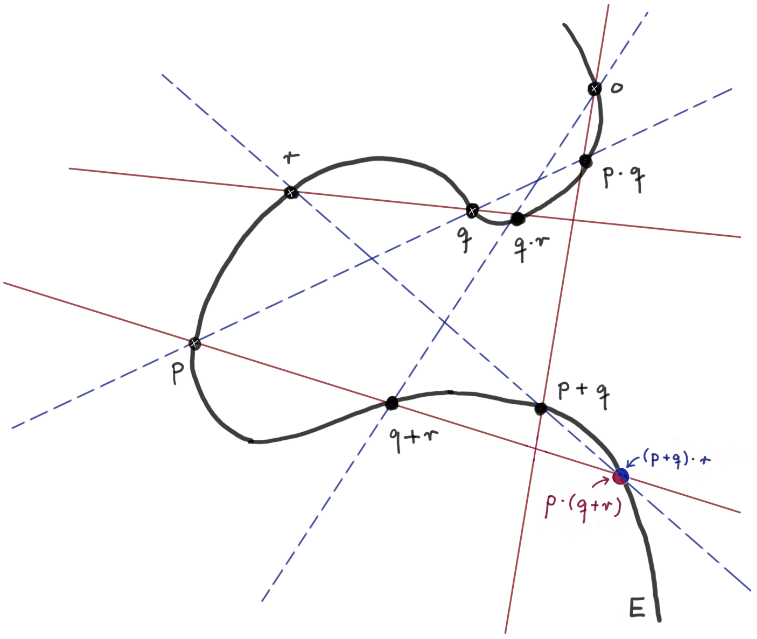
$$p + (-p) \;=\; p + p \cdot (o \cdot o) \;=\; o \cdot (p \cdot (p \cdot (o \cdot o)) \;=\; o \cdot (o \cdot o) \;=\; o$$

by unravelling the definitions. So it only remains to check associativity, for which we must show

$$p \cdot (q + r) \;=\; (p + q) \cdot r$$

for all $p, q, r \in E(K)$. For this we consider the the following picture which involves three cubics:

- the smooth cubic $E$ which is drawn in black,
- the union $C_1 = \langle q, r\rangle \cup \langle p, q+r\rangle \cup \langle o, p \cdot q\rangle$ of the three solid red lines,
- the union $C_2 = \langle p, q\rangle \cup \langle p+q, r\rangle \cup \langle o, q \cdot r\rangle$ of the three dashed blue lines.



By construction we have

$$\begin{aligned} E \cdot C_1 &= [o] + [p] + [q] + [r] + [p \cdot q] + [q \cdot r] + [p+q] + [p+r] + [p \cdot (q+r)], \\ E \cdot C_2 &= [o] + [p] + [q] + [r] + [p \cdot q] + [q \cdot r] + [p+q] + [p+r] + [(p+q) \cdot r]. \end{aligned}$$

These two intersections agree in eight of the nine points. Hence by corollary 3.9 the ninth point also agrees, so $p \cdot (q + r) = (p + q) \cdot r$ as required. □

## 5. Abel's theorem and Riemann-Roch

Over $k = \mathbb{C}$ the theory of elliptic functions has given for any lattice $\Lambda \subset \mathbb{C}$ an isomorphism

$$\begin{aligned} \mathbb{C}/\Lambda \;&\xrightarrow{\sim}\; E \;\subset\; \mathbb{P}^2 \\ z \mod \Lambda \;&\mapsto\; [w:x:y] \;=\; [1:\wp(z):\wp'(z)]. \end{aligned}$$

Moreover, we have seen that any meromorphic function on the complex torus $\mathbb{C}/\Lambda$ is a rational function in the Weierstrass function and its derivative, which in terms of the above isomorphism become

$$\begin{aligned} \wp(z) \;&=\; x/w, \\ \wp'(z) \;&=\; y/w. \end{aligned}$$

On the right hand side we have functions on $E \subset \mathbb{P}^2$ which arise as the restriction of rational homogenous function of degree zero in the variables $x, y, z$. This explains what should be the right algebraic replacement for the notion of "meromorphic functions" on curves over an arbitrary field $k$:

Definition 5.1. Consider a plane curve $C = C_f \subset \mathbb{P}^2$ cut out by an irreducible homogenous polynomial $f \in k[w,x,y]$. After a projective linear coordinate change we may assume that the curve intersects the standard affine chart. Let $f^\flat \in k[x,y]$ be the dehomogenization as in remark 1.3, then $k[x,y]/(f^\flat)$ is an integral domain whose quotient field

$$k(C) \;=\; \operatorname{Quot}\big(k[x,y]/(f^\flat)\big)$$

is called the *function field* of the curve. Its elements are called *rational functions* on the curve. It is a simple exercise to check that in terms of homogenous polynomials the function field can be written as the residue field $k(C) = \mathscr{O}_{\mathbb{P}^2,C}/(f)$ of the local ring

$$\mathscr{O}_{\mathbb{P}^2,C} = \big\{ g/h \in k(w,x,y) \;\mid\; g,h \in k[w,x,y] \text{ homogenous}, \deg(g) = \deg(h), f \nmid h \big\},$$

and we write

$$(g/h)|_C \;=\; (g/h \mod (f)) \;\in\; k(C)$$

for the rational function which is the residue class of an element $g/h \in \mathscr{O}_{\mathbb{A}^2,C}$.

Definition 5.2. Let $C \subset \mathbb{P}^2$ be a smooth curve. If $F = (g/h)|_C \in k(C)^*$ is a rational function other than the zero function, we define its *order of zeroes or poles* at $p \in C(k)$ by

$$\operatorname{ord}_p(F) \;=\; i_p(f,g) - i_p(f,h).$$

One easily checks that this is independent of the chosen representative $g/h$. The first step in the proof of corollary 3.7 shows that the following three properties are equivalent:

- We have $\operatorname{ord}_p(F) > 0$.
- For any $g, h$ with $F = (g/h)|_C$ we have $g^\flat \in (f^\flat, h^\flat)_p \trianglelefteq \mathscr{O}_{\mathbb{A}^2,p}$.
- There exist $\tilde{g}, \tilde{h}$ with $F = (\tilde{g}/\tilde{h})|_C$ such that moreover $\tilde{h}(p) \neq 0$.

If these properties hold, we say that $F$ is *defined at* $p$ and put $F(p) = \tilde{g}(p)/\tilde{h}(p) \in k$.

The above notion of the order of vanishing or poles of rational functions is an algebraic version of the analogous notion for meromorphic functions:

Example 5.3. (a) Over the complex numbers we have seen in theorem I.4.2 that the sum of all pole and zero orders of a meromorphic function vanishes. In the present algebraic setting the analogue is provided by Bézout's theorem 3.4 which gives

$$\sum_{p \in C_f(\overline{k})} \mathrm{ord}_p((g/h)|_{C_f}) \;=\; \deg(f) \cdot (\deg(g) - \deg(h)) \;=\; 0.$$

(b) Let $f(1,x,y) = y^2 - \varphi(x)$ for a cubic polynomial $\varphi(x) = x^3 + ax + b \in k[x]$ without multiple roots. Then the definitions imply that for the point $p = [0,0,1]$ one has

$$\begin{aligned} i_p(f,x) &= 1 && \text{since } k[w,x]/(f(w,x,1),x) \simeq k, \\ i_p(f,y) &= 0 && \text{since } y \neq 0 \text{ near the point } p \in C_f(k), \\ i_p(f,w) &= 3 && \text{since } k[w,x]/(f(w,x,1),w) \simeq k[x]/(x^3). \end{aligned}$$

Hence

$$\mathrm{ord}_p\big((x/w)|_{C_f}\big) \;=\; -2 \quad \text{and} \quad \mathrm{ord}_p\big((y/w)|_{C_f}\big) \;=\; -3$$

as expected from the pole orders of the Weierstrass function and its derivative.

Definition 5.4. A *divisor* on a smooth plane curve $C \subset \mathbb{P}^2$ is a finite formal sum

$$D = \sum_{p \in C(\overline{k})} n_p \cdot [p]$$

of points over the algebraic closure with coefficients $n_p = n_p(D) \in \mathbb{Z}$. Let $\mathrm{Div}(C)$ denote the abelian group of divisors. The *degree* of a divisor is its image under the homomorphism

$$\deg : \quad \mathrm{Div}(C) \;\longrightarrow\; \mathbb{Z}, \quad D \;\mapsto\; \sum_p n_p(D).$$

Put $\mathrm{Div}^0(C) = \ker(\deg)$. Sending a rational function to its divisor of zeroes gives a homomorphism

$$\mathrm{div} : \quad k(C)^* \;\longrightarrow\; \mathrm{Div}^0(C), \quad F \;\mapsto\; \sum_p \mathrm{ord}_p(F) \cdot [p]$$

with image is the subgroup of *principal divisors* denoted by $\mathrm{PDiv}(C)$. We denote the quotient by

$$\mathrm{Pic}^0(C) \;=\; \frac{\mathrm{Div}^0(C)}{\mathrm{PDiv}(C)}$$

and to keep track of the base field we write

$$\begin{aligned} \mathrm{Div}^0(C)(k) &= \{D \in \mathrm{Div}^0(C) \mid n_p(D) = 0 \; \forall p \in C(\overline{k}) \setminus C(k)\}, \\ \mathrm{PDiv}(C)(k) &= \mathrm{PDiv}(C) \cap \mathrm{Div}^0(C)(k), \\ \mathrm{Pic}^0(C)(k) &= \mathrm{Div}^0(C)(k)/\mathrm{PDiv}^0(C)(k). \end{aligned}$$

We say that two divisors $D_1, D_2$ are *linearly equivalent over* $k$ and write $D_1 \sim D_2$ if their difference is a principal divisor in $\mathrm{PDiv}(C)(k)$.

THEOREM 5.5 (Abel). *Let $(E, o)$ be an elliptic curve over $k$. Then we have a natural isomorphism of groups*

$$E(k) \xrightarrow{\sim} Pic^0(E)(k), \quad p \mapsto [p] - [o].$$

*Proof. (a) Surjectivity.* By induction on the number of points that enter with nonzero multiplicity in a given divisor, it suffices to show that for all $p, q \in E(k)$ there exists $r \in E(k)$ with $[p] + [q] \sim [r] - [o]$. For this denote by

- $\ell$ the line through $p$ and $q$ (tangent to $E$ if $p = q$),
- $p \cdot q \in E(k)$ the third point of intersection in $\ell \cap E$,
- $m$ the line through $o$ and $p \cdot q$,
- $r = o \cdot (p \cdot q)$ the third point in $m \cap E$.

Let $g, h \in k[w, x, y]$ be homogenous linear forms defining the lines $\ell$ respectively $m$, then the rational function $F = g/h \in k(E)^*$ has zeroes and poles precisely at the points of intersection of the above two lines with the given elliptic curve and hence we get

$$\operatorname{div}(F) \;=\; [p] + [q] + [p \cdot q] - [p] - [p \cdot q] - [r] \;=\; [p] + [q] - [r] - [o].$$

*(b) Injectivity.* Given $p, q \in E(k)$ with $[p] - [o] \sim [q] - [o]$, there exists $F \in k(E)^*$ with divisor

$$\operatorname{div}(F) \;=\; [p] - [q].$$

Then $F = g/h$ for homogenous polynomials $g, h \in k[w, x, y]$ of the same degree $d$ such that

$$\begin{aligned} E \cdot C_g \;&=\; [p] + [p_1] + \cdots + [p_{3d-1}] \\ E \cdot C_h \;&=\; [q] + [p_1] + \cdots + [p_{3d-1}] \end{aligned}$$

for certain $p_1, \dots, p_{3d-1} \in E(\overline{k})$. Let $L \in k[w, x, y]$ be a homogenous linear form over $k$ with

$$E \cdot C_L \;=\; [o] + [p] + [o \cdot p],$$

then $(C_{hL} - C_g) \cdot E = [o] + [q] + [o \cdot p] \geq 0$, so corollary 3.7 gives a line $\ell \subset \mathbb{P}^2$ such that

$$\ell \cdot E \;=\; [o] + [q] + [o \cdot p].$$

By definition of the point $o \cdot p \in E(k)$ it follows that $q = p$ as required.

*(c) Additivity.* It is clear from the proof of surjectivity that the above bijection sends the group law from theorem 4.4 to the addition of divisors in $Pic^0(E)(k)$; note that here we have only used the definition of the group law, hence this gives an independent argument that + is indeed associative and commutative. □

Often we are not interested in the precise zeroes and poles but only want to bound them from below. For this we make the following

DEFINITION 5.6. For $D \in \operatorname{Div}(E)(k)$ the space of rational functions with poles bounded by $D$ is

$$\mathscr{L}(D) \;=\; \big\{ F \in k(E)^* \mid \operatorname{div}(F) \geq -D \big\}$$

Since Abel's theorem allows to decide when two divisors are linearly equivalent, the following gives complete control on rational functions with given divisors:

Theorem 5.7 (Riemann-Roch for elliptic curves). *Let $D \in \mathrm{Div}(E)(k)$.*

(1) *We have $\mathscr{L}(D) = \varnothing$ if $\deg(D) < 0$ or if $\deg(D) = 0$ but $D \not\sim 0$.*

(2) *Otherwise we have*

$$\dim_k \mathscr{L}(D) = \begin{cases} 1 & \text{if } D \sim 0, \\ d & \text{if } d = \deg(D) > 0. \end{cases}$$

*Proof.* If $\mathscr{L}(D) \neq \varnothing$, pick any $F \in \mathscr{L}(D)$. Then $\mathrm{div}(F) \geq -D$. Since principal divisors have degree zero, it follows in particular that $\deg(D) \geq 0$ and equality can only hold if $D = \mathrm{div}(F)$, in which case $D \sim 0$. In this last case we have an isomorphism

$$\mathscr{L}(D) \xrightarrow{\sim} \mathscr{L}(0), \quad G \mapsto F \cdot G$$

where $\mathscr{L}(0) = k$ because the only rational functions without zeroes or poles are the constant functions (exercise). Let us next discuss the case $\deg(D) = 1$. Here Abel's theorem gives a rational function $F \in k(E)$ whose divisor has the form $\mathrm{div}(F) = D - [p]$, where $p \in E(k)$ is the sum of the points in $D$ counted with multiplicities. Then

$$\mathscr{L}(D) \xrightarrow{\sim} \mathscr{L}([p]), \quad G \mapsto F \cdot G$$

and $\mathscr{L}([p]) = k$ since any rational function with at most a simple pole must be constant, for instance by the injectivity in Abel's theorem. The cases $\deg(D) \geq 2$ can now be treated by induction on the degree: Pick $p \neq q \in E(k)$. Without loss of generality we may assume $D \not\sim [p] + [q]$ (exercise). By induction on the degree there exists

$$G \ \in \ \mathscr{L}(D - [q]) \setminus \mathscr{L}(D - [p] - [q]) \ \subset \ \mathscr{L}(D).$$

Then the map

$$\varphi : \quad \mathscr{L}(D) \twoheadrightarrow k, \quad F \mapsto (F/G)(p)$$

is surjective because it is $k$-linear and satisfies $\varphi(G) = 1$. The kernel of this map $\varphi$ is

$$\ker(\varphi) \ = \ \{F \in \mathscr{L}(D) \mid \mathrm{ord}_p(F) > \mathrm{ord}_p(G) = n_p(D)\} \ = \ \mathscr{L}(D - [p]),$$

hence $\dim_k \mathscr{L}(D) = 1 + \dim_k \mathscr{L}(D - [p])$ and we are done by induction. □

## 6. Weierstrass normal forms

When talking about curves, we have so far only looked at coordinate changes induced by projective linear transformations of the ambient projective plane, see remark 1.4. For more flexibility we should allow more general morphisms:

Definition 6.1. Consider the plane curve $C_f \subset \mathbb{P}^2$ cut out by an irreducible homogenous polynomial $f \in k[w, x, y]$. Given rational functions $\varphi_0, \varphi_1, \varphi_2 \in k(C_1)$, not all three zero, with

$$g(\varphi_0, \varphi_1, \varphi_2) \ = \ 0$$

for some other irreducible homogenous polynomial $g \in k[r, s, t]$, we obtain a family of maps

$$C_f(K) \setminus \Sigma(K) \ \longrightarrow \ C_g(K), \quad p \ \mapsto \ \big[\varphi_0(p) : \varphi_1(p) : \varphi_2(p)\big]$$

where $K/k$ ranges over all extensions of the base field and where $\Sigma(K) \subset C_f(K)$ is the finite subset of points at which some $\varphi_i$ has a pole or at which all three of them vanish. Up to modifications of these finite subsets, the family of maps only depends on the point $\varphi = [\varphi_0 : \varphi_1 : \varphi_2] \in \mathbb{P}^2(k(C_f))$. We call $\varphi$ a *rational map* and write

$$\varphi \ = \ [\varphi_0 : \varphi_1 : \varphi_2] : \quad C_f \ \dashrightarrow \ C_g$$

to indicate that we think of this as a map between subsets of the respective two curves. There is an obvious notion of composition of rational maps; we say that a rational map $\varphi$ is *birational* if has an inverse $\varphi^{-1}$ for this notion of composition.

EXERCISE 6.2. Show that any non-constant rational map $\varphi : C_f \dashrightarrow C_g$ induces an embedding

$$\varphi^* : \quad k(C_g) \hookrightarrow k(C_f), \quad F \mapsto F(\varphi_0, \varphi_1, \varphi_2),$$

of function fields, and that this embedding is an isomorphism iff $\varphi$ is birational.

In the definition of rational maps there was a considerable ambiguity. We say that a rational map $\varphi : C_f \dashrightarrow C_g$ is *defined at* $p \in C_f(K)$ if it can be written in the form

$$\varphi \;=\; [\psi_0 : \psi_1 : \psi_2]$$

where the $\psi_i$ are rational functions defined at $p$ in the sense of 5.2 and $\psi_i(p) \neq 0$ for at least one index $i \in \{0, 1, 2\}$. We say that the rational map $\varphi$ is a *morphism* if it is defined at every point $p \in C_f(K)$, for all $K/k$. Finally, the rational map $\varphi$ is called an *isomorphism* if it is birational and both $\varphi$ and $\varphi^{-1}$ are morphisms. The following is an analog of Riemann's removable singularities theorem:

LEMMA 6.3. *If $C_f \subset \mathbb{P}^2$ is smooth, then any rational map $\varphi : C_f \dashrightarrow C_g$ is a morphism. Hence any birational map between smooth curves is an isomorphism.*

*Proof.* Suppose that the curve $C_f \subset \mathbb{P}^2$ is smooth. Given $p \in C_f(K)$ for some extension field $K/k$, pick $i_0 \in \{0, 1, 2\}$ such that the rational function $\varphi_{i_0} \not\equiv 0$ is not the zero function but the order of vanishing $\mathrm{ord}_p(\varphi_{i_0})$ is smallest possible. Then we have $\varphi = [\psi_0 : \psi_1 : \psi_2]$ where $\psi_i = \varphi_i/\varphi_{i_0}$. Since $\mathrm{ord}_p(\psi_i) \geq 0$ and $\psi_{i_0} \equiv 1$, it follows that the rational map $\varphi$ is defined at $p$. □

The smoothness is essential for the above: For $f(w, x, y) = wy^2 - x^3 \in k[w, x, y]$ the map

$$\varphi : \quad C_f \dashrightarrow \mathbb{P}^1, \quad [1 : x : y] \mapsto [x : y]$$

is birational, but it cannot be extended to a morphism. Coming back to smooth curves, the above notion of isomorphism is much more flexible than projective linear transformations. For smooth cubics it allows to pass to the following normal form[1] where any given point can be made into a flex point:

THEOREM 6.4. *Let $(E, o)$ be an elliptic curve with marked point $o \in E(k)$.*

(1) *There exists an isomorphism $\varphi : E \xrightarrow{\sim} C_f \subset \mathbb{P}^2$ onto a smooth plane cubic cut out in affine coordinates by an equation in the long Weierstrass form*

$$f(1, x, y) \;=\; y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6)$$

*with $a_i \in k$ such that the point $o \in E(k)$ is sent to $\varphi(o) = [0 : 0 : 1]$.*

(2) *If $char(k) \neq 2, 3$, then we can moreover assume $a_1 = a_2 = a_3 = 0$ so that the equation of the elliptic curve is transformed into the short Weierstrass form*

$$y^2 \;=\; x^3 - a_4x - a_6.$$

[1] The labelling of the coefficients is chosen in such a way that if we attach the weight $i$ to the coefficient $a_i$ and let $x, y$ have weights $2, 3$ respectively, then each summand in the equation will be homogenous of total weight six.

*Proof.* (1) Iteratively using the Riemann-Roch theorem 5.7, we get $X, Y \in k(E)$ such that

$$\begin{aligned} \mathscr{L}(1[o]) &= \langle 1 \rangle, \\ \mathscr{L}(2[o]) &= \langle 1, X \rangle, \\ \mathscr{L}(3[o]) &= \langle 1, X, Y \rangle, \\ \mathscr{L}(4[o]) &= \langle 1, X, Y, X^2 \rangle, \\ \mathscr{L}(5[o]) &= \langle 1, X, Y, X^2, XY \rangle, \end{aligned}$$

as a vector space over $k$, with $\operatorname{ord}_o(X) = -2$ and $\operatorname{ord}_o(Y) = -3$. At the next step we find

$$1, X, Y, X^2, XY, X^3, Y^2 \ \in \ \mathscr{L}(6[o])$$

but the right hand side has dimension six, so there is a nontrivial linear relation between the seven elements on the left. In this relation the coefficients of $y^2$ and $x^3$ cannot both vanish, indeed $\mathscr{L}(5[o])$ has dimension five. To normalize these two nonvanishing coefficients, we pick any rational function $Z \in k(E)^*$ with $\operatorname{ord}_o(Z) = 1$ and rescale $X, Y$ so that

$$\begin{aligned} (XZ^2)(o) &= +1, \\ (YZ^3)(o) &= -1. \end{aligned}$$

Taking $Z^6$ times our previous relation and evaluationg at the point $o \in E(k)$, we see that the coefficients of $y^2$ and $x^3$ are negative to each other. Without loss of generality we may take them to be $\pm 1$. Then our relation reads $f(X, Y) = 0$ for a polynomial

$$f(x,y) \ = \ y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6) \ \in \ k[x,y],$$

so we get a rational map

$$\varphi = [1 : X : Y] : \quad E \dashrightarrow C_f \ \subset \ \mathbb{P}^2$$

and by lemma 6.3 this will be an isomorphism if we can show that the curve $C = C_f$ is smooth and that $\varphi$ is birational.

One easily sees that $f(x, y) \in k[x, y]$ is irreducible, so we can talk about the function field of $C$ and the birationality of $\varphi$ follows from proposition 6.5 below. It remains to discuss smoothness. By direct inspection $[0 : 0 : 1] \in C(\overline{k})$ is a smooth point. Any other point in $C(\overline{k})$ has the form

$$p \ = \ [1 : x_0 : y_0] \ \in \ C(\overline{k})$$

with $x_0, y_0 \in \overline{k}$. Since we defined the notion of smoothness by passing to the algebraic closure of the base field, we may assume from the start that $x_0, y_0 \in k$ and then after a translation $(x, y) \mapsto (x - x_0, y - y_0)$ that $x_0 = y_0 = 0$. Note that the polynomial $f(x, y)$ will change under this translation, but the new polynomial will still have long Weierstrass form. Now if $[1 : 0 : 0] \in C(k)$ is a singular point of $C$, then the constant term and the linear terms in the polynomial $f(x, y)$ must vanish so that

$$f(x,y) \ = \ y^2 + a_1xy - (x^3 - a_2x^2).$$

Then $C = C_f$ is a nodal or cuspidal cubic, and from exercise 6.2 one easily sees that

$$\psi : \quad C \dashrightarrow \mathbb{P}^1 \quad \text{with} \quad \psi_i \ = \ \begin{cases} 1, & \text{for } i = 0, \\ (x/w)|_C & \text{for } i = 1, \\ (y/w)|_C & \text{for } i = 2, \end{cases}$$

is birational. Then $\psi \circ \varphi : E \dashrightarrow \mathbb{P}^1$ is a birational map between smooth curves, hence an isomorphism by lemma 6.3. But this is impossible since isomorphisms of smooth curves preserve the pole order of rational functions (exercise) and on $\mathbb{P}^1$ there exist rational functions with only a single simple pole. This contradiction shows that the curve $C = C_f$ must be smooth.

(2) Consider a cubic equation $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ in long Weierstrass form. If the base field has odd characteristic, then completing the square via the substitution $y \mapsto y - \frac{1}{2}(a_1x + a_3)$ we can make $a_1 = a_3 = 0$. If we also assume that the characteristic of the base field is not three, then completing the cube via a further substitution $x \mapsto x - \frac{1}{3}a_2$ we can make $a_2 = 0$. □

In the above proof we have used the following description for the function field of elliptic curves, which is the analog of theorem I.4.7:

Proposition 6.5. *Let $(E, o)$ be an elliptic curve. Let $X, Y \in k(E)$ be rational functions with no poles except at the point $o \in E(k)$ where their pole orders are given by* $\operatorname{ord}_o(X) = -2$ *and* $\operatorname{ord}_o(Y) = -3$. *Then any rational function $F \in k(E)$ can be written as*

$$F \;=\; h_1(X) + h_2(X) \cdot Y \quad \textit{with} \quad h_1, h_2 \;\in\; \overline{k}(x).$$

*If $F$ has poles at most in the point $o$, then we can in fact take $h_1, h_2 \in k[x]$.*

*Proof.* We first deal with the case that $F \in k(E)$ has no poles except possibly at $o \in E(k)$. If $F$ has no poles at all, then it is a constant and we are done. So we may assume that the pole order satisfies $-\operatorname{ord}_o(F) \geq 2$. Take $m \in \mathbb{N}_0$, $n \in \{0, 1\}$ with

$$\operatorname{ord}_o(F) \;=\; -2m - 3n \;=\; \operatorname{ord}_o(X^mY^n),$$

then for a suitable constant $c \in k^*$ the rational function $G = F - c \cdot X^mY^n$ will have $\operatorname{ord}_o(G) > \operatorname{ord}_o(F)$ and we are done by induction on the pole order.

To deal with the general case, if $\operatorname{ord}_a(F) < 0$ for some $a \in E(\overline{k}) \setminus \{o\}$, then the function

$$F_1 \;=\; (X - X(a))^N \cdot F$$

for $N \gg 0$ has the set of poles bounded by $F_1^{-1}(\infty) \subseteq F^{-1}(\infty) \setminus \{a\}$. Inductively we find points $a_1, \dots, a_n \in E(\overline{k}) \setminus \{0\}$ and natural numbers $N_1, \dots, N_n \in \mathbb{N}$ such that the rational function

$$F_n \;=\; F \cdot \prod_{i=1}^{N} (X - X(a_i))^{N_i} \;\in\; \overline{k}(E)$$

has poles only in the point $o$, which reduces us to the case considered above. □

## 7. The $j$-invariant

In the last section we have seen that every elliptic curve is isomorphic to an elliptic curve in Weierstrass normal form, with the flex at infinity as its marked point. Here by an isomorphism of elliptic curves we mean an isomorphism of cubic curves sending the marked points to each other. To classify elliptic curves up to isomorphism, it remains to see to what extent the Weierstrass form in theorem 6.4 is unique. The only ambiguity comes from projective linear transformations:

Lemma 7.1. *Let $E, E' \subset \mathbb{P}^2$ be elliptic curves defined by two equations in long Weierstrass form, with the point at infinity $o = [0 : 0 : 1]$ as the marked point. Then* *any isomorphism* $\varphi : (E, o) \xrightarrow{\sim} (E', o)$ *is given in affine charts by a coordinate change*

$$\begin{aligned} x &\mapsto \lambda^2 x + a \\ y &\mapsto \lambda^3 y + b \cdot \lambda^2 x + c \qquad \textit{with } \lambda \in k^* \textit{ and } a, b, c \in k. \end{aligned}$$

*Proof.* Let $X, Y \in k(E)$ and $X', Y' \in k(E')$ be the rational functions on the elliptic curves obtained as the restriction of the rational functions $x/w, y/w$ on the ambient projective plane. We have already remarked earlier that isomorphisms between smooth curves preserve the order of poles and zeroes of rational functions, hence

$$\begin{aligned} \varphi^*(X') &\in \mathscr{L}(2[o]) = \langle 1, X \rangle, \\ \varphi^*(Y') &\in \mathscr{L}(3[o]) = \langle 1, X, Y \rangle. \end{aligned}$$

So we have

$$\begin{aligned} \varphi^*(X') &= \alpha X + a, \\ \varphi^*(Y') &= \beta X + \gamma Y + c, \end{aligned}$$

with $\alpha, \beta, \gamma, a, c \in k$. Taking the Weierstrass equation satisfied by $(X, Y)$ and the pullback of the Weierstrass equation satisfied by $(X', Y')$, we get in the function field $k(E)$ two equations

$$\begin{aligned} Y^2 - X^3 - \cdots &= 0, \\ \gamma^2 Y^2 - \alpha^3 X^3 - \cdots &= 0, \end{aligned}$$

where $\cdots$ are terms of pole order less than six. These equations imply $\gamma^2 = \alpha^3$ and hence

$$\begin{aligned} \gamma &= \gamma^3/\gamma^2 = \gamma^3/\alpha^3 = \lambda^3 \\ \alpha &= \alpha^3/\alpha^2 = \gamma^2/\alpha^2 = \lambda^2 \end{aligned}$$

for $\lambda = \gamma/\alpha$. Now the result follows with $b = \lambda^{-2} \cdot \beta$. □

Of course for any given pair of elliptic curves only very few values of $(\lambda, a, b, c)$ actually occur in the above lemma. Let us be more specific in the case of short Weierstrass equations:

COROLLARY 7.2. *Let* $E = C_f$ *and* $E' = C_g$ *be defined by short Weierstrass polynomials*

$$\begin{aligned} f(1, x, y) &= y^2 - (x^3 - a_4 x - a_6), \\ g(1, x, y) &= y^2 - (x^3 - b_4 x - b_6), \end{aligned}$$

*and put* $o = [0 : 0 : 1]$*. If* $\operatorname{char}(k) \neq 2, 3$*, then any isomorphism* $\varphi : (E, o) \xrightarrow{\sim} (E', o)$ *has the form*

$$(x, y) \mapsto (\lambda^2 x, \lambda^3 y) \quad \textit{for some } \lambda \in k^* \textit{ with} \quad \begin{cases} a_4 = \lambda^4 b_4, \\ a_6 = \lambda^6 b_6. \end{cases}$$

*Proof.* With notations as in the previous lemma the isomorphism $\varphi$ must be a linear map with coefficients $(\lambda, a, b, c)$. Since $\varphi(E) \subset C' = C_g$, it follows that the cubic

$$h(x, y) = g(1, \lambda^2 x + a, \lambda^3 y + b \cdot \lambda^2 x + c) \in k[x, y]$$

restricts to zero on the affine part of the smooth cubic curve $E = C_f$, which can happen only if

$$h(x, y) = \gamma \cdot f(1, x, y) \quad \text{for some } \gamma \in k \setminus \{0\}.$$

Since the right hand side is in short Weierstrass form, it follows that $h(x,y)$ does not involve any of the monomials $xy, y, x^2$ and for $\mathrm{char}(k) \neq 2,3$ we get by direct inspection that $a = b = c = 0$. But then $\gamma = \lambda^6$ and the claim follows. □

In what follows we will assume for simplicity that $\mathrm{char}(k) \neq 2,3$. We have seen in theorem 6.4 that then any elliptic curve is isomorphic to one cut out by a short Weierstrass equation

$$f(1,x,y) \;=\; y^2 \;-\; (x^3 + a_4 x + a_6) \;=\; 0$$

with $a_4, a_6 \in k$. Note that after the substitution $y \mapsto 2y$ this Weierstrass equation becomes

$$y^2 \;=\; 4x^3 - g_2 x - g_3 \quad \text{where} \quad \begin{cases} g_2 \;=\; -4a_4, \\ g_3 \;=\; -4a_6. \end{cases}$$

So in analogy with the complex case, we define the *discriminant* and the *j-invariant* for $\mathrm{char}(k) \neq 2,3$ by

$$\begin{aligned} \Delta(f) \;&=\; g_2^3 - 27g_3^2 \\ j(f) \;&=\; 12^3\, g_2^3/\Delta(f) \end{aligned}$$

where for the latter we use

Lemma 7.3. *We have $\Delta(f) \neq 0$ iff the cubic $C_f \subset \mathbb{P}^2$ is smooth.*

*Proof.* This follows from lemma 4.2, because the discriminant vanishes iff the polynomial $4x^3 - g_2 x - g_3$ has a double root in $\overline{k}$. □

We can now classify elliptic curves up to isomorphism over any algebraically closed field in the same way as we did over the complex numbers:

Theorem 7.4. *(a) If the base field $k = \overline{k}$ is algebraically closed, the $j$-invariant induces a bijection*

$$j: \quad \big\{ \text{isomorphism classes of elliptic curves } (E,o) \text{ over } \overline{k} \,\big\} \;\xrightarrow{\sim}\; \mathbb{A}^1(\overline{k}).$$

*(b) For $\mathrm{char}(k) \neq 2,3$ the automorphism groups of elliptic curves over $\overline{k}$ are given by*

$$Aut_{\overline{k}}(E,o) \;\simeq\; \begin{cases} \mathbb{Z}/4\mathbb{Z} & \text{for } j(E) = 0, \\ \mathbb{Z}/6\mathbb{Z} & \text{for } j(E) = 12^3, \\ \mathbb{Z}/2\mathbb{Z} & \text{otherwise.} \end{cases}$$

*Proof.* We give the proof only for $\mathrm{char}(k) \neq 2,3$, although statement (a) holds more generally. We first show that if $f, h \in k[w,x,y]$ are homogenous cubics in short Weierstrass form such that the corresponding elliptic curves $(C_f, o)$, $(C_g, o)$ are isomorphic over $\overline{k}$, then

$$j(f) \;=\; j(g).$$

Indeed, by definition the $j$-invariant does not change under extensions of the base fields and so we may assume that the two elliptic curves are already isomorphic over $k$. Since $\mathrm{char}(k) \neq 2,3$, corollary 7.2 then says that there exists $\lambda \in k^*$ such that

$$g_2(f) \;=\; \lambda^4\, g_2(g) \quad \text{and} \quad g_3(f) \;=\; \lambda^6\, g_3(g),$$

so the two cubics have the same $j$-invariant. Thus for an elliptic curve $(E,o)$ we can put

$$j(E) \;:=\; j(f) \quad \text{for any short Weierstrass cubic } f \text{ with } E \;\simeq\; C_f.$$

So far we have attached to any elliptic curve a $j$-invariant which depends only on the isomorphism class of the curve over the algebraic closure $\overline{k}$. We next claim that conversely the isomorphism class of $(E,o)$ over $\overline{k}$ is determined uniquely by the value $j = j(E)$. We may assume that the elliptic curve is cut out by a cubic equation in the short Weierstrass form $y^2 = x^3 + a_4x + a_6$. The idea is to look for $\lambda \in \overline{k}$ such that

$$E \xrightarrow{\sim} C_f, \quad (x,y) \mapsto (\lambda^2 x,\, \lambda^3 y)$$

where $f \in \overline{k}[w,x,y]$ is a short Weierstrass cubic depending only on $j$:

*Case 1.* For $j \neq 0, 12^3$ we have $a_4a_6 \neq 0$. Taking $\lambda \in \overline{k}^*$ with $\lambda^4 a_6 = \lambda^6 a_4$ we arrive at

$$f(x,y) \;=\; y^2 - (x^3 - cx - c)$$

for some $c \in \overline{k}$. One easily checks

$$j \;=\; j(f) \;=\; 12^3 \cdot \frac{4c}{4c-27} \quad\Longleftrightarrow\quad 4c \;=\; \frac{27j}{j-12^3}$$

so $f(x,y)$ only depends on $j$. Moreover, corollary 7.2 shows $Aut_{\overline{k}}(C_f, o) = \{\pm 1\}$.

*Case 2.* For $j = 12^3$ we have $a_6 = 0$ and $a_4 \neq 0$. Taking $\lambda \in \overline{k}^*$ with $\lambda^4 = a_4$ we arrive at

$$f(x,y) \;=\; y^2 - (x^3 + x),$$

and in this case corollary 7.2 shows $Aut_{\overline{k}}(C_f, o) = \{\lambda \in \overline{k} \mid \lambda^4 = 1\} \simeq \mathbb{Z}/4\mathbb{Z}$.

*Case 3.* For $j = 0$ we have $a_4 = 0$ and $a_6 \neq 0$. Taking $\lambda \in \overline{k}^*$ with $\lambda^6 = a_6$ we arrive at

$$f(x,y) \;=\; y^2 - (x^3 + 1)$$

and in this case corollary 7.2 shows $Aut_{\overline{k}}(C_f, o) \;=\; \{\lambda \in \overline{k} \mid \lambda^6 = 1\} \simeq \mathbb{Z}/6\mathbb{Z}$. □

Remark 7.5. Even if two elliptic curves $E, E' \subset \mathbb{P}^2$ are both defined by short Weierstrass equations over the base field $k$ and have $j(E) = j(E')$, they are not necessarily isomorphic over $k$. However, the same argument like in the above proof shows that they will become isomorphic over any field extension $K \supset k$ containing a solution $\lambda \in K^*$ of the equations $\lambda^4 = b_4/a_4$ and $\lambda^6 = b_6/a_6$.

Remark 7.6. For $\operatorname{char}(k) = 2, 3$ the above definitions do not work, but there is a more general definition that works in all cases. To motivate it, consider a long Weierstrass equation

$$f(1,x,y) \;=\; y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6) \;=\; 0.$$

If $\operatorname{char}(k) \neq 2$, then completing the square by the substitution $\tilde{y} = \frac{1}{2}(y - a_1x - a_3)$ we get

$$\tilde{y}^2 \;=\; 4x^3 + b_2x^2 + 2b_4x + b_6 \quad \text{where} \quad \begin{cases} b_2 \;=\; a_1^2 + 4a_2, \\ b_4 \;=\; a_1a_3 + 2a_4, \\ b_6 \;=\; a_3^2 + 4a_6. \end{cases}$$

Let $\alpha_1, \alpha_2, \alpha_3 \in \overline{k}$ denote the roots of the cubic polynomial $4x^3 + b_2x^2 + 2b_4x + b_6$, then for the discriminant of this cubic polynomial one can check with some patience that

$$\Delta(f) \;:=\; 16 \cdot \prod_{i<j} (\alpha_i - \alpha_j)^2 \;\in\; \mathbb{Z}[a_1, \dots, a_6]$$

is an integer polynomial in the coefficients $a_1, \dots, a_6$ of the original Weierstrass equation. This polynomial makes sense regardless of the base field and gives a general definition of the discriminant; similarly one can define the $j$-invariant. For $\operatorname{char}(k) \neq 2, 3$ this agrees with our previous definition, and part (a) of theorem 7.4 holds in general, though part (b) and corollary 7.2 have to be modified, see the books by Husemöller or Silverman.

CHAPTER III

# Arithmetic of elliptic curves

## 1. Rational points on elliptic curves

The theory from the preceding chapter was mostly about geometric properties that depend only on the algebraic closure of the base field. In this chapter we will discuss some more arithmetic results about the group of rational points on elliptic curves over number fields. Let $E$ be an elliptic curve over a subfield $k \subset \mathbb{Q}$. By the first chapter

$$E(\mathbb{C}) \;\simeq\; \mathbb{C}/\Lambda$$

is a complex torus, the quotient of the complex plane by a lattice $\Lambda \subset \mathbb{C}$. Over the reals we have

Lemma 1.1. *Let $E$ be the elliptic curve cut out by $y^2 = g(x)$ where $g(x) \in \mathbb{R}[x]$ is a real cubic, then*

$$E(\mathbb{R}) \;=\; \begin{cases} S^1 & \textit{if } g(x) \textit{ has only one real root,} \\ S^1 \times \mathbb{Z}/2\mathbb{Z} & \textit{if } g(x) \textit{ has three real roots.} \end{cases}$$

*Proof.* We have an embedding as a closed submanifold $E(\mathbb{R}) \subset \mathbb{P}^2(\mathbb{R})$. Since the projective plane is a compact real manifold of dimension two, it follows that $E(\mathbb{R})$ is a compact one-dimensional real Lie group. Now it is well known that the unit circle $S^1$ is the only connected compact abelian Lie group. Furthermore, as a real manifold the set $E(\mathbb{R})$ has at most two connected components:

It is connected if $g(x)$ has a single root, otherwise it has two components. □

If the elliptic curve $E \subset \mathbb{P}^2$ is defined by an equation over $k = \mathbb{Q}$, then we may ask about the structure of the set $E(\mathbb{Q})$ of its rational points. This is a more subtle problem:

Example 1.2. Let $E \subset \mathbb{P}^2$ be the elliptic curve defined by $y^2 = x^3 - 432$. Then we have

$$p \;=\; (12, 36) \;\in\; E(\mathbb{Q}),$$

and a direct computation shows that this is a point of order three on the elliptic curve. To see whether there are any other points in $E(\mathbb{Q})$, note that the coordinate transform

$$(x, y) \;\mapsto\; (u, v) \;=\; \left(\tfrac{6}{x} + \tfrac{y}{6},\, \tfrac{6}{x} - \tfrac{y}{6}\right)$$

gives an isomorphism between the elliptic curve $E$ and the Fermat cubic cut out by $u^3 + v^3 = 1$. So

$$E(\mathbb{Q}) \;=\; \{o, \pm p\} \;\simeq\; \mathbb{Z}/3\mathbb{Z},$$

but knowing this is equivalent to Fermat's last theorem for the exponent $p = 3$!

Once we have guessed one nontrivial point on an elliptic curve, we obtain other points by taking its multiples: For instance, let $E \subset \mathbb{P}^2$ be the elliptic curve defined by $y^2 = x^3 - 43x + 166$. Trying some small integer values we find $p = (3, 8) \in E(\mathbb{Q})$, and a direct computation shows

$$\begin{aligned} 2p &= (-1, -2), \\ 3p &= (3, -6), \\ 4p &= (3, +6) \;=\; -3p, \end{aligned}$$

Here $p$ is a torsion point of order seven. In general, the torsion points on any elliptic curve over $\mathbb{Q}$ can be computed by first applying a coordinate change to put it into Weierstrass form with integer coefficients, and then using the following

Theorem 1.3 (Lutz-Nagell). *Let $E$ be the elliptic curve cut out by a Weierstrass equation*

$$y^2 \;=\; x^3 + ax + b$$

*with integer coefficients $a, b \in \mathbb{Z}$. Then any rational torsion point $(x_0, y_0) \in E(\mathbb{Q})$ satisfies*

(1) *$x_0, y_0 \in \mathbb{Z}$, and*

(2) *$y_0^2$ divides $\Delta = 4a^2 + 27b^2$.*

In fact the main content of the theorem is the statement that any rational torsion point has the integrality property (1), we will show this in section 2 by considering the reduction of the elliptic curve modulo a prime. The divisibility property (2) is then an easy consequence:

Lemma 1.4. *Let $E$ be an elliptic curve cut out by a Weierstrass equation with integer coefficients as in theorem 1.3. Let $p = (x_0, y_0) \in E(\mathbb{Q})$ be a rational point such that both $p$ and $2p$ have integer coordinates. Then we must have $y_0^2 \mid \Delta$.*

*Proof.* We may clearly assume $y_0 \neq 0$. Writing the Weierstrass equation for $E$ as $y^2 = g(x)$ with $g(x) \in \mathbb{Z}[x]$, we have

$$T_pE \;=\; \{(x, y) \mid y = \alpha x + \beta\} \quad \text{where} \quad \alpha \;=\; \frac{g'(x_0)}{2y_0}, \quad \beta \;=\; y_0 - \alpha x_0.$$

TODO As we have seen in the exercises, the point $2p = (x_1, x_2)$ can then be computed by the formula

$$x_1 \;=\; \alpha^2 - 2x_0.$$

By assumption $x_1 \in \mathbb{Z}$, hence we obtain that $\alpha \in \mathbb{Z}$ and therefore $y_0 \mid g'(x_0)$. On the other hand we also know that $y_0 \mid y_0^2 = g(x_0)$. Now it is a general fact that the discriminant of a polynomial can be written as a linear combination of the polynomial and its derivative, with the coefficients themselves polynomials. In our case we have

$$\Delta \;=\; -27(x^3 + ax - b)g(x) + (3x^2 + 4a)g'(x)^2$$

in $\mathbb{Z}[x]$. Inserting $x = x_1$ we obtain $y_1^2 \mid \Delta$ as required. □

Remark 1.5. (a) It is *not* true that any multiple of an integer point on an elliptic curve has integer coordinates: On the elliptic curve $E$ cut out by $y^2 = x^3 - x + 1$ the point $p = (0,1) \in E(\mathbb{Q})$ has $2p = (1/4, 7/8)$. In such cases the Lutz-Nagell theorem shows that we have found a point of infinite order in $E(\mathbb{Q})$.

(b) By the Lutz-Nagell theorem, to find the torsion subgroup $E(\mathbb{Q})_{tors} \subset E(\mathbb{Q})$ we only need to check for which divisors $y_0 \mid \Delta$ the cubic $g(x) - y_0^2$ has an integer root. Note that the discriminant depends on the Weierstrass equation, and to ease computations we should choose coordinates which make it minimal. For instance, if $E$ is defined by $y^2 = x^3 - 1$, then $\Delta = 27$ and the Lutz-Nagell theorem implies that any rational torsion point must have the form $(x_0, y_0)$ with $y_0 \in \{0, \pm 1, \pm 3\}$, and we get

$$E(\mathbb{Q})_{tors} \;=\; \{o, (0, \pm 1), (-1, 0), (2, \pm 3)\} \;\simeq\; \mathbb{Z}/6\mathbb{Z}.$$

(c) The Lutz-Nagell theorem already implies that the torsion group $E(\mathbb{Q})_{tors}$ is finite. In fact a much stronger assertion holds: A deep theorem of Mazur[1] says the only options are

$$E(\mathbb{Q})_{tors} \;\simeq\; \begin{cases} \mathbb{Z}/n\mathbb{Z} & \text{with } 1 \leq n \leq 10 \text{ or } n = 12, \\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2n\mathbb{Z} & \text{with } 1 \leq n \leq 4. \end{cases}$$

Conversely, each of these occurs for infinitely many elliptic curves over $\mathbb{Q}$.

Let us now pass from the torsion subgroup to the entire group $E(\mathbb{Q})$. In contrast with the former, the latter is in general not easy to compute. However, it turns out that all points in $E(\mathbb{Q})$ can be obtained from some finite set of points by means of the geometric group law. The result holds more generally over any number field:

Theorem 1.6 (Mordell-Weil). *If $E$ is an elliptic curve over a number field $K$, then the group of its rational points is a finitely generated abelian group.*

In other words, we have an isomorphism $E(K) \simeq E(K)_{tors} \times \mathbb{Z}^r$ for a unique integer $r = r_{E/K} \in \mathbb{N}_0$ which is called the *rank* of the elliptic curve. This rank is one of the most mysterious invariants of elliptic curves:

Even for elliptic curves defined by Weierstrass equations $y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Q}$, there is no known *effective* algorithm to determine the rank or even the question whether $E(\mathbb{Q})$ is finite or not from $a, b$. It is still unknown whether there exist elliptic curves of arbitrarily high rank, though this is believed to be the case. Finally, the conjecture of Birch and Swinnerton-Dyer gives a deep connection to analytic number theory: It predicts in particular that the rank is the order of vanishing

$$r_{E/\mathbb{Q}} \;=\; \mathrm{ord}_{s=1} L_{E/\mathbb{Q}}(s)$$

of a certain analytic function obtained as the analytic continuation of an Euler product $L_{E/\mathbb{Q}}(s) = \prod_p L_p(s)$, where $p$ runs over all primes and the local factors depend only on the number of points of the reduction $\overline{E}$ of the elliptic curve over $\mathbb{F}_p$ in the sense of section 2 below. If $E$ is cut out by $y^2 = g(x)$ with $g(x) \in \mathbb{Z}[x]$, it implies

$$r_{E/\mathbb{Q}} \;>\; 0 \quad\Longleftrightarrow\quad \lim_{n\to\infty} \textstyle\prod_{p \leq n} \frac{N_p}{p} \;=\; \infty$$

for the number of points $N_p \;=\; \#\overline{E}(\mathbb{F}_p) = 1 + \#\big\{(x,y) \in \mathbb{A}^2(\mathbb{F}_p) \mid y^2 = g(x)\big\}$.

[1] Barry Mazur, *Rational isogenies of prime degree*, Invent. Math. 44 (1978) 129–162.

We will prove the Mordell-Weil theorem in sections 4 and 5. The idea is to construct a suitable measure for the size of rational points, for which we use the following notion:

Definition 1.7. A *norm* on an abelian group $G$ is a function $|\cdot|: G \to \mathbb{R}_{\geq 0}$ such that

(1) $|mp| = |m||p|$ for all $p \in G$ and $m \in \mathbb{Z}$,
(2) $|p+q| \leq |p| + |q|$ for all $p, q \in G$,
(3) for each $c \in \mathbb{R}$ the set $G_c = \{p \in G \mid |p| \leq c\}$ is finite.

Any finitely generated abelian group $G$ can be endowed with a norm by taking an isomorphism $G/G_{tors} \simeq \mathbb{Z}^r$ and composing it with the standard Euclidean norm on the target. Conversely the existence of a norm on an abelian group forces its torsion subgroup to be finite but does *not* imply finite generation, e.g. take an infinite orthogonal sum $G = \bigoplus_{n \in \mathbb{N}} \mathbb{Z}$ where the $n$-th summand is endowed with $n$ times the standard Euclidean norm. However:

Lemma 1.8. *An abelian group $G$ is finitely generated iff*

(1) *it admits a norm, and*
(2) *$G/mG$ is finite for some integer $m > 1$.*

*Proof.* If $n = \#G/mG$ is finite, write $G = \biguplus_{i=1}^{n}(r_i + mG)$ where the $r_i \in G$ form a set of representatives for the cosets in the group. If the group comes with a norm $|\cdot|$, consider the finite subset

$$G_c \ \supseteq\ \{r_1, \dots, r_n\} \quad \text{for} \quad c \ =\ \max_i |r_i| + 1.$$

We claim that this finite subset generates the entire group. Indeed, let $g \in G \backslash G_c$ be any element in the complement. Since the $g_i$ form a full set of coset representatives, we have $g - r_{i_1} = mg_1 \in mG$ for some $i_1 \in \{1, \dots, n\}$, $g_1 \in G$. But

$$\begin{aligned} |r_{i_1}| \ &<\ c && \text{by definition of } c \\ &<\ |g| && \text{by assumption on } g \\ &\leq\ (m-1)|g| && \text{since } m > 1. \end{aligned}$$

and hence

$$|g_1| \ =\ \frac{|mg_1|}{m} \ =\ \frac{|g - r_{i_1}|}{m} \ \leq\ \frac{|g| + |r_{i_1}|}{m} \ <\ \frac{|g| + (m-1)|g|}{m} \ =\ |g|.$$

Inductively we find

$$\begin{aligned} g \ &=\ r_{i_1} + mg_1 \quad \text{with } |g_1| < |g|, \\ g_1 \ &=\ r_{i_2} + mg_2 \quad \text{with } |g_2| < |g_1|, \\ &\vdots \\ g_{\nu-1} \ &=\ r_{i_\nu} + mg_\nu \quad \text{with } |g_\nu| < |g_{\nu-1}|, \end{aligned}$$

until at some point we arrive at the situation where $|g_\nu| \leq c$. But then $g_\nu \in G_c$ and writing

$$g \ =\ r_{i_1} + mr_{i_2} + m^2 r_{i_3} + \cdots + m^{\nu-1} r_{i_\nu} + m^\nu g_\nu$$

we obtain a representation of the given element as a sum of elements of $G_c$. □

The procedure in the above proof is called the method of *infinite descent* and goes back to Fermat. Note that the argument is constructive, one we know a system of representatives for the cosets in $G/mG$.

So the proof of the Mordell-Weil theorem for an elliptic curve $E$ over a number field $K$ naturally falls into two steps:

- In section 4 we show $E(K)/mE(K)$ is finite, using group cohomology.
- In section 5 we construct a norm on $E(K)$, using the notion of heights.

The argument for the first step also gives explicit upper bounds for $\#E(K)/mE(K)$ which sometimes suffice to determine the rank, but there is still no known effective algorithm that works in general: Often the best one can do is to give lower bounds for the rank by constructing many linearly independent rational points. The current record is an elliptic curve of rank $r_{E/\mathbb{Q}} \geq 28$ that was found by Elkies in 2016.

## 2. Reduction modulo primes and torsion points

Reduction of rational points modulo prime numbers is a very powerful technique in arithmetic geometry. We here only explain the most basic version. Let $R$ be a unique factorization domain and $k$ its quotient field, e.g. $R = \mathbb{Z}$, $k = \mathbb{Q}$. For a prime element $p \in R$ we denote by $\nu_p(a) \in \mathbb{N}_0$ the multiplicity with which it enters in the prime factorization of a given element $a \in R \setminus \{0\}$. This extends uniquely to a homomorphism $\nu_p : k^* \to \mathbb{Z}$ which is called the *p-adic valuation* of the field $k$, a discrete valuation whose residue field we denote by $\mathbb{F}_p = R/pR$. Formally we also put $v_p(0) = \infty$. For $n \in \mathbb{N}$ we define the *reduction map*

$$\begin{aligned} \mathbb{P}^n(k) &\longrightarrow \mathbb{P}^n(\mathbb{F}_p), \\ x = [x_0 : \cdots : x_n] &\mapsto \overline{x} = [\overline{y}_0 : \cdots : \overline{y}_n] \end{aligned}$$

where $y_i = p^{-\nu} x_i$ for $\nu = \min\{\nu_p(x_i) \mid 0 \leq i \leq n\}$, $\overline{y}_i = y_i \bmod p \in \mathbb{F}_p$. For elliptic curves we proceed similarly, but some care is needed with our choice of Weierstrass equations:

Definition 2.1. A Weierstrass equation for an elliptic curve $E$ over $k$ is *integral* if it has the form

$$g(x,y) \;=\; y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6) \;=\; 0 \quad \text{with} \quad a_i \;\in\; R.$$

It is called *minimal* if moreover its discriminant $\Delta \in R$ has the smallest possible valuation $\nu_p(\Delta) \in \mathbb{N}_0$ among all integral Weierstrass equations for the given elliptic curve. We then take the *reduction of the elliptic curve* to be the curve $\overline{E}$ over $\mathbb{F}_p$ cut out by the reduction $\overline{g}(x,y) \in \mathbb{F}_p[x,y]$. Thus we obtain a reduction map on rational points by restricting the reduction map for the projective plane:

$$\begin{tikzcd} E(k) \arrow[r] \arrow[d, phantom, "\cap" description] & E(\mathbb{F}_p) \arrow[d, phantom, "\cap" description] \\ \mathbb{P}^2(k) \arrow[r] & \mathbb{P}^2(\mathbb{F}_p) \end{tikzcd}$$

To see that these notions are well-defined one has to work a bit:

Lemma 2.2. *Any elliptic curve $E$ over $k$ has a minimal Weierstrass equation, and any two such equations are related to each other by a coordinate change of the form $(x,y) \mapsto (\lambda^2 x + a, \lambda^3 y + b \cdot \lambda^2 x + c)$ with $\nu_p(a), \nu_p(b), \nu_p(c) \geq 0$ and $\nu_p(\lambda) = 0$.*

*Proof.* We know from lemma II.7.1 that any two Weierstrass equations are related by a coordinate change of the above form with $a, b, c, \lambda \in k$, $\lambda \neq 0$. The transformation formula for the coefficients and discriminant under such a change shows that if both equations are minimal, then $\nu_p(\lambda) = 0$ and $\nu_p(a), \nu_p(b), \nu_p(c) \geq 0$. For details of the computation see [Silverman, prop. VII.1.3]. □

We say that an elliptic curve $E$ over $k$ has *good reduction* at a prime $p$ if its reduction $\overline{E}$ is smooth, which happens iff $p \nmid \Delta$. In this case the reduction is again an elliptic curve and

$$\rho : \quad E(k) \longrightarrow E(\mathbb{F}_p)$$

is a group homomorphism because the reduction map for the projective plane sends lines to lines. We put

$$\begin{aligned} E^1(k) &= \ker(\rho) \\ &= \left\{ [w:x:y] \in E(k) \,\middle|\, w,x,y \in R,\ p \mid w,\ p \mid x,\ p \nmid y \right\} \\ &= \left\{ [w:x:y] \in E(k) \,\middle|\, y \neq 0,\ \nu_p(x/y) \geq 1,\ \nu_p(w/y) \geq 1 \right\} \end{aligned}$$

and more generally

$$E^n(k) \;=\; \{[w:x:y] \in E^1(k) = \ker(\rho) \mid \nu_p(x/y) \geq n\}$$

for $n \in \mathbb{N}$. This gives a decreasing filtration

$$E(k) \supseteq E^1(k) \supseteq E^2(k) \supseteq \cdots$$

which is called the *p-adic filtration.* We have

Remark 2.3. $\bigcap_{n\in\mathbb{N}} E^n(k) = \{0\}$.

*Proof.* If $[w:x:y] \in E^n(k)$ for all $n \in \mathbb{N}$, then ...

Remark 2.4. Let $[1:x:y] \in E(k)$. If $\nu_p(x) < 0$, then there exists $n \in \mathbb{N}$ such that

$$\nu_p(x) \;=\; -2n \quad \textit{and} \quad \nu_p(y) \;=\; -3n.$$

*Proof.* A look at the integral Weierstrass equation shows that if $\nu_p(x) < 0$, then necessarily $\nu_p(y^2) = \nu_p(x^3)$. □

Proposition 2.5. *The $E^n(k) \subset E(k)$ are subgroups for all $n \in \mathbb{N}$. Furthermore, the map*

$$t: \quad E^n(k) \;\to\; p^n R, \quad [w:x:y] \;\mapsto\; x/y$$

*induces an embedding*

$$E^n(k)/E^{n+i}(k) \;\hookrightarrow\; p^nR/p^{n+i}R \;\simeq\; R/p^iR \quad \textit{for any} \quad i \;\in\; \{1,2,\dots,4n\}.$$

*Proof.* Both claims will follow if we can show for all $p_1, p_2 \in E^n(k)$ that

$$t(p_1+p_2) \;\equiv\; t(p_1)+t(p_2) \mod p^{5n}R.$$

check $p = 2$

For this we look at the group law in a neighborhood of the origin $o = [0:0:1] \in E(k)$, like in the study of Lie groups and Lie algebras. Note that since we assumed $2 \nmid p$, we can assume that our chosen minimal Weierstrass equation for $E$ has the short form

$$y^2 \;=\; x^3 + ax + b$$

with $a, b \in R$. Then in the affine coordinates $(s,t) = (1/y, x/y)$ around the point $o$, the elliptic curve is cut out by

$$(\star) \qquad s \;=\; t^3 + ats^2 + bs^3$$

as one sees by putting $(x,y) = (t/s, 1/s)$ in the previous Weierstrass equation. In these coordinates, the line through two points $p_1 = (s_1,t_1), p_2 = (s_2,t_2) \in E(k)$ can be written in the form

$$\ell \;=\; \{(s,t) \in \mathbb{A}^2(k) \mid s = \alpha t + \beta\} \quad \text{with} \quad \alpha \;=\; \begin{cases} \frac{s_2-s_1}{t_2-t_1} & \text{if } t_2 \neq t_1, \\ \cdots & \text{otherwise,} \end{cases}$$

and some $\beta \in k$. To describe the group law, let $(s,t) \in (E \cap \ell)(k)$ be the third point of intersection of the line with the elliptic curve. Inserting the equation for the line in $\star$ we get

$$\alpha t + \beta \; = \; t^3 + at \cdot (\alpha t + \beta)^2 + b(\alpha t + \beta)^3$$

In this equation,

- the coefficient of $t^3$ is $= 1 + a\alpha^2 + \alpha^3 b$,
- the coefficient of $t^2$ is $= 2a\alpha\beta + 3b\alpha^2\beta$

## 3. An intermezzo on group cohomology

## 4. The weak Mordell-Weil theorem

## 5. Heights and the Mordell-Weil theorem