Plane Algebraic Curves

Andreas Gathmann

Class Notes RPTU Kaiserslautern 2023

Contents

0. Introduction

These notes are meant as a gentle introduction to *algebraic geometry*, a combination of *linear algebra* and *algebra*:

(a) In linear algebra (as e. g. in the "Foundations of Mathematics" class [G2]), we study systems of linear equations in several variables over a fixed ground field K.

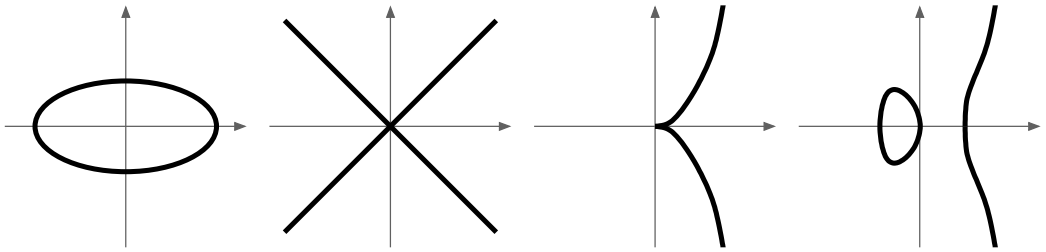
(b) In algebra (as e. g. in the "Algebraic Structures" or "Introduction to Algebra" classes [G1, G3]), a central topic are polynomials in one variable over K.

Algebraic geometry combines this by studying systems of polynomial equations in several variables over K. Of course, such polynomials in several variables occur in many places both in pure mathematics and in applications. Consequently, algebraic geometry has become a very large and active field of mathematics with deep connections to many other areas, such as commutative algebra, computer algebra, number theory, cryptography, topology, and complex analysis, just to name a few.

On the one hand, all these connections make algebraic geometry into a very interesting field to study – but on the other hand they may also make it hard for the beginner to get started. So to keep everything digestible, we will restrict ourselves here to the first case that is covered by neither (a) nor (b) above: *one polynomial equation in two variables*. Its set of solutions in K^2 can then be thought of as a curve in the plane, we can draw it (at least in the case $K = \mathbb{R}$), ask geometric questions about it, and try to answer them with algebraic methods. This restriction will significantly reduce the required theoretical background, but still leads to many interesting results that we will discuss in these notes.

To get a feeling for the kind of problems that one may ask about plane curves, we will now mention a few of them in this introductory chapter. Their flavor differs quite a bit depending on the chosen ground field K.

Example 0.1 (Curves over $\mathbb{R}$)**.** The following picture shows some real plane curves. Note that they can have many different "shapes": The curve (a) lies in a bounded region of the plane, whereas the others do not. The curve (b) consists of two components in the sense that it can be decomposed into two subsets (given by $x+y=0$ and $x-y=0$) that are given by polynomial equations themselves. The curve (c) has a so-called singularity at the origin, i. e. a point where it does not locally look like a smoothly deformed real line (in fact, (b) has a singularity at the origin as well). Finally, the image in (d) consists of two disconnected parts, but these parts are *not* given by separate polynomial equations themselves, as we will see in Exercise 1.8.



$\frac{x^2}{4}+y^2-1=0$ (a) $\quad x^2-y^2=0$ (b) $\quad y^2-x^3=0$ (c) $\quad y^2+x-x^3=0$ (d)

It is a main goal of algebraic geometry to prove such properties of curves just from looking at the polynomials, i. e. without drawing and referring to a picture (which would not be an exact proof anyway). Other related questions we might ask are: In how many points can two curves intersect? How many singularities can a curve have?

Example 0.2 (Curves over $\mathbb{C}$)**.** Over the complex numbers, pictures of curves will look different since a 1-dimensional complex object is real 2-dimensional, i. e. a surface. Note that we cannot draw such a surface as a subset of $K^2 = \mathbb{C}^2 = \mathbb{R}^4$ since we would need four dimensions for that. But we can still get a correct topological picture of the curve itself if we disregard this embedding. Let us show informally how to do this for the curve with the equation $y^2 + x - x^3 = 0$ as in Example 0.1 (d) above; for more details see 5.16.

Note that in this case it is actually possible to write down all the points of the curve explicitly, because the given equation

$$y^2 = x^3 - x = x(x-1)(x+1)$$

is (almost) solved for y already: We can pick x to be any complex number, and then get two values for y, namely the two square roots of $x(x-1)(x+1)$ – unless $x \in \{-1, 0, 1\}$, in which case there is only one value for y (namely 0).

So one might think that the curve looks like two copies of the complex plane, glued together at the three points $-1, 0, 1$: The complex plane parametrizes the values for x, and the two copies of it correspond to the two possible values for y, i. e. to the two roots of the number $x(x-1)(x+1)$.

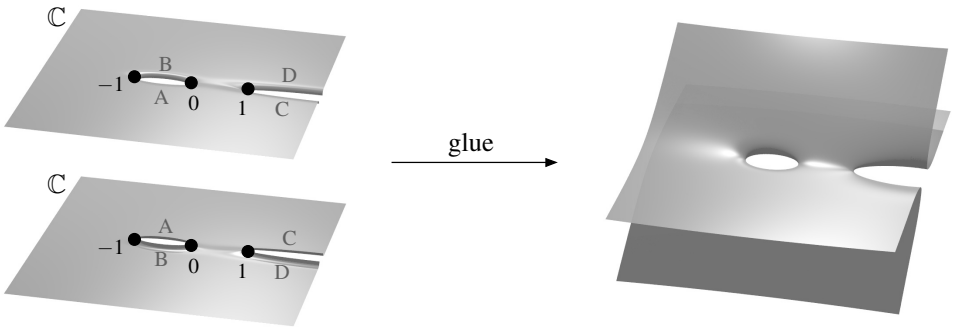
This is not the correct topological picture however, because a non-zero complex number does not have a distinguished first and second root that could correspond to the first and second copy of the complex plane. Rather, the two roots of a complex number get exchanged if we run around the origin once: If we consider a closed path

$$z = r\mathrm{e}^{\mathrm{i}\varphi} \qquad \text{for } 0 \leq \varphi \leq 2\pi \text{ and fixed } r > 0$$

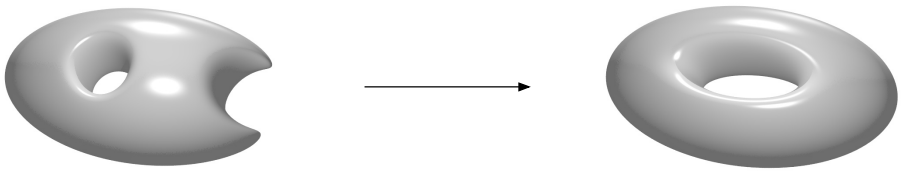
around the complex origin, the square root of this number would have to be defined by

$$\sqrt{z} = \sqrt{r}\mathrm{e}^{\frac{\mathrm{i}\varphi}{2}},$$

which gives opposite values at $\varphi = 0$ and $\varphi = 2\pi$. In other words, if x runs around one of the points $-1, 0, 1$ (i. e. around a point at which y is the square root of 0), we go from one copy of the plane to the other. One way to draw this topologically is to cut the two planes along the real intervals $(-1, 0)$ and $(1, \infty)$, and to glue the two planes along these edges as in the following picture on the left, where edges with the same letter are meant to be identified. The gluing itself is then visualized best by first turning one of the planes upside down; this is shown in the picture on the right.

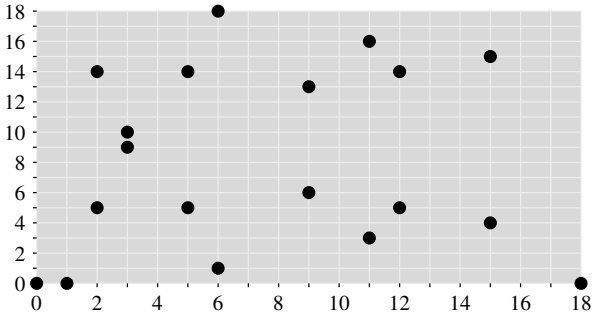


This is now actually a topologically correct picture of the given curve. To make the situation a little nicer, we can compactify it by adding a point at infinity, which corresponds to identifying the two planes at their infinitely far points as well (the precise construction will be described in Chapter 3). This is shown in the picture below, and leads topologically to a torus.



We will show in Proposition 5.16 how such topological pictures can be obtained immediately from the given equation of the curve.

Example 0.3 (Curves over finite fields)**.** Of course, over a finite field such as $\mathbb{Z}/p\mathbb{Z}$ for a prime number p, a curve just consists of finitely many points, and hence once again looks completely different. The following picture shows again the curve given by the equation $y^2+x-x^3=0$ as in our previous two examples, but this time over the ground field $\mathbb{Z}/19\mathbb{Z}$.



Note that we can still see the symmetry $y \leftrightarrow -y$ in the picture, but apart from that we just have a seemingly random collection of points. In fact, we will see in Example 7.9 that such curves have important applications in modern cryptography.

Example 0.4 (Curves over $\mathbb{Q}$)**.** The most famous application of algebraic geometry to ground fields other than just the real or complex numbers is probably Fermat's Last Theorem. This is the statement that, for $n \in \mathbb{N}_{\geq 3}$, the curve given by the equation $x^n+y^n-1=0$ *over the rational numbers* has only the trivial solutions where $x=0$ or $y=0$, or equivalently (by setting $x=\frac{a}{c}$ and $y=\frac{b}{c}$ for $a,b,c \in \mathbb{Z}$ with $c \neq 0$), that the equation $a^n+b^n=c^n$ has no non-trivial solutions over $\mathbb{Z}$ at all. Note that this picture is again very different from the cases of the other ground fields considered above. But as one might expect, a large part of the theory of algebraic curves works over arbitrary ground fields, and in fact the proof of Fermat's Last Theorem uses concepts of algebraic geometry in many places. So, in some sense, we can view (algebraic) number theory as a part of algebraic geometry.

Example 0.5 (Relations to complex analysis)**.** We have just seen in the examples above that algebraic geometry has deep relations to e. g. topology and number theory, and it should not come as a surprise that there are many relations to algebraic fields of mathematics such as commutative algebra and computer algebra as well. Although it is not within the scope of these notes, let us finish this introductory chapter by showing interesting relations to complex analysis as well.

Consider a (sufficiently nice) compactified complex curve, such as a torus as in Example 0.2. Of course, in algebraic geometry one does not only study curves for themselves but also maps between them; and hence we will have to consider "nice" functions on such curves (where "nice" will translate into "locally a quotient of polynomials"). What do such functions f look like if they are defined globally on the whole curve? As the curve is compact, note that the image of f must be a compact subset of the complex plane, which means that the absolute value $|f|$ must take a maximum somewhere. But locally the curve just looks like the complex plane, and by the Maximum Modulus Principle [G4, Proposition 6.14] the absolute value of a nice (read: holomorphic) function on the complex plane cannot have a local maximum unless it is constant. So we conclude that f must be a constant function: There are actually no non-trivial nice global functions on a compact curve.

In fact, we will prove this statement in Corollary 6.29 using only algebraic methods, and hence over arbitrary (algebraically closed) ground fields. In a similar way, many interesting results over the ground field $\mathbb{C}$ can be obtained using both algebraic geometry and complex analysis, with completely different methods, and thus give a close relation between these two branches of mathematics as well.

But let us now start with our study of plane curves. In order to keep these notes as accessible as possible, we will only assume a basic knowledge of groups, rings, and fields as about to the

extent of the "Algebraic Structures" class [G1], but a little more experience in dealing with these structures would certainly be advantageous. Very occasionally we will need to assume results from commutative algebra that go beyond these prerequisites (marked as "Facts" in the notes), but they will always be clearly stated and motivated, and provided with a reference. However, in order not to lose this very interesting part of the subject we will nevertheless quite frequently explore the relations of our results to other fields of mathematics in side remarks and excursions (that will then not be needed afterwards to follow the remaining parts of the notes).

1. Affine Curves

In this first chapter we will introduce plane curves both from an algebraic and a geometric point of view. As explained in the introduction, they will be given as solutions of polynomial equations. So let us start by fixing the corresponding notations.

Rings are always assumed to be commutative with a multiplicative neutral element 1. The multiplicative group of units of a ring R will be denoted by R^*.

Notation 1.1 (Polynomials)**.** Throughout these notes, K will always denote a fixed ground field. By $K[x_1,\dots,x_n]$ we will denote the *polynomial ring* in n variables $x_1,\dots,x_n$ over K, i. e. the ring of finite formal sums

$$f = \sum_{i_1,\dots,i_n\in\mathbb{N}} a_{i_1,\dots,i_n}\, x_1^{i_1}\cdot\dots\cdot x_n^{i_n}$$

with all $a_{i_1,\dots,i_n}\in K$ (see e. g. [G1, Chapter 9] how this concept of "formal sums" can be defined in a mathematically rigorous way). Note that we can regard it as an iterated univariate polynomial ring since $K[x_1,\dots,x_n] = K[x_1,\dots,x_{n-1}][x_n]$. Of course, for a polynomial f as above and a point $P=(c_1,\dots,c_n)\in K^n$, the *value* of f at P is defined as

$$f(P) := \sum_{i_1,\dots,i_n\in\mathbb{N}} a_{i_1,\dots,i_n}\, c_1^{i_1}\cdot\dots\cdot c_n^{i_n} \quad \in K.$$

Unless stated otherwise, the *degree* of a term $a_{i_1,\dots,i_n}\, x_1^{i_1}\cdot\dots\cdot x_n^{i_n}$ as above is meant to be the total degree $i_1+\dots+i_n$ in all variables together. The maximum degree occurring in a term with non-zero coefficient of a polynomial $f\neq 0$ is called the *degree* $\deg f$ of f. We call F *homogeneous* if all its terms have the same degree.

It is easy to see that $K[x_1,\dots,x_n]$ is an integral domain, and that $\deg(fg)=\deg f+\deg g$ holds for all non-zero polynomials f,g. The units of $K[x_1,\dots,x_n]$ are just the non-zero constant polynomials, which we can identify with $K^*=K\backslash\{0\}$.

Fact 1.2 (Factorial rings)**.** The polynomial ring $K[x_1,\dots,x_n]$ is a *factorial ring* (also called a *unique factorization domain*) [G6, Proposition 8.1 and Remark 8.4]. This means that prime and irreducible elements agree, and that every non-zero non-unit has a decomposition as a product of irreducible polynomials in a unique way (up to permutations, and up to multiplication with units). In the following, we will usually use this unique factorization property without mentioning. Note however that, as it is already the case for the integers $\mathbb{Z}$, performing such factorizations in $K[x_1,\dots,x_n]$ explicitly or even determining if a given polynomial is irreducible is usually hard.

Definition 1.3 (Affine varieties)**.**

(a) For $n\in\mathbb{N}$ we call $\mathbb{A}^n := \mathbb{A}^n_K := K^n$ the **affine n-space** over K.

It is customary to use the different notation $\mathbb{A}^n$ for K^n here since K^n is also a K-vector space and a ring. We will usually write $\mathbb{A}^n_K$ if we want to ignore these additional structures: For example, addition and scalar multiplication are defined on K^n, but not on $\mathbb{A}^n_K$. The affine space $\mathbb{A}^n_K$ will be the ambient space for our zero loci of polynomials below.

(b) For a subset $S\subset K[x_1,\dots,x_n]$ of polynomials we call

$$V(S) := \{P\in\mathbb{A}^n : f(P)=0 \text{ for all } f\in S\} \quad \subset \mathbb{A}^n$$

the (affine) **zero locus** of S. Subsets of $\mathbb{A}^n$ of this form are called **(affine) varieties**. If $S=\{f_1,\dots,f_k\}$ is a finite set, we will write $V(S)=V(\{f_1,\dots,f_k\})$ also as $V(f_1,\dots,f_k)$.

In these notes we will mostly restrict ourselves to zero loci of a single polynomial in two variables. We will then usually call these variables x and y instead of x_1 and x_2.

Remark 1.4. Obviously, for two polynomials $f, g \in K[x,y]$ we have ...

(a) $V(f) \cup V(g) = V(fg)$, as $fg(P) = 0$ for a point $P \in \mathbb{A}^2$ if and only if $f(P) = 0$ or $g(P) = 0$;

(b) $V(f) \cap V(g) = V(f,g)$ by definition.

One would probably expect now that a plane curve is just the zero locus of a polynomial in two variables. However, for our purposes it turns out to be more convenient to define a (plane) curve as such a polynomial itself rather than as its zero locus – this will simplify many statements and proofs later on when we want to study curves algebraically, i. e. in terms of their polynomials. Often, we will denote polynomials by capital instead of small letters if we want to think of them in this way. However, as it is obvious that two polynomials F and G with $F = \lambda G$ for some $\lambda \in K^*$ have the same zero locus (and thus determine the same geometric object), we incorporate this already in the definition of a curve:

Definition 1.5 (Affine curves)**.**

(a) An **(affine plane algebraic) curve** (over K) is a non-constant polynomial $F \in K[x,y]$ modulo units, i. e. modulo the equivalence relation $F \sim G$ if $F = \lambda G$ for some $\lambda \in K^*$. We will write it just as F, not indicating this equivalence class in the notation – this will not lead to any confusion.

We call $V(F) = \{P \in \mathbb{A}^2 : F(P) = 0\}$ the **set of points** of F.

(b) The **degree** of a curve is its degree as a polynomial. Curves of degree $1, 2, 3, \dots$ are usually referred to as **lines**, **quadrics/conics**, **cubics**, and so on.

(c) A curve F is called **irreducible** if it is as a polynomial, and **reducible** otherwise. Similarly, if $F = {F_1}^{a_1} \cdot \dots \cdot {F_k}^{a_k}$ is the irreducible decomposition of F as a polynomial (see Fact 1.2), we will also call this the **irreducible decomposition** of the curve F. The curves $F_1, \dots, F_k$ are then called the **(irreducible) components** of F and $a_1, \dots, a_k$ their **multiplicities**.

A curve F is called **reduced** if all its irreducible components have multiplicity 1.

Remark 1.6.

(a) Obviously, the notions of Definition 1.5 are well-defined, i. e. they do not change when multiplying a polynomial with a unit in K^*. All our future constructions with curves will also have this property, and it will be equally obvious in all these cases as well. In the following, we will therefore not mention this fact any more.

(b) In the literature, a curve often refers to the set of points $V(F)$ as in Definition 1.5 (a), i. e. to the geometric object in $\mathbb{A}^2$ rather than to the polynomial F.

Example 1.7. Especially in the case of the ground field $K = \mathbb{R}$, we will usually visualize a curve F by drawing its set of points $V(F)$ in the plane – although this does not contain the full information on the curve, as we will see below.

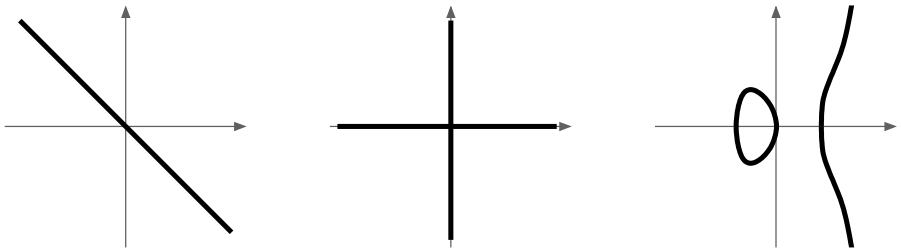
(a) The curve $x + y$ is a line, and hence irreducible (as a polynomial of degree 1 cannot be a product of two non-constant polynomials). Its square $(x+y)^2$ has the same set of points as $x + y$, but it is a quadric. It is neither irreducible nor reduced.

More generally, it is obvious that curves with the same irreducible components, just with different multiplicities, have the same set of points.

(b) The quadric xy is reducible as well, but it is reduced since it has two irreducible components x and y of multiplicity 1.

(c) In contrast to its appearance (see the picture below), the cubic $F = y^2 + x - x^3$ is irreducible: If we had $F = GH$ for some non-constant G and H, and thus $V(F) = V(G) \cup V(H)$ by Remark 1.4 (a), then one of these factors would have to be a line and the other one a quadric. But F does not contain a line as we can see from the picture.

(d) The set of points of the real curve $F = x^2 + y^2 + 1$ is empty, but by our definition F is nevertheless a curve – and also different from the curve $x^2 + y^2 + 2$, whose set of points is also empty. If we consider F over the complex numbers however, it has a non-empty set of points, but it is hard to visualize as it lies in $\mathbb{A}^2_{\mathbb{C}} = \mathbb{A}^4_{\mathbb{R}}$.



(a) (b) (c)

Exercise 1.8. Prove algebraically that the curve $y^2 + x - x^3$ of Example 1.7 (c) is irreducible.

Even if we defined a curve to be a polynomial (modulo scalars), we would of course rather like to think of it as a geometric object in $\mathbb{A}^2$ as in the pictures in Examples 0.1 or 1.7. For the rest of this chapter we will therefore study to what extent the set of points $V(F)$ determines back F, i. e. whether we can "draw $V(F)$ in the plane to specify F". We have already seen two reasons why this does not work in general:

- If a curve F is non-reduced as in Example 1.7 (a), we cannot determine the multiplicities on its components from $V(F)$.
- If (as in the case $K = \mathbb{R}$) there are non-constant polynomials without zeros, the set of points $V(F)$ might be empty as in Example 1.7 (d), and thus does not determine back F.

We now want to see that these are essentially the only two problems that can arise, and simultaneously prove that the intersection of two curves without a common component is finite. For this, we need two algebraic prerequisites.

Remark 1.9 (Algebraically closed fields)**.** A field K is called *algebraically closed* if every non-constant polynomial $F \in K[x]$ in one variable has a zero. The most prominent example is clearly $K = \mathbb{C}$ [G4, Proposition 6.20] – but it can be shown that every field is contained in an algebraically closed one, so that considering only curves over algebraically closed fields would not be a serious restriction. In fact, many textbooks on algebraic geometry restrict to this case altogether. In these notes however we will at least develop the general theory for arbitrary ground fields up to Chapter 5 in order not to exclude e. g. the geometrically most intuitive case of real curves from the very beginning.

Note that any algebraically closed field is necessarily infinite: If $K = \{c_1, \ldots, c_n\}$ was finite, the polynomial $F = \prod_{i=1}^{n}(x - c_i) + 1$ would have no zero.

Construction 1.10 (Quotient fields)**.** For any integral domain R, there is an associated *quotient field*

$$\operatorname{Quot} R = \left\{ \frac{a}{b} : a, b \in R \text{ with } b \neq 0 \right\},$$

where the "fraction" $\frac{a}{b}$ denotes the equivalence class of the pair (a, b) under the relation

$$(a, b) \sim (a', b') \;\Leftrightarrow\; ab' = a'b.$$

It is in fact a field with the standard addition and multiplication rules for fractions. The ring R is then a subring of $\operatorname{Quot} R$ by identifying $a \in R$ with $\frac{a}{1} \in \operatorname{Quot} R$ [G6, Example 6.5 (b)].

The easiest example is $R = \mathbb{Z}$, in which case we just have $\operatorname{Quot} R = \mathbb{Q}$. For our purposes the most important example is the polynomial ring $R = K[x_1, \ldots, x_n]$, for which $\operatorname{Quot} R$ is denoted $K(x_1, \ldots, x_n)$ and called the *field of rational functions* over K. Note that, despite its name, its elements are not defined as functions, but rather as formal quotients of polynomials as e. g. $\frac{x_1 + x_2}{x_1 - x_2} \in K(x_1, x_2)$. They do, however, define functions on the subset of $\mathbb{A}^n$ where the denominator is non-zero.

Lemma 1.11. *Let F be an affine curve.*

(a) *If K is algebraically closed then $V(F)$ is infinite.*

(b) *If K is infinite then $\mathbb{A}^2_K \setminus V(F)$ is infinite.*

Proof. As F is not a constant polynomial, it has positive degree in at least one of the variables x and y. By symmetry we may assume that this is x, so that $F = a_n x^n + \cdots + a_0$ for some $a_0, \ldots, a_n \in K[y]$ with $n > 0$ and $a_n \neq 0$.

Being non-zero, the polynomial $a_n \in K[y]$ has only finitely many zeros. But K is in any case infinite by Remark 1.9, hence there are infinitely many $y \in K$ with $a_n(y) \neq 0$. For each such y, the polynomial $F(x,y)$ is non-constant in x, so in case (a) there is an $x \in K$ with $F(x,y) = 0$, and in case (b) there is an $x \in K$ with $F(x,y) \neq 0$ (as $F(\cdot, y)$ has only finitely many zeros). □

01

Proposition 1.12 (Finiteness of the intersection of curves)**.** *Let F and G be two curves without a common component.*

(a) *The ideal $\langle F, G \rangle$ in $K[x,y]$ contains a non-zero polynomial that depends only on x (and hence by symmetry also a non-zero polynomial that depends only on y).*

(b) *The intersection $V(F,G)$ of the two curves is finite.*

Proof.

(a) By assumption, F and G are coprime in $K[x,y]$. We claim that they are then also coprime in $K(x)[y]$. In fact, if they had a common factor in $K(x)[y]$ then after clearing denominators we would have $aF = HF'$ and $aG = HG'$ for some $H, F', G' \in K[x,y]$ and non-zero $a \in K[x]$, where H has a positive y-degree. But then every irreducible factor of a must divide H or both F' and G' in $K[x,y]$. So by replacing H or both F' and G' by these quotients we arrive at a new decomposition $F = HF'$ and $G = HG'$ with $H, F', G' \in K[x,y]$ and H of positive y-degree, in contradiction to F and G being coprime in $K[x,y]$.

Now the ring $K(x)[y]$ as a univariate polynomial ring over a field $K(x)$ is a principal ideal domain [G1, Example 10.23]. So as $F, G \in K(x)[y]$ are coprime we can write 1 as a linear combination of F and G with coefficients in $K(x)[y]$ [G1, Proposition 10.13 (b)], which means after clearing denominators again that $c = DF + EG$ for some $D, E \in K[x,y]$ and a non-zero $c \in K[x]$. Hence, $c \in \langle F, G \rangle$ is a non-zero polynomial that depends only on x.

(b) Continuing the above notation, if $P \in V(F,G)$ we have $c(P) = D(P)F(P) + E(P)G(P) = 0$. This restricts the x-coordinate of all points $P \in V(F,G)$ to the finitely many zeros of c. By symmetry, we then also have only finitely many choices for the y-coordinate, i. e. $V(F,G)$ is finite. □

Corollary 1.13. *Let F be a curve over an algebraically closed field. Then for any irreducible curve G we have*

$$G \mid F \;\Leftrightarrow\; V(G) \subset V(F).$$

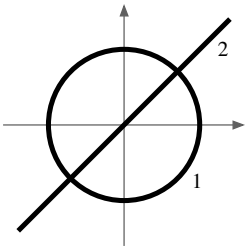
In particular, the irreducible components of F (but not their multiplicities, see Example 1.7 (a)) can be recovered from $V(F)$.

Proof.

"⇒" Assume that $F = GH$ for some curve H. If $P \in V(G)$, i. e. $G(P) = 0$, then we also have $F(P) = G(P)H(P) = 0$, and hence $P \in V(F)$.

"⇐" Now assume that $V(G) \subset V(F)$. Then $V(F,G) = V(G)$ is infinite by Lemma 1.11 (a). By Proposition 1.12 (b) this means that F and G must have a common component. As G is irreducible, this is only possible if $G \mid F$. □

Remark 1.14 (Specifying a curve by its set of points)**.** By Corollary 1.13, over an algebraically closed field we can specify a curve by giving its set of points together with a multiplicity on each irreducible component. For example, the picture on the right (where the circle has radius 1 and the numbers at the components are their multiplicities) represents the curve $(x^2+y^2-1)(x-y)^2$. (Note however that this is a real picture, but Corollary 1.13 would only hold over $\mathbb{C}$.)



If we do not specify multiplicities in a picture, we usually mean the corresponding reduced curve, i. e. where all multiplicities are 1.

Notation 1.15. Due to the above correspondence between a curve F and its set of points $V(F)$, we will sometimes write:

(a) $P \in F$ instead of $P \in V(F)$, i. e. $F(P) = 0$ ("P lies on the curve F");

(b) $F \cap G$ instead of $V(F,G)$ for the points that lie on both F and G;

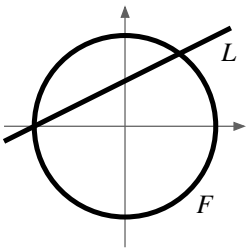
(c) $F \cup G$ for the curve FG (see Remark 1.4 (a));

(d) $G \subset F$ instead of $G \mid F$.

Exercise 1.16 (Pythagorean triples in algebraic geometry)**.** Let $F = x^2+y^2-1 \in K[x,y]$ be the "unit circle" over K. Assume that the characteristic of K is not 2, i. e. that $1+1 \neq 0$ in K.

(a) Considering the intersection points of an arbitrary line L (with slope t) through $(-1,0)$ with F, show that the set of points of F is

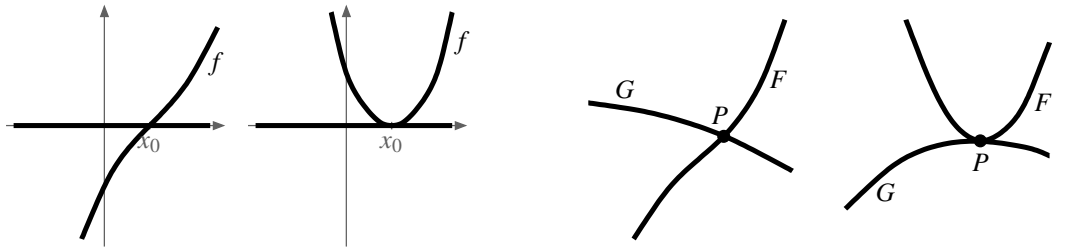
$$V(F) = \{(-1,0)\} \cup \left\{ \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2} \right) : t \in K \text{ with } 1+t^2 \neq 0 \right\}.$$



(b) Using (a), prove that the integer solutions (a,b,c) of the equation $a^2+b^2=c^2$ (the so-called Pythagorean triples) are, up to a permutation of a and b, exactly the triples of the form $\lambda(u^2-v^2, 2uv, u^2+v^2)$ with $\lambda, u, v \in \mathbb{Z}$.

2. Intersection Multiplicities

Let us start our study of curves by introducing the concept of intersection multiplicity, which will be central throughout these notes. It generalizes the well-known notion of multiplicity of a zero of a univariate polynomial: If $f \in K[x]$ is a polynomial and $x_0 \in K$ such that $f = (x-x_0)^m g$ for a polynomial $g \in K[x]$ with $g(x_0) \neq 0$, then f is said to have multiplicity m at x_0. As in the following two pictures on the left, a zero of multiplicity 1 means that the graph of f intersects the x-axis transversely, whereas in the case of multiplicity (at least) 2 it is tangent to it. Roughly speaking, higher multiplicities would correspond to graphs for which the x-axis is an even better approximation around x_0.



multiplicity 1 $f = (x-x_0)\,g$ | multiplicity 2 $f = (x-x_0)^2\,g$ | multiplicity 1 | multiplicity 2

In this geometric interpretation, we have considered how the graph of f intersects the horizontal axis locally at the given point, i. e. how the two curves $F = y - f$ and $G = y$ intersect. As in the picture above on the right, this concept should thus also make sense for arbitrary curves F and G at an intersection point P: If they intersect transversely, i. e. with different tangent directions, we want to say that they have an intersection multiplicity of 1 at P, whereas equal tangents correspond to higher multiplicities. But of course, the curves F and G might also have "singularities" as e. g. the origin in Example 0.1 (b) and (c), in which case it is not clear a priori how their intersection multiplicity can be interpreted or even defined.

So our first task must be to actually construct the intersection multiplicity for arbitrary curves. For this we need the following algebraic object that allows us to capture the local geometry of the plane around a point.

Definition 2.1 (Local rings of $\mathbb{A}^2$)**.** Let $P \in \mathbb{A}^2$ be a point.

(a) The **local ring** of $\mathbb{A}^2$ at P is defined as

$$\mathscr{O}_P := \mathscr{O}_{\mathbb{A}^2,P} := \left\{ \frac{f}{g} : f,g \in K[x,y] \text{ with } g(P) \neq 0 \right\} \quad \subset K(x,y).$$

(b) It admits a well-defined ring homomorphism

$$\mathscr{O}_P \to K,\ \frac{f}{g} \mapsto \frac{f(P)}{g(P)}$$

which we will call the **evaluation map**. Its kernel will be denoted by

$$I_P := I_{\mathbb{A}^2,P} := \left\{ \frac{f}{g} : f,g \in K[x,y] \text{ with } f(P) = 0 \text{ and } g(P) \neq 0 \right\} \quad \subset \mathscr{O}_P.$$

Remark 2.2 (Geometric and algebraic interpretation of local rings)**.** Intuitively, $\mathscr{O}_P$ describes "nice" (i. e. rational) functions that have a well-defined value at P (determined by the evaluation map), and thus also in a neighborhood of P. Note however that $\mathscr{O}_P$ does not admit similar evaluation maps

at other points $Q \neq P$ since the denominator of the fractions might vanish there. This explains the name "local ring" from a geometric point of view. The ideal I_P in $\mathcal{O}_P$ describes exactly those local functions that have the value 0 at P.

Algebraically, $\mathcal{O}_P$ is a subring of $K(x,y)$ that contains $K[x,y]$. As a subring of a field it is an integral domain, and its units are precisely the fractions $\frac{f}{g}$ for which both f and g are non-zero at P. Moreover, just like $K[x,y]$ it is a factorial ring, with the irreducible elements being the irreducible polynomials that vanish at P (since the others have become units).

For those who know some commutative algebra we should mention that $\mathcal{O}_P$ is also a local ring in the algebraic sense, i. e. that it contains exactly one maximal ideal, namely I_P [G6, Definition 6.9]: If I is any ideal in $\mathcal{O}_P$ that is not a subset of I_P then it must contain an element $\frac{f}{g}$ with $f(P) \neq 0$ and $g(P) \neq 0$. But this is then a unit since $\frac{g}{f} \in \mathcal{O}_P$ as well, and hence we have $I = \mathcal{O}_P$.

In fact, in the algebraic sense $\mathcal{O}_P$ is just the localization of the polynomial ring $K[x,y]$ at the maximal ideal $\langle x - x_0, y - y_0 \rangle$ associated to the point $P = (x_0, y_0)$ – which also shows that it is a local ring [G6, Corollary 6.10].

Definition 2.3 (Intersection multiplicities)**.** For a point $P \in \mathbb{A}^2$ and two curves (or polynomials) F and G we define the **intersection multiplicity** of F and G at P to be

$$\mu_P(F,G) := \dim \mathcal{O}_P/\langle F,G\rangle \quad \in \mathbb{N} \cup \{\infty\},$$

where dim denotes the dimension as a vector space over K.

As this definition is rather abstract, we should of course figure out how to compute this number, what its properties are, and why it captures the geometric idea given above. In fact, it is not even clear whether $\mu_P(F,G)$ is finite. But let us start with a few simple statements and examples.

Remark 2.4.

(a) It is clear from the definitions that an invertible *affine coordinate transformation* from (x,y) to

$$(x',y') = (ax+by+c, dx+ey+f) \quad \text{for } a,b,c,d,e,f \in K \text{ with } ae-bd \neq 0$$

gives us an isomorphism between the local rings $\mathcal{O}_P$ and $\mathcal{O}_{P'}$, where P' is the image point of P; and between $\mathcal{O}_P/\langle F,G\rangle$ and $\mathcal{O}_{P'}/\langle F',G'\rangle$, where F' and G' are F and G expressed in the new coordinates x' and y'. We will often use this invariance to simplify our calculations by picking suitable coordinates, e. g. such that $P = 0$ is the origin.

(b) The intersection multiplicity is symmetric: We have $\mu_P(F,G) = \mu_P(G,F)$ for all F and G.

(c) For all F,G,H we have $\langle F, G+FH\rangle = \langle F,G\rangle$, and thus $\mu_P(F,G+FH) = \mu_P(F,G)$.

In Definition 2.3, we have not required a priori that P actually lies on both curves F and G. However, the intersection multiplicity is at least 1 if and only if it does:

Lemma 2.5. *Let $P \in \mathbb{A}^2$, and let F and G be two curves (or polynomials). We have:*

(a) $\mu_P(F,G) \geq 1$ *if and only if* $P \in F \cap G$*;*

(b) $\mu_P(F,G) = 1$ *if and only if* $\langle F,G\rangle = I_P$ *in* $\mathcal{O}_P$*.*

Proof. Assume first that $F(P) \neq 0$. Then F is a unit in $\mathcal{O}_P$, and thus $\langle F,G\rangle = \mathcal{O}_P$, i. e. $\mu_P(F,G) = 0$. Moreover, we then have $P \notin F$ and $F \notin I_P$, proving both (a) and (b) in this case. Of course, the case $G(P) \neq 0$ is analogous.

So we may now assume that $F(P) = G(P) = 0$, i. e. $P \in F \cap G$. Then the evaluation map at P induces a well-defined and surjective map $\mathcal{O}_P/\langle F,G\rangle \to K$. It follows that $\mu_P(F,G) \geq 1$, proving (a) in this case. Moreover, we have $\mu_P(F,G) = 1$ if and only if this map is an isomorphism, i. e. if and only if $\langle F,G\rangle$ is exactly the kernel I_P of the evaluation map. □

Example 2.6 (Intersection multiplicity of coordinate axes)**.** The kernel I_0 of the evaluation map at 0 consists exactly of the fractions $\frac{f}{g}$ such that f does not have a constant term, which is just the ideal $\langle x,y\rangle$ in $\mathscr{O}_0$. By Lemma 2.5 (b) this means that $\mu_0(x,y) = 1$, i. e. (as expected) that the two coordinate lines have intersection multiplicity 1 at the origin.

Regarding the finiteness of the intersection multiplicity, the following two exercises show that $\mu_P(F,G)$ is finite if and only if F and G do not have a common component through P. This should not come as a surprise since an infinite intersection multiplicity should mean that the two curves "touch at P to infinite order", i. e. that they agree locally around P in the irreducible case, resp. share a common component in the general case. By Remark 2.4 (a) it suffices to consider the case when $P = 0$ is the origin.

Exercise 2.7 (Finiteness of the intersection multiplicity)**.** Let F and G be two curves without a common component that passes through the origin. Show:

(a) There is a number $n \in \mathbb{N}$ such that $x^n = y^n = 0$ in $\mathscr{O}_0/\langle F,G\rangle$.

(b) Every element of $\mathscr{O}_0/\langle F,G\rangle$ has a polynomial representative.

(c) $\mu_0(F,G) < \infty$.

Exercise 2.8 (Infinite intersection multiplicities)**.** Let F and G be two curves that pass through the origin. Show:

(a) If F and G have no common component then the family $(F^n)_{n\in\mathbb{N}}$ is linearly independent in $\mathscr{O}_0/\langle G\rangle$.

(b) If F and G have a common component that passes through the origin then $\mu_0(F,G) = \infty$.

For the last important basic property of intersection multiplicities we first need another easy algebraic tool.

Construction 2.9 (Short exact sequences)**.** We say that a sequence

$$0 \longrightarrow U \xrightarrow{\varphi} V \xrightarrow{\psi} W \longrightarrow 0$$

of linear maps between vector spaces (where 0 denotes the zero vector space) is **exact** if the image of each map equals the kernel of the next, i. e. if

(a) $\ker\varphi = 0$ (i. e. φ is injective);

(b) $\operatorname{im}\varphi = \ker\psi$; and

(c) $\operatorname{im}\psi = W$ (i. e. ψ is surjective).

In this case, we get a dimension formula

$$\begin{aligned}\dim U + \dim W &\overset{\text{(a),(c)}}{=} \dim\operatorname{im}\varphi + \dim\operatorname{im}\psi = \dim\operatorname{im}\varphi + \dim V/\ker\psi \overset{\text{(b)}}{=} \dim\operatorname{im}\varphi + \dim V/\operatorname{im}\varphi\\ &= \dim V.\end{aligned}$$

Proposition 2.10 (Additivity of intersection multiplicities)**.** *Let $P \in \mathbb{A}^2$, and let F,G,H be any three curves (or polynomials).*

(a) *If F and G have no common component through P there is an exact sequence*

$$0 \longrightarrow \mathscr{O}_P/\langle F,H\rangle \xrightarrow{\cdot G} \mathscr{O}_P/\langle F,GH\rangle \xrightarrow{\pi} \mathscr{O}_P/\langle F,G\rangle \longrightarrow 0,$$

where π is the natural quotient map.

(b) *We have $\mu_P(F,GH) = \mu_P(F,G) + \mu_P(F,H)$.*

Proof.

(a) We may assume that F and G have no common component at all, since components that do not pass through P are units in $\mathscr{O}_P$ and can therefore be dropped in the ideals.

It is checked immediately that both non-trivial maps in this sequence are well-defined, and that conditions (b) and (c) of Construction 2.9 hold. Hence we just have to show that the first multiplication map is injective: Assume that $\frac{f}{g}$ is in the kernel of this map, i. e. that

$$\frac{f}{g} \cdot G = \frac{f'}{g'} \cdot F + \frac{f''}{g''} \cdot GH$$

for certain $f', f'', g', g'' \in K[x,y]$ with $g'(P)$ and $g''(P)$ non-zero. We may assume without loss of generality that all three fractions have the same denominator, and multiply by it to obtain the equation $fG = f'F + f''GH$ in $K[x,y]$. Now G clearly divides fG and $f''GH$, hence also $f'F$, and consequently f' as F and G have no common component. So we have $f' = aG$ for some $a \in K[x,y]$, and we see that $fG = aFG + f''GH$. Dividing by G, it follows that $f = aF + f''H$, so that f and hence also $\frac{f}{g}$ are zero in $\mathcal{O}_P/\langle F,H \rangle$. This shows the injectivity of the first map.

(b) If F and G have no common component through P the statement follows immediately from (a) by taking dimensions as in Construction 2.9. Otherwise the equation is true as $\infty = \infty$ by Exercise 2.8 (b). □

02

Touching the mathematical field of computer algebra, we are now ready to explicitly compute the intersection multiplicity $\mu_P(F,G)$ of two arbitrary curves F and G at a point P where they do not have a common component. By Remark 2.4 (a) it suffices to do this at the origin $P = 0$. Let us start with the simple case when one of the curves is the horizontal axis; this will be needed in the general algorithm afterwards.

Example 2.11 (Intersection multiplicity with the horizontal axis)**.** Let F be an affine curve that does not contain the horizontal axis y. We want to compute the intersection multiplicity $\mu_0(y,F)$ with this axis at the origin.

By Remark 2.4 (c) we may remove all multiples of y from F, i. e. replace F by the polynomial $F(x,0) \in K[x]$, which is not the zero polynomial since y is not a component of F. We can write $F(x,0) = x^m g$ where $g \in K[x]$ is non-zero at the origin, so that m is the multiplicity of 0 in $F(x,0)$. Hence we obtain

$$\begin{aligned}
\mu_0(y,F) &= \mu_0(y,F(x,0)) && \text{(Remark 2.4 (c))} \\
&= \mu_0(y,x^m g) && \\
&= m\,\mu_0(y,x) + \mu_0(y,g) && \text{(Proposition 2.10 (b))} \\
&= m && \text{(Example 2.6 and Lemma 2.5 (a)).}
\end{aligned}$$

Note that this coincides with the expectation from the beginning of this chapter: If $f \in K[x]$ is a univariate polynomial with a zero x_0 of multiplicity m (which is just $x_0 = 0$ in our current case) then the intersection multiplicity of its graph $y - f$ with the x-axis at the point $(x_0, 0)$ is m.

Algorithm 2.12 (Computation of the intersection multiplicity $\mu_0(F,G)$)**.** Let F and G be two curves (or polynomials) without common component through the origin. We then repeat the following procedure recursively to compute the intersection multiplicity $\mu_0(F,G)$:

(a) If $F(0) \neq 0$ or $G(0) \neq 0$, i. e. if one of the curves does not pass through the origin, we stop with $\mu_0(F,G) = 0$ by Lemma 2.5 (a).

(b) Otherwise, if F and G both contain a monomial independent of y, we write

$$\begin{aligned}
F &= ax^m + (\text{terms involving } y \text{ or with a lower power of } x), \\
G &= bx^n + (\text{terms involving } y \text{ or with a lower power of } x)
\end{aligned}$$

for some $a, b \in K^*$ and $m, n \in \mathbb{N}_{>0}$, where we may assume (by possibly swapping F and G) that $m \geq n$. Similarly to a standard polynomial long division we then set

$$F' := F - \frac{a}{b} x^{m-n} G,$$

hence canceling the x^m-term in F. By Remark 2.4 (c) we then have $\mu_0(F,G) = \mu_0(F',G)$, so we can replace F by F' (which also passes through the origin) and repeat this step (b). As this procedure makes the number $m+n$ strictly smaller in each step, we will eventually arrive at a situation with one of the polynomials not having a monomial independent of y, leading to the final case:

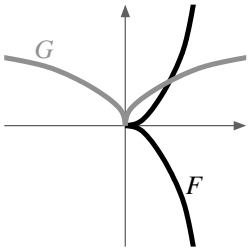
(c) If one of the polynomials F and G, say F, does not contain a monomial independent of y, we can factor $F = yF'$ and obtain by Proposition 2.10 (b)

$$\mu_0(F,G) = \mu_0(y,G) + \mu_0(F',G).$$

In this expression, the multiplicity $\mu_0(y,G)$ can be computed directly by Example 2.11: It is the lowest power of x in a term of G independent of y. Note that this number is non-zero as $G(0) = 0$. Hence we have $\mu_0(F',G) < \mu_0(F,G)$; so if we now repeat the algorithm recursively to compute $\mu_0(F',G)$ it will terminate in finitely many steps.

Example 2.13. Let us compute the intersection multiplicity $\mu_0(F,G)$ at the origin of the two curves $F = y^2 - x^3$ and $G = x^2 - y^3$ as in the picture below on the right. We follow Algorithm 2.12 and indicate which step we performed each time:

$$\begin{aligned}
\mu_0(y^2-x^3, x^2-y^3) &\overset{\text{(b)}}{=} \mu_0(y^2 - x^3 + x(x^2-y^3), x^2-y^3) \\
&= \mu_0(y^2 - xy^3, x^2-y^3) \\
&\overset{\text{(c)}}{=} \underbrace{\mu_0(y, x^2-y^3)}_{=2 \text{ by } 2.11} + \mu_0(y - xy^2, x^2-y^3) \\
&\overset{\text{(c)}}{=} 2 + \underbrace{\mu_0(y, x^2-y^3)}_{=2 \text{ by } 2.11} + \underbrace{\mu_0(1-xy, x^2-y^3)}_{=0 \text{ by (a)}} \\
&= 4.
\end{aligned}$$



Remark 2.14 (Curves with common components)**.** If F and G have a common component through 0, Algorithm 2.12 still performs correct computations, but it might not terminate. For example, for the curves $F = x^2$ and $G = xy - x$ with common component x it yields

$$\begin{aligned}
\mu_0(x^2, xy-x) &\overset{\text{(b)}}{=} \mu_0(x^2 + x(xy-x), xy-x) \\
&= \mu_0(x^2y, xy-x) \\
&\overset{\text{(c)}}{=} \underbrace{\mu_0(y, xy-x)}_{=1 \text{ by } 2.11} + \mu_0(x^2, xy-x),
\end{aligned}$$

leading to an infinite loop. However, if for arbitrary given F and G it does terminate with a finite answer, then by Exercise 2.8 (b) we have proven simultaneously with this computation that F and G have no common component through the origin. In contrast, if the algorithm does not seem to terminate we will find in Remark 4.8 (c) a rigorous way to decide whether F and G have a common component through 0.

Exercise 2.15. Draw the real curves $F = x^2 + y^2 + 2y$ and $G = y^3x^6 - y^6x^2$, determine their irreducible decompositions, their intersection points, and their intersection multiplicities at these points.

Following our algorithm, we can now also give an easy and important criterion for when the intersection multiplicity is 1.

Notation 2.16 (Homogeneous parts of polynomials)**.** For a polynomial $F \in K[x,y]$ of degree d and $i = 0,\dots,d$, we define the *degree-i part* of F to be the sum of all terms of F of degree i. Hence all F_i are homogeneous, and we have $F = F_0 + \dots + F_d$. We call F_0 the *constant part*, F_1 the *linear part*, and F_d the *leading part* of F.

Proposition 2.17 (Intersection multiplicity 1)**.** *Let F and G be two curves (or polynomials) through the origin. Then $\mu_0(F,G) = 1$ if and only if the linear parts F_1 and G_1 are linearly independent.*

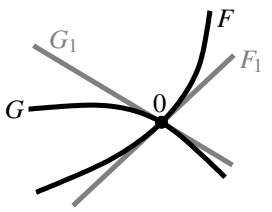
Proof. We prove the statement following Algorithm 2.12, using the notation from there.

By assumption F and G pass through the origin, so we are not in case (a) of the algorithm. In case (b), note that F_1' and G_1 are linearly independent if and only if F_1 and G_1 are, as either $F_1' = F_1$ (if $m > n$) or $F_1' = F_1 - \frac{a}{b}G_1$ (if $m = n$). Hence we can consider the first time we reach case (c). As $\mu_0(y, G) > 0$ we then have

$$\begin{aligned}
\mu_0(F,G) = 1 &\Leftrightarrow \mu_0(y,G) = 1 \text{ and } \mu(F',G) = 0 \\
&\Leftrightarrow G \text{ contains a monomial } x^1y^0 \text{ and } F' \text{ contains a constant term} \\
&\qquad \text{(by Example 2.11 and Lemma 2.5 (a))} \\
&\Leftrightarrow G_1 = ax + by \text{ for some } a \in K^*, b \in K, \text{ and } F_1 = cy \text{ for some } c \in K^* \\
&\Leftrightarrow F_1 \text{ and } G_1 \text{ are linearly independent,}
\end{aligned}$$

where the last implication "$\Leftarrow$" follows since $F = yF'$ clearly does not contain a monomial x^1y^0. □

In fact, Proposition 2.17 has an easy geometric interpretation in the spirit of the beginning of this chapter: F_1 and G_1 can be thought of as the linear approximations of F and G around the origin. If these approximations are non-zero, hence lines, they can be thought of as the tangents to the curves as in the picture on the right, and the proposition states that the intersection multiplicity is 1 if and only if these tangent directions are not the same.



In general, it is the lowest non-zero terms of a curve F that can be considered as the best local approximation of F around 0. We can use this idea to define tangents to arbitrary curves (i. e. even if the linear approximation F_1 vanishes) as follows.

Definition 2.18 (Tangents and multiplicities of points)**.** Let F be a curve.

(a) The smallest $m \in \mathbb{N}$ for which the homogeneous part F_m is non-zero is called the **multiplicity** $m_0(F)$ of F at the origin. Any linear factor of F_m (considered as a curve) is called a **tangent** to F at the origin.

(b) For a general point $P = (x_0, y_0) \in \mathbb{A}^2$, tangents at P and the multiplicity $m_P(F)$ are defined by first shifting coordinates to $x' = x - x_0$ and $y' = y - y_0$, and then applying (a) to the origin $(x', y') = (0, 0)$.

Exercise 2.19. Given a linear coordinate transformation that maps the origin to itself and a curve F to F', show that $m_0(F) = m_0(F')$, and that the transformation maps any tangent of F to a tangent of F'. In particular, despite its appearance, Definition 2.18 is independent of the choice of coordinates on $\mathbb{A}^2$.

By definition, we clearly have $m_P(F) > 0$ if and only if $P \in F$. The most important case of Definition 2.18 is then $m_P(F) = 1$, i. e. if there is a non-zero local linear approximation for F around P. There is a special terminology for this case.

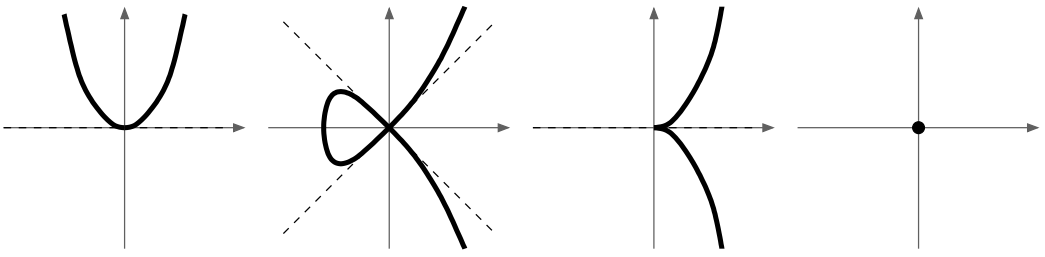
Definition 2.20 (Smooth and singular points)**.** Let F be a curve.

(a) A point $P \in F$ is called **smooth** or **regular** if $m_P(F) = 1$. Note that F has then a unique tangent at P, which we will denote by T_PF. For $P = 0$, it is simply given by the linear part F_1 of F.

If P is not a smooth point, i. e. if $m_P(F) > 1$, we say that P is a **singular** point or a **singularity** of F. As a special case, a singularity with $m_P(F) = 2$ such that F has (exactly) two different tangents there is called a **node**.

(b) The curve F is said to be **smooth** or **regular** if all its points are smooth. Otherwise, F is called **singular**.

Example 2.21. Let us consider the origin in the real curves in the following picture.



For the case (a), the curve $F = y - x^2$ in (a) has (no constant but) a linear term y. Hence, we have $m_0(F) = 1$, the origin is a smooth point of the curve, and its tangent there is $T_0F = y$.

For the other three curves, the origin is a singular point of multiplicity 2. In (b), this singularity is a node, since the quadratic term is $y^2 - x^2 = (y-x)(y+x)$, and thus we have the two tangents $y-x$ and $y+x$, shown as dashed lines in the picture. The curve in (c) has only one tangent y which is of multiplicity 2. Finally, in (d) there is no tangent at all since x^2+y^2 does not contain a linear factor over $\mathbb{R}$. Note that, in any case, knowing the tangents of F at the origin (which are easy to compute) tells us to some extent what the curve looks like locally around 0.

With these notations we can now reformulate Proposition 2.17.

Corollary 2.22 (Transverse intersections)**.** *Let P be a point in the intersection of two curves F and G. Then $\mu_P(F,G) = 1$ if and only if P is a smooth point of both F and G, and $T_PF \neq T_PG$.*

*We say in this case that F and G intersect **transversely** at P.*

Remark 2.23 (Additivity of point multiplicities)**.** Note that $m_P(FG) = m_P(F) + m_P(G)$. Hence, any point that lies on at least two (not necessarily distinct) irreducible components has multiplicity at least 2, and is thus a singular point. In particular, all points on a component of multiplicity at least 2 (in the sense of Definition 1.5 (c)) are always singular.

To check if a given curve F is smooth, i. e. whether every point $P \in F$ is a smooth point of F, there is a simple criterion that does not require to shift P to the origin first. It uses the (partial) derivatives $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ of F, which can be defined purely formally over an arbitrary ground field and then satisfy the usual rules of differentiation [G1, Exercise 9.10].

Proposition 2.24 (Affine Jacobi Criterion)**.** *Let $P = (x_0, y_0)$ be a point on an affine curve F.*

(a) *P is a singular point of F if and only if $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = 0$.*

(b) *If P is a smooth point of F the tangent to F at P is given by*

$$T_PF = \frac{\partial F}{\partial x}(P) \cdot (x - x_0) + \frac{\partial F}{\partial y}(P) \cdot (y - y_0).$$

Proof. Substituting $x = x' + x_0$ and $y = y' + y_0$, i. e. $x' = x - x_0$ and $y' = y - y_0$, we can consider F as a polynomial in x' and y'. If we expand

$$F = ax' + by' + (\text{higher order terms in } x' \text{ and } y'),$$

then by definition F is singular at $(x', y') = (0,0)$, i. e. at P, if and only if $a = b = 0$. But by the chain rule of differentiation we have

$$a = \frac{\partial F}{\partial x'}(0) = \frac{\partial F}{\partial x}(P) \quad \text{and} \quad b = \frac{\partial F}{\partial y'}(0) = \frac{\partial F}{\partial y}(P),$$

so that (a) follows. Moreover, if F is smooth at P then its tangent is just the term of F linear in x' and y', i. e.

$$ax' + by' = \frac{\partial F}{\partial x}(P) \cdot (x - x_0) + \frac{\partial F}{\partial y}(P) \cdot (y - y_0),$$

as claimed in (b). □

Example 2.25. Consider again the real curve $F = y^2 - x^2 - x^3$ from Example 2.21 (b). To determine its singular points, we compute the partial derivatives

$$\frac{\partial F}{\partial x} = -2x - 3x^2 \quad \text{and} \quad \frac{\partial F}{\partial y} = 2y.$$

Its common zeros are $(0,0)$ and $(-\frac{2}{3},0)$. But the latter does not lie on the curve, and so we conclude that the origin is the only singular point of F.

Smoothness of a curve F at a point P has another important algebraic consequence: It means that the containment of ideals containing F in $\mathcal{O}_P$ (or in other words of ideals in $\mathcal{O}_P/\langle F\rangle$) can be checked by a simple comparison of intersection multiplicities.

Proposition 2.26 (Comparing ideals using intersection multiplicities)**.** *Let P be a smooth point on a curve F. Then for any two curves G and H that do not have a common component with F through P we have*

$$\langle F,G\rangle \subset \langle F,H\rangle \text{ in } \mathcal{O}_P \quad \Leftrightarrow \quad \mu_P(F,G) \geq \mu_P(F,H).$$

In particular, we have $\langle F,G\rangle = \langle F,H\rangle$ in $\mathcal{O}_P$ if and only if $\mu_P(F,G) = \mu_P(F,H)$. 03

Proof.

"$\Rightarrow$": Clearly, if $\langle F,G\rangle \subset \langle F,H\rangle$ then $\mu_P(F,G) = \dim \mathcal{O}_P/\langle F,G\rangle \geq \dim \mathcal{O}_P/\langle F,H\rangle = \mu_P(F,H)$.

"$\Leftarrow$": Let L be a line through P which is not the tangent T_PF. Then $\mu_P(F,L) = 1$ by Corollary 2.22, and hence $\mu_P(F,L^n) = n$ for all $n \in \mathbb{N}$ by Proposition 2.10. Let n be the maximum number such that $\langle F,G\rangle \subset \langle F,L^n\rangle$ in $\mathcal{O}_P$ (this exists since $\langle F,G\rangle \subset \mathcal{O}_P = \langle F,L^0\rangle$, and $\langle F,G\rangle \subset \langle F,L^n\rangle$ requires $n \leq \mu_P(F,G)$ by the direction "$\Rightarrow$" that we have already shown).

We claim that then $\langle F,G\rangle = \langle F,L^n\rangle$ in $\mathcal{O}_P$, i. e. that $L^n \in \langle F,G\rangle$. To see this, note that $\langle F,G\rangle \subset \langle F,L^n\rangle$ implies $G = aF + bL^n$ for some $a,b \in \mathcal{O}_P$. If we had $b(P) = 0$ it would follow that $b \in I_P = \langle F,L\rangle$ by Lemma 2.5 (b), i. e. $b = cF + dL$ for some $c,d \in \mathcal{O}_P$, which means that $G = aF + (cF + dL)L^n \in \langle F,L^{n+1}\rangle$ and thus contradicts the maximality of n. Hence $b(P) \neq 0$, i. e. b is a unit in $\mathcal{O}_P$, and we obtain $L^n = \frac{1}{b}(G - aF) \in \langle F,G\rangle$ as desired.

Of course, now $\langle F,G\rangle = \langle F,L^n\rangle$ implies that $\mu_P(F,G) = \mu_P(F,L^n) = n$, so that we obtain $\langle F,G\rangle = \langle F,L^{\mu_P(F,G)}\rangle$. But the same holds for H instead of G, and so the inequality $\mu_P(F,G) \geq \mu_P(F,H)$ yields

$$\langle F,G\rangle = \langle F,L^{\mu_P(F,G)}\rangle \subset \langle F,L^{\mu_P(F,H)}\rangle = \langle F,H\rangle. \qquad \square$$

Example 2.27. Proposition 2.26 is false without the smoothness assumption on F: For the real curve $F = x^2 - y^2 = (x-y)(x+y)$ (i. e. the union of the two diagonals in $\mathbb{A}^2$, with singular point 0), $G = x$, and $H = y$, we have $\langle F,G\rangle = \langle x,y^2\rangle$ and $\langle F,H\rangle = \langle y,x^2\rangle$. Hence $\mu_0(F,G) = \mu_0(F,H) = 2$, but $\langle F,G\rangle \neq \langle F,H\rangle$ (since $y \notin \langle x,y^2\rangle$, as otherwise we would have $\langle x,y^2\rangle = \langle x,y\rangle$, in contradiction to $\mu_0(x,y^2) = 2 \neq 1 = \mu_0(x,y)$).

Remark 2.28 (Geometric interpretation of smooth curves)**.** Mainly for the ground field $K = \mathbb{R}$, our results on smooth curves have an intuitive interpretation:

(a) The Jacobi Criterion of Proposition 2.24 (a) states that P is a smooth point of a real curve F if and only if the Implicit Function Theorem [G2, Proposition 27.10] can be applied to the equation $F = 0$ around P, so that $V(F)$ is a 1-dimensional submanifold of $\mathbb{R}^2$ [G2, Definition 27.18]. Hence, in this case $V(F)$ is locally the graph of a differentiable function (expressing y as a function of x or vice versa), and thus we arrive at the intuitive interpretation of smoothness as "having no sharp corners".

(b) To interpret Proposition 2.26, let us continue the picture of (a) and consider a local (analytic) coordinate z around P on the 1-dimensional manifold $V(F)$. In accordance with the idea of intersection multiplicity at the beginning of this chapter, a curve G should have intersection multiplicity n with F at P if on F it is locally a function of the form az^n in this coordinate, with a a non-zero function at P (corresponding to a unit in $\mathcal{O}_P$). Now if $n = \mu_P(F,G) \geq$

$\mu_P(F,H) = m$ then in the same way H is of the form bz^m for a function b non-zero at P, so that $bz^m = H$ divides $az^n = G$. This means that $\langle G \rangle \subset \langle H \rangle$ in $\mathcal{O}_P/\langle F \rangle$ (i. e. as functions on F, a point of view that we will discuss in detail starting in Chapter 6) and thus that $\langle F,G \rangle \subset \langle F,H \rangle$ in $\mathcal{O}_P$.

(c) In fact, the analytic idea of (b) has a direct counterpart in commutative algebra that can then be applied over arbitrary ground fields: For a smooth curve F the ring $\mathcal{O}_P/\langle F \rangle$ is a so-called *discrete valuation ring* [G6, Chapter 12]. This means that the non-zero elements of this ring have a valuation – a natural number that can be interpreted as the order of the zero as a function on F, and hence as the local intersection multiplicity with F. It is then a result in commutative algebra that the non-zero ideals in a discrete valuation ring are in one-to-one correspondence with these valuations as above [G6, Corollary 12.17]. This is precisely the statement of Proposition 2.26.

Exercise 2.29 (Cusps)**.** Let P be a point on an affine curve F. We say that P is a **cusp** if $m_P(F) = 2$, there is exactly one tangent L to F at P, and $\mu_P(F,L) = 3$.

(a) Give an example of a real curve with a cusp, and draw a picture of it.

(b) If F has a cusp at P, prove that F has only one irreducible component passing through P.

(c) If F and G have a cusp at P, what is the minimum possible value for the intersection multiplicity $\mu_P(F,G)$?

Exercise 2.30.

(a) Find all singular points of the curve $F = (x^2+y^2-1)^3 + 10x^2y^2 \in \mathbb{R}[x,y]$, and determine the multiplicities and tangents to F at these points.

(b) Show that an irreducible curve F over a field of characteristic 0 has only finitely many singular points.

Can you find weaker assumptions on F that also imply that F has only finitely many singular points?

(c) Show that an irreducible cubic can have at most one singular point, and that over an algebraically closed field this singularity must be a node or a cusp as in Exercise 2.29.

3. Projective Curves

In the last chapter we have studied the local intersection behavior of curves. Our next major goal will be to consider the global situation and ask how many intersection points two curves can have in total, i. e. how many common zeros we find for two polynomials $F, G \in K[x,y]$ (where we will count each such zero with its intersection multiplicity).

For polynomials in one variable, the corresponding question would simply be how many zeros a single polynomial $f \in K[x]$ has. At least if K is algebraically closed, so that f is a product of linear factors, the answer is then of course that we always get $\deg f$ zeros (counted with multiplicities). Hence, in our current case of two polynomials $F, G \in K[x,y]$ we would also hope for a result that depends only on $\deg F$ and $\deg G$, and not on the chosen polynomials.

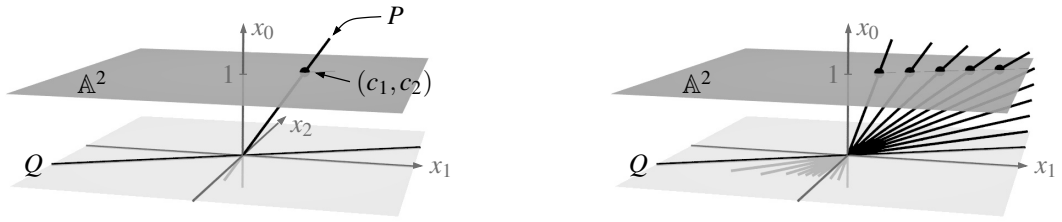
However, even in the simplest case when F and G are two distinct lines this will not work, since F and G might intersect in one point or be parallel (and hence have no intersection point). To fix this situation, the geometric idea is to add points at infinity to the affine plane $\mathbb{A}^2$, so that two lines that are parallel in $\mathbb{A}^2$ will meet there. On the other hand, two non-parallel lines (that intersect already in $\mathbb{A}^2$) should not meet at infinity any more as this would then lead to two intersection points. Hence, we have to add one point at infinity for each direction in the affine plane, so that parallel lines with the same direction meet there, whereas others do not.

This new space with the added points at infinity will be called the *projective plane*. In the case $K = \mathbb{R}$ we can also think of it as a compactification of the affine plane $\mathbb{A}^2$. It is the goal of this chapter to study this process in detail, leading to plane curves that are "compactified" by points at infinity. For two such compactified curves we will then compute the number of intersection points in the next chapter, and the answer will then indeed depend only on the degrees of the curves.

Remark 3.1 (Geometric idea of projective spaces)**.** Algebraically, the idea for adding points at infinity is to embed the affine space $\mathbb{A}^n$ in the vector space K^{n+1} by prepending a new coordinate (typically called x_0) equal to 1, i. e. by the map

$$\mathbb{A}^n \to K^{n+1},\ (x_1,\dots,x_n) \mapsto (1,x_1,\dots,x_n),$$

and considering the 1-dimensional linear subspace in K^{n+1} spanned by this vector. For example, in this way a point $(c_1,c_2) \in \mathbb{A}^2$ corresponds to the line through the origin and $(1,c_1,c_2) \in K^3$, denoted by P in the picture below on the left.



We will define the projective plane as the set of all such 1-dimensional linear subspaces of K^3. It then consists of all lines through the origin coming from points of $\mathbb{A}^2$ as above – together with lines contained in the plane where $x_0 = 0$ that do not arise in this way, such as Q in the picture above. As shown on the right, these lines can be thought of as limits of lines coming from an unbounded sequence of points in $\mathbb{A}^2$. They can therefore be interpreted as the "points at infinity" that we were looking for.

Let us now turn this idea into a precise definition.

Definition 3.2 (Projective spaces)**.** For $n \in \mathbb{N}$, we define the **projective n-space** over K as the set of all 1-dimensional linear subspaces of K^{n+1}. It is denoted by $\mathbb{P}^n_K$ or simply $\mathbb{P}^n$.

Notation 3.3 (Homogeneous coordinates)**.** Obviously, a 1-dimensional linear subspace of K^{n+1} is uniquely determined by a spanning non-zero vector in K^{n+1}, with two such vectors giving the same linear subspace if and only if they are scalar multiples of each other. In other words, we have

$$\mathbb{P}^n = (K^{n+1} \backslash \{0\}) / \sim$$

with the equivalence relation

$$(x_0, \dots, x_n) \sim (y_0, \dots, y_n) \quad :\Leftrightarrow \quad x_i = \lambda y_i \text{ for some } \lambda \in K^* \text{ and all } i.$$

The equivalence class of $(x_0, \dots, x_n)$ is usually denoted by $(x_0 : \dots : x_n) \in \mathbb{P}^n$. We call $x_0, \dots, x_n$ the **homogeneous** or **projective coordinates** of the point $(x_0 : \dots : x_n)$. Hence, in this notation for a point in $\mathbb{P}^n$ the numbers $x_0, \dots, x_n$ are not all zero, and they are defined only up to a common scalar multiple.

Remark 3.4 (Geometric interpretation of $\mathbb{P}^n$)**.** There are two ways to interpret the projective space $\mathbb{P}^n$ geometrically:

(a) As in Remark 3.1, we can embed the affine space $\mathbb{A}^n$ in $\mathbb{P}^n$ by the map

$$\mathbb{A}^n \to \mathbb{P}^n, \ (x_1, \dots, x_n) \mapsto (1 : x_1 : \dots : x_n)$$

whose image is the subset $U_0 := \{(x_0 : \dots : x_n) : x_0 \neq 0\}$ of $\mathbb{P}^n$. We will often consider $\mathbb{A}^n$ as a subset of $\mathbb{P}^n$ in this way, i. e. by setting $x_0 = 1$. The other coordinates $x_1, \dots, x_n$ are then called the **inhomogeneous** or **affine coordinates** on U_0.

The remaining points of $\mathbb{P}^n$ are of the form $(0 : x_1 : \dots : x_n)$. By forgetting their coordinate x_0 (which is zero anyway) they form a set that is naturally bijective to $\mathbb{P}^{n-1}$, corresponding to the 1-dimensional linear subspaces of K^n. As in Remark 3.1 we can regard them as *points at infinity*; there is hence one such point for each direction in K^n. In short-hand notation, one often writes this decomposition as $\mathbb{P}^n = \mathbb{A}^n \cup \mathbb{P}^{n-1}$ and calls $\mathbb{A}^n$ and $\mathbb{P}^{n-1}$ the *affine* and *infinite part* of $\mathbb{P}^n$, respectively.

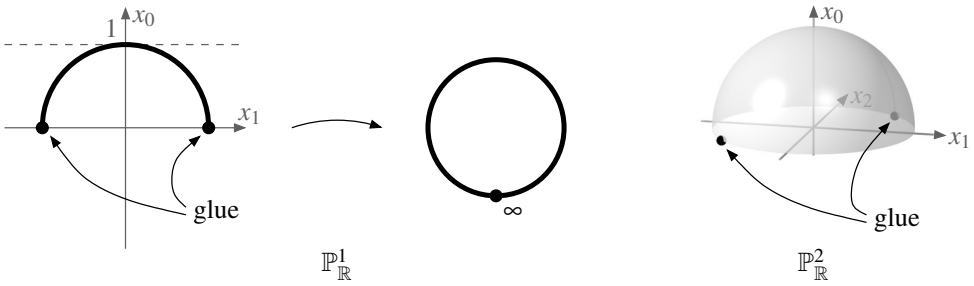
(b) By the symmetry of the homogeneous coordinates, the subsets $U_i := \{(x_0 : \dots : x_n) : x_i \neq 0\}$ of $\mathbb{P}^n$ are naturally bijective to $\mathbb{A}^n$ for all $i = 0, \dots, n$, in the same way as for $i = 0$ in (a). As every point of $\mathbb{P}^n$ has at least one non-zero coordinate, it lies in one of the U_i, and hence in a subset of $\mathbb{P}^n$ that just looks like the ordinary affine space $\mathbb{A}^n$. In this sense we can say that projective space "looks everywhere the same"; the fact that we interpreted the points with $x_0 = 0$ as points as infinity above was just due to our special choice of $i = 0$ in (a).

Example 3.5. By Remark 3.4 (a), we have $\mathbb{P}^1 = \mathbb{A}^1 \cup \mathbb{P}^0$. The affine part consists of the points $(1 : x_1)$ for $x_1 \in K$, and the infinite part contains the single point $(0 : 1)$. Denoting this point at infinity by ∞, we can therefore write $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$.

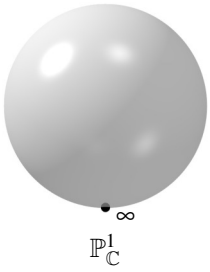
Remark 3.6 (Topology of projective spaces over $\mathbb{R}$ and $\mathbb{C}$)**.** Over the real or complex numbers, every point in $\mathbb{P}^n$ has a representative on the unit sphere $\{(x_0, \dots, x_n) : |x_0|^2 + \dots + |x_n|^2 = 1\}$ by normalizing. In other words, $\mathbb{P}^n$ can be written as the image of this compact unit sphere under the quotient map $(x_0, \dots, x_n) \mapsto (x_0 : \dots : x_n)$. In accordance with our motivation at the beginning of this chapter, this means that $\mathbb{P}^n$ is itself compact (with the quotient topology [G5, Definition 5.3 and Corollary 5.8 (c)]).

(a) For $K = \mathbb{R}$, every 1-dimensional linear subspace of K^{n+1} meets the unit sphere in exactly two points, which are negatives of each other. Hence, all points $(x_0 : \dots : x_n) \in \mathbb{P}^n$ have a representative on the upper half of the unit sphere, i. e. where $x_0 \geq 0$, and this representative is unique except for points on its boundary where $x_0 = 0$ (i. e. for points at infinity). As in the following picture, we can therefore visualize $\mathbb{P}^n_{\mathbb{R}}$ as the space obtained from the upper half unit sphere by identifying opposite points on the boundary. For $n = 1$ we have only one pair of gluing points, corresponding to one point at infinity as in Example 3.5, and obtain

topologically a circle. For $n = 2$, each point on the boundary of the upper half unit sphere has to be identified with its negative, which leads to a space that cannot be embedded in $\mathbb{R}^3$.



(b) For $K = \mathbb{C}$, only $\mathbb{P}^1_{\mathbb{C}}$ can be visualized in $\mathbb{R}^3$. By Example 3.5 it is just the complex plane together with a point ∞. It is therefore topologically a sphere as in the picture on the right.



Having studied projective spaces, we now want to consider subsets of $\mathbb{P}^n$ given by polynomial equations. However, polynomials in homogeneous coordinates are not well-defined functions on $\mathbb{P}^n$: For example, for the polynomial $f = x_0^2 + x_1$ we have $f(1,-1) = 0$ and $f(-1,1) = 2$ although $(1:-1) = (-1:1) \in \mathbb{P}^1$. We can solve this problem by using homogeneous polynomials as follows.

Remark 3.7. Let

$$f = \sum_{i_0+\cdots+i_n=d} a_{i_0,\ldots,i_n}\, x_0^{i_0} \cdot \cdots \cdot x_n^{i_n} \quad \in K[x_0,\ldots,x_n]$$

be a homogeneous polynomial of degree d. Then

$$f(\lambda x_0,\ldots,\lambda x_n) = \sum_{i_0+\cdots+i_n=d} a_{i_0,\ldots,i_n}\, \lambda^{i_0+\cdots+i_n} x_0^{i_0} \cdot \cdots \cdot x_n^{i_n} = \lambda^d f(x_0,\ldots,x_n)$$

for all $\lambda \in K$. In particular, we see:

(a) Although f is not a well-defined function on $\mathbb{P}^n$, its zero locus is well-defined on $\mathbb{P}^n$, i. e. we have

$$f(\lambda x_0,\ldots,\lambda x_n) = 0 \quad \Leftrightarrow \quad f(x_0,\ldots,x_n) = 0$$

for all $\lambda \in K^*$. In the following, we will therefore write this condition simply as $f(P) = 0$ for $P = (x_0 : \cdots : x_n)$.

(b) If g is another homogeneous polynomial of degree d then

$$\frac{f(\lambda x_0,\ldots,\lambda x_n)}{g(\lambda x_0,\ldots,\lambda x_n)} = \frac{\lambda^d f(x_0,\ldots,x_n)}{\lambda^d g(x_0,\ldots,x_n)} = \frac{f(x_0,\ldots,x_n)}{g(x_0,\ldots,x_n)},$$

and so the quotient $\frac{f}{g}$ is a well-defined function on the subset of $\mathbb{P}^n$ where g does not vanish.

Definition 3.8 (Projective varieties)**.** For a subset $S \subset K[x_0,\ldots,x_n]$ of homogeneous polynomials we call

$$V(S) := \{P \in \mathbb{P}^n : f(P) = 0 \text{ for all } f \in S\} \quad \subset \mathbb{P}^n$$

the (projective) **zero locus** of S. Subsets of $\mathbb{P}^n$ that are of this form are called **(projective) varieties**. If $S = \{f_1,\ldots,f_k\}$ is a finite set, we will write $V(S) = V(\{f_1,\ldots,f_k\})$ also as $V(f_1,\ldots,f_k)$. To distinguish the projective from the affine zero locus of Definition 1.3 (b), we will sometimes denote it by $V_{\mathrm{p}}(S) \subset \mathbb{P}^n$ as opposed to $V_{\mathrm{a}}(S) \subset \mathbb{A}^{n+1}$.

In this class we will mostly restrict ourselves to the case of the projective plane $\mathbb{P}^2$. We will then usually denote the homogeneous coordinates by x, y, and z, with z corresponding to the variable x_0 defining the points at infinity as in Remark 3.4 (a).

Remark 3.9. The properties of Remark 1.4 hold analogously for the projective zero locus: For any two homogeneous polynomials $f, g \in K[x,y,z]$ we have

(a) $V(f) \cup V(g) = V(fg)$;
(b) $V(f) \cap V(g) = V(f,g)$.

Exercise 3.10. By a *projective coordinate transformation* we mean a map $f\colon \mathbb{P}^n \to \mathbb{P}^n$ of the form

$$(x_0 : \cdots : x_n) \mapsto (f_0(x_0,\ldots,x_n) : \cdots : f_n(x_0,\ldots,x_n))$$

for linearly independent homogeneous linear polynomials $f_0,\ldots,f_n \in K[x_0,\ldots,x_n]$.

Now let $P_1,\ldots,P_{n+2} \in \mathbb{P}^n$ be points such that any $n+1$ of them are linearly independent in K^{n+1}, and in the same way let $Q_1,\ldots,Q_{n+2} \in \mathbb{P}^n$ be points such that any $n+1$ of them are linearly independent. Show that there is a projective coordinate transformation f with $f(P_i) = Q_i$ for all $i = 1,\ldots,n+2$.

Exercise 3.11. Show:

(a) If $F, G \in K[x_0,\ldots,x_n]$ are polynomials such that $F \mid G$ and G is homogeneous, then F is homogeneous.
(b) Every homogeneous polynomial in two variables over an algebraically closed field is a product of linear polynomials.

04

The definition of projective plane curves is now completely analogous to the affine case in Definition 1.5.

Definition 3.12 (Projective curves)**.**

(a) A **(projective plane algebraic) curve** (over K) is a non-constant homogeneous polynomial $F \in K[x,y,z]$ modulo units. We call $V(F) = \{P \in \mathbb{P}^2 : F(P) = 0\}$ its **set of points**.
(b) The **degree** of a projective curve is its degree as a polynomial. As in the affine case, curves of degree $1,2,3,\ldots$ are called **lines**, **quadrics/conics**, **cubics**, and so on. The line z is referred to as the **line at infinity**.
(c) The notions of irreducible/reducible/reduced curves, as well as of irreducible components and their multiplicities, are defined in the same way as for affine curves in Definition 1.5 (c) (note that irreducible factors of homogeneous polynomials are always homogeneous by Exercise 3.11 (a)).

To study projective curves, we will often want to relate them to affine curves. For this we need the following construction.

Construction 3.13 (Homogenization and dehomogenization)**.**

(a) For a non-zero polynomial

$$f = \sum_{i+j\le d} a_{i,j}\, x^i y^j \quad \in K[x,y]$$

of degree d we define the **homogenization** of f as

$$f^{\mathrm{h}} := \sum_{i+j\le d} a_{i,j}\, x^i y^j z^{d-i-j} \quad \in K[x,y,z].$$

Note that f^{h} is homogeneous of degree $\deg f^{\mathrm{h}} = \deg f = d$, and that $z \nmid f^{\mathrm{h}}$ since f contains a term with $i+j=d$.

(b) For a non-zero homogeneous polynomial

$$f = \sum_{i+j+k=d} a_{i,j,k}\, x^i y^j z^k \quad \in K[x,y,z]$$

of degree d we define the **dehomogenization** of f to be

$$f^{\mathrm{i}} := f(z=1) = \sum_{i+j+k=d} a_{i,j,k}\, x^i y^j \quad \in K[x,y].$$

In general, f^{i} will be an inhomogeneous polynomial. If $z \nmid f$, i. e. if f contains a monomial without z, then this monomial will also be present in f^{i}, and thus $\deg f^{\mathrm{i}} = \deg f = d$.

In particular, there is a bijective correspondence

$$\left\{\begin{array}{c}\text{polynomials of degree } d \\ \text{in } K[x,y]\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{homogeneous polynomials of degree } d \\ \text{in } K[x,y,z] \text{ not divisible by } z\end{array}\right\}$$
$$\begin{array}{ccc} f & \longmapsto & f^{\mathrm{h}} \\ f^{\mathrm{i}} & \longmapsfrom & f. \end{array}$$

Example 3.14. For $f = y - x^2 \in K[x,y]$ we have $f^{\mathrm{h}} = yz - x^2 \in K[x,y,z]$, and then back again $(f^{\mathrm{h}})^{\mathrm{i}} = y - x^2 = f$.

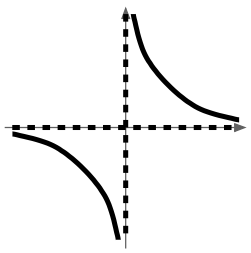
Construction 3.15 (Affine parts and projective closures)**.**

(a) For a projective curve F its affine set of points is $V_{\mathrm{p}}(F) \cap \mathbb{A}^2 = V_{\mathrm{a}}(F(z=1)) = V_{\mathrm{a}}(F^{\mathrm{i}})$. We will therefore call F^{i} the **affine part** of F. The points at infinity of F are given by $V_{\mathrm{p}}(F(z=0)) \subset \mathbb{P}^1$.

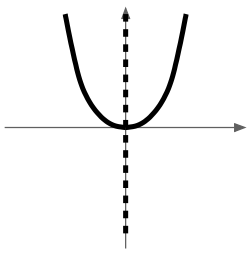
(b) For an affine curve F we call F^{h} its **projective closure**. By Construction 3.13 it is a projective curve whose affine part is again F, and that does not contain the line at infinity as a component.

However, F^{h} may contain points at infinity: If $F = F_0 + \cdots + F_d$ is the decomposition into homogeneous parts as in Notation 2.16, we have $F^{\mathrm{h}} = z^d F_0 + z^{d-1} F_1 + \cdots + F_d$ and hence $F^{\mathrm{h}}(z=0) = F_d$. So the points at infinity of F are given by the projective zero locus of the leading part of F.

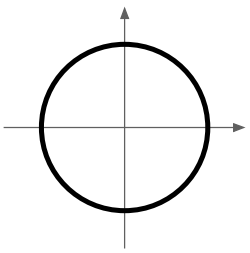
Example 3.16 (Visualization of projective curves)**.** To visualize a (real) projective curve F (that does not have the line at infinity as a component), we will often just draw its affine set of points $V_{\mathrm{a}}(F^{\mathrm{i}})$, and if desired in addition its points at infinity as directions in $\mathbb{A}^2$. The following picture shows in this way the projective closures of the three types of real conics – a hyperbola, a parabola, and an ellipse (resp. a circle) – where the dashed lines correspond to the points at infinity. We see that the hyperbola has two points at infinity (namely $(0:1:0)$ and $(1:0:0)$ in the case below), the parabola has one ($(0:1:0)$ below), and the circle no such point. Note that, including these additional points, all three cases become topologically a loop, as the unbounded ends of the affine curves meet up at the corresponding points at infinity. In fact, up to a change of coordinates, we will see in Exercise 3.28 that there is essentially only one type of real projective conic.



(a) hyperbola
$F = xy - 1$
$F^{\mathrm{h}} = xy - z^2$
points at infinity: $xy = 0$



(b) parabola
$F = y - x^2$
$F^{\mathrm{h}} = yz - x^2$
points at infinity: $x^2 = 0$



(c) ellipse
$F = x^2 + y^2 - 1$
$F^{\mathrm{h}} = x^2 + y^2 - z^2$
points at infinity: $x^2 + y^2 = 0$

Remark 3.17 (Spaces of curves as projective spaces)**.** For $d \in \mathbb{N}_{>0}$, the vector space of homogeneous polynomials of degree d in $K[x,y,z]$ has dimension $\binom{d+2}{2}$, hence it is isomorphic to K^{n+1} with $n = \binom{d+2}{2} - 1$. By definition, a projective curve of degree d is then a non-zero point of this vector space modulo scalars. Hence, the space of all such curves is just the projective space $\mathbb{P}^n$, and thus itself a projective variety.

It is in fact very special to algebraic geometry – and very powerful – that the spaces of (certain) varieties are again varieties, and thus can be studied with exactly the same methods as the initial objects themselves. In other categories this is usually far from being true: The space of all groups is not a group, the space of all vector spaces is not a vector space, the space of all topological spaces is not a topological space, and so on.

For the rest of this chapter, let us transfer our results on affine curves from Chapters 1 and 2 to the projective case.

Remark 3.18 (Finiteness of zero loci)**.** Let F and G be two projective curves. The finiteness results of Lemma 1.11 and Proposition 1.12 (b) hold for the affine parts of F and G (for any choice of coordinate determining the line at infinity), and thus for F and G themselves: $V(F)$ is infinite if K is algebraically closed, $\mathbb{P}^2 \backslash V(F)$ is infinite if K is infinite, and $V(F,G)$ is finite if F and G have no common component.

Remark 3.19 (Recovering F from $V(F)$)**.** Let F be a projective curve over an algebraically closed field. We can write it as $F = z^m G$ for some $m \in \mathbb{N}$ and a curve G with $z \nmid G$. Then G can be recovered from G^{i} since $G = (G^{\mathrm{i}})^{\mathrm{h}}$ by Construction 3.13, and G^{i} can be recovered from $V_{\mathrm{a}}(G^{\mathrm{i}}) = V_{\mathrm{p}}(G) \cap \mathbb{A}^2$ and a multiplicity on each component by Remark 1.14.

As the components of F are just the components of G plus possibly the line at infinity z (with multiplicity m), this means that F can be reconstructed from $V(F)$ and a multiplicity on each component, just as in the affine case.

Construction 3.20 (Local rings of $\mathbb{P}^2$)**.** For $P \in \mathbb{P}^2$ we define the **local ring** of $\mathbb{P}^2$ at P according to Remark 3.7 (b) as

$$\mathscr{O}_P := \mathscr{O}_{\mathbb{P}^2,P} := \left\{ \frac{f}{g} : f,g \in K[x,y,z] \text{ homogeneous of the same degree with } g(P) \neq 0 \right\} \cup \{0\} \subset K(x,y,z).$$

As in Definition 2.1, these rings admit a well-defined **evaluation map**

$$\mathscr{O}_P \to K, \ \frac{f}{g} \mapsto \frac{f(P)}{g(P)}$$

with kernel

$$I_P := I_{\mathbb{P}^2,P} := \left\{ \frac{f}{g} \in \mathscr{O}_P : f(P) = 0 \right\} \subset \mathscr{O}_P.$$

For a point $P = (x_0 : y_0 : 1)$ in the affine part of $\mathbb{P}^2$ it is easily checked that there is an isomorphism

$$\mathscr{O}_{\mathbb{P}^2,(x_0:y_0:1)} \to \mathscr{O}_{\mathbb{A}^2,(x_0,y_0)}, \ \frac{f}{g} \mapsto \frac{f^{\mathrm{i}}}{g^{\mathrm{i}}}$$

compatible with the evaluation maps, and thus taking $I_{\mathbb{P}^2,(x_0:y_0:1)}$ to $I_{\mathbb{A}^2,(x_0,y_0)}$. Hence the local rings are still the same as in the affine case – which is of course expected, as objects that are local around a point in $\mathbb{A}^2$ should not be affected by adding points at infinity.

Construction 3.21 (Intersection multiplicities)**.** Note that homogeneous polynomials are not elements of the local ring $\mathscr{O}_{\mathbb{P}^2,P}$. But for $F_1,\dots,F_k$ homogeneous we can still define a generated ideal

$$\langle F_1,\dots,F_k \rangle = \left\{ \frac{f_1}{g_1}F_1 + \dots + \frac{f_k}{g_k}F_k : f_i = 0 \text{ or } f_i,g_i \in K[x,y,z] \text{ homogeneous with } g_i(P) \neq 0 \text{ and } \deg(f_iF_i) = \deg g_i \text{ for all } i \right\}$$

in $\mathscr{O}_P$. As in the affine case we can therefore define the **intersection multiplicity** of two curves F,G at a point $P \in \mathbb{P}^2$ as

$$\mu_P(F,G) := \dim \mathscr{O}_P/\langle F,G \rangle \quad \in \mathbb{N} \cup \{\infty\}. \tag{$*$}$$

For a point $P = (x_0 : y_0 : 1)$ in the affine part of $\mathbb{P}^2$ one can verify directly that the isomorphism $\mathscr{O}_{\mathbb{P}^2,(x_0:y_0:1)} \cong \mathscr{O}_{\mathbb{A}^2,(x_0,y_0)}$ of Construction 3.20 takes $\langle F,G \rangle$ to $\langle F^{\mathrm{i}},G^{\mathrm{i}} \rangle$. Hence we have

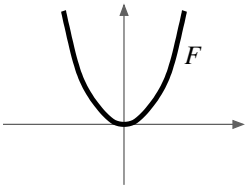
$\mu_{(x_0:y_0:1)}(F,G) = \mu_{(x_0,y_0)}(F^{\mathrm{i}},G^{\mathrm{i}})$, i. e. intersection multiplicities in the affine part can be computed exactly as in Chapter 2. At other points, the multiplicity can be computed similarly by choosing another (non-zero) coordinate to define the line at infinity as in Remark 3.4 (b). We will therefore probably never use the global definition $(*)$ of the multiplicity above for actual computations; its only purpose is to ensure that the result does not depend on the choice of coordinate defining the line at infinity.

Moreover, in the same way as in Remark 2.4 (a) intersection multiplicities are invariant under projective coordinate transformations as in Exercise 3.10, and they satisfy all the other properties of the multiplicities in Remark 2.4, Lemma 2.5, and Proposition 2.10.

Example 3.22. Let us compute the intersection multiplicity of the curve $F = yz - x^2$ (whose affine part is shown on the right) with the line $G = z$ at infinity at the common point $P = (0:1:0)$. For this we choose the affine part given by $y = 1$ and affine coordinates x and z. We then obtain

$$\mu_P(F,G) = \mu_{(0,0)}(z - x^2, z) = 2$$

by Example 2.11.



Construction 3.23 (Tangents and multiplicities of points, smooth and singular points)**.** The remaining concepts of Chapter 2 are also transferred easiest to a projective curve F using affine parts. So for a point $P = (x_0:y_0:1) \in \mathbb{P}^2$ in the affine part $\mathbb{A}^2$, we define the **multiplicity** $m_P(F)$ of F at P to be $m_{(x_0,y_0)}(F^{\mathrm{i}})$ in the sense of Definition 2.18. A **tangent** to F at P is the projective closure of a tangent to F^{i} at (x_0,y_0). If P is not in the affine part, we choose a different coordinate for the line at infinity as in Example 3.22 (it can be checked that this does not depend on the choice of coordinate).

We say that $P \in F$ is a **smooth** or **regular** point if $m_P(F) = 1$; its unique tangent is denoted $T_P F$. Otherwise, P is called a **singular** point of F. The curve F is said to be **smooth** or **regular** if all its points are smooth; otherwise F is called **singular**.

As in the affine case, there is a simple criterion to determine all singular points of a given projective curve. To prove it, we need a simple lemma first.

Lemma 3.24. *For any homogeneous polynomial $F \in K[x,y,z]$ of degree d we have*

$$x\frac{\partial F}{\partial x} + y\frac{\partial F}{\partial y} + z\frac{\partial F}{\partial z} = dF.$$

Proof. For $F = \sum_{i+j+k=d} a_{i,j,k}x^i y^j z^k$ we have $x\frac{\partial F}{\partial x} = \sum_{i+j+k=d} i\,a_{i,j,k}x^i y^j z^k$. An analogous formula holds for the other partial derivatives, and hence we conclude

$$x\frac{\partial F}{\partial x} + y\frac{\partial F}{\partial y} + z\frac{\partial F}{\partial z} = \sum_{i+j+k=d} (i+j+k)\,a_{i,j,k}x^i y^j z^k = dF. \qquad \square$$

Proposition 3.25 (Projective Jacobi Criterion)**.** *Let P be a point on a projective curve F.*

(a) *P is a singular point of F if and only if $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$.*

(b) *If P is a smooth point of F the tangent to F at P is given by*

$$T_P F = \frac{\partial F}{\partial x}(P)\cdot x + \frac{\partial F}{\partial y}(P)\cdot y + \frac{\partial F}{\partial z}(P)\cdot z.$$

Proof. Without loss of generality we may assume that $P = (x_0:y_0:1)$ is in the affine part of F.

(a) By the affine Jacobi criterion of Proposition 2.24 (a) we know that P is a singular point of F if and only if $\frac{\partial F^{\mathrm{i}}}{\partial x}(x_0,y_0) = \frac{\partial F^{\mathrm{i}}}{\partial y}(x_0,y_0) = 0$. As dehomogenizing F (which is just setting $z = 1$) commutes with taking partial derivatives with respect to x and y, this is equivalent to $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = 0$. This is in turn equivalent to $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$ by Lemma 3.24 since $F(P) = 0$ by assumption.

(b) By Proposition 2.24 (b) the affine tangent to F at P is given by

$$\begin{aligned}&\frac{\partial F^{\mathrm{i}}}{\partial x}(x_0,y_0)\cdot(x-x_0)+\frac{\partial F^{\mathrm{i}}}{\partial y}(x_0,y_0)\cdot(y-y_0)\\ &\quad=\frac{\partial F}{\partial x}(P)\cdot x+\frac{\partial F}{\partial y}(P)\cdot y-\left(\frac{\partial F}{\partial x}(P)\cdot x_0+\frac{\partial F}{\partial y}(P)\cdot y_0\right)\\ &\quad\overset{3.24}{=}\frac{\partial F}{\partial x}(P)\cdot x+\frac{\partial F}{\partial y}(P)\cdot y+\frac{\partial F}{\partial z}(P).\end{aligned}$$

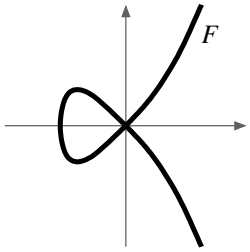
By definition, T_PF is now obtained by taking the projective closure, i. e. the homogenization of this polynomial. □

Remark 3.26. If the ground field K has characteristic 0, Lemma 3.24 tells us for any point $P \in \mathbb{P}^2$ that the conditions $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$ already imply $F(P) = 0$. In contrast to the affine case in Example 2.25, we therefore do not have to check explicitly that the point lies on the curve when computing singular points with the Jacobi criterion.

05

Example 3.27. Let $F = y^2z - x^2z - x^3$ be the projective closure of the real affine curve $y^2 - x^2 - x^3$ of Example 2.21 (b). We have

$$\frac{\partial F}{\partial x} = -2xz - 3x^2, \quad \frac{\partial F}{\partial y} = 2yz, \quad \frac{\partial F}{\partial z} = y^2 - x^2.$$



It is checked immediately that the only common zero of these three polynomials is the point $(0:0:1)$, i. e. the origin of the affine part of F. So by Proposition 3.25 this is the only singular point of F (note that we have already seen in Example 2.25 using the affine Jacobi criterion that the origin is the only singular point of the affine part of F).

In particular, the point $(0:1:0) \in F$ at infinity is a smooth point of F, and the tangent to F there is by Proposition 3.25

$$\frac{\partial F}{\partial x}(0:1:0)\cdot x+\frac{\partial F}{\partial y}(0:1:0)\cdot y+\frac{\partial F}{\partial z}(0:1:0)\cdot z = z,$$

i. e. the line at infinity.

Exercise 3.28. Let F and G be two real smooth projective conics with non-empty set of points. Show that there is a projective coordinate transformation of $\mathbb{P}^2$ as in Exercise 3.10 that takes F to G.

Exercise 3.29. For a projective curve F in the homogeneous coordinates x_0, x_1, x_2 we define the associated *Hessian* to be $H_F := \det\left(\frac{\partial^2 F}{\partial x_i \partial x_j}\right)_{i,j=0,1,2}$.

(a) Show that the Hessian is compatible with coordinate transformations, i. e. if a projective coordinate transformation as in Exercise 3.10 takes F to F' then up to multiplication with a unit it takes H_F to $H_{F'}$.

(b) Let $P \in F$ be a smooth point, and assume that the characteristic of the ground field K is 0. Show that $H_F(P) = 0$ if and only if $\mu_P(F, T_PF) \geq 3$. Such a point is called an *inflection point* of F.

Hint: By part (a) and Exercise 3.10 you may assume after a coordinate transformation that $P = (0:0:1)$ and $T_PF = x_1$.

4. Bézout's Theorem

Let F and G be two projective curves without common component. We have seen already in Remark 3.18 that the intersection $F \cap G$ is finite in this case. Bézout's Theorem, which is the main goal of this chapter, will determine the number of these intersection points, where each such point P will be counted with its intersection multiplicity $\mu_P(F,G)$.

In the same way as for the number of zeros of a univariate polynomial, the result will only be nice (i. e. depend only on the degree of the polynomials) if we assume that the underlying ground field is algebraically closed. To use this assumption we will need the following result from commutative algebra that extends the defining property of an algebraically closed field to polynomials in several variables.

Fact 4.1 (Hilbert's Nullstellensatz)**.** Recall that a field K is called algebraically closed if every univariate polynomial $f \in K[x]$ without a zero in K is constant.

An obvious generalization of this statement to the multivariate case (which can be proven easily by induction of the number of variables) would be that every polynomial $f \in K[x_1,\dots,x_n]$ without a zero in $\mathbb{A}^n$ is constant. However, there is a much stronger statement that also applies to several polynomials at once, or more precisely to the ideal generated by them: *Any ideal I in $K[x_1,\dots,x_n]$ with $V(I) = \emptyset$ over an algebraically closed field K is the unit ideal $I = \langle 1 \rangle$.* This statement is called by its German name **Hilbert's Nullstellensatz** ("theorem of the zeros") [G6, Remark 10.12]. Obviously, in the case $n = 1$ of polynomials in one variable, the ideal I must be generated by a single polynomial f as $K[x_1]$ is a principal ideal domain, and thus Hilbert's Nullstellensatz just reduces to the original statement that f must be constant if it does not have a zero.

Although Bézout's Theorem requires projective curves (as we have already motivated at the beginning of Chapter 3), it is actually more convenient to perform almost all steps required in its proof for the affine case. Our first step will be to compute the sum $\sum_{P \in F \cap G} \mu_P(F,G)$ of the local intersection multiplicities of two affine curves F and G and express it in terms of one global object. In fact, in the same way as $\mu_P(F,G)$ is by definition the dimension of the quotient of the *local* ring $\mathcal{O}_P$ by the ideal $\langle F,G \rangle$, the sum of these multiplicities is just the dimension of the quotient of the *global* polynomial ring $K[x,y]$ by $\langle F,G \rangle$:

Lemma 4.2 (Summing up intersection multiplicities)**.** *Let F and G be two affine curves over K with no common component (so that $F \cap G$ is finite by Remark 3.18). We consider the natural ring homomorphism*

$$\varphi \colon K[x,y]/\langle F,G \rangle \to \prod_{P \in F \cap G} \mathcal{O}_P/\langle F,G \rangle$$

that sends the class of a polynomial $f \in K[x,y]$ to the class of $f \in \mathcal{O}_P$ in each factor $\mathcal{O}_P/\langle F,G \rangle$.

(a) *The morphism φ is surjective.*

(b) *If K is algebraically closed then φ is an isomorphism.*

In particular, we have $\sum_P \mu_P(F,G) \le \dim K[x,y]/\langle F,G \rangle$, with equality if K is algebraically closed.

Proof.

(a) Let $F \cap G = \{P_0,\dots,P_m\}$ with $P_i = (x_i,y_i)$ for $i = 0,\dots,m$. By Exercise 2.7 (a) there is a number $n \in \mathbb{N}$ such that $(x-x_i)^n = (y-y_i)^n = 0 \in \mathcal{O}_{P_i}/\langle F,G \rangle$ for all i. For the polynomial

$$f := \prod_{i : x_i \neq x_0} (x-x_i)^n \cdot \prod_{i : y_i \neq y_0} (y-y_i)^n \quad \in K[x,y]$$

we then have $f(P_0) \neq 0$, so by Exercise 2.7 (b) there is a polynomial representative $g \in K[x,y]$ for $\frac{1}{f} \in \mathcal{O}_{P_0}/\langle F,G \rangle$. The polynomial fg is then mapped by φ ...

- in the component $\mathcal{O}_{P_0}/\langle F,G\rangle$ to $fg = f\cdot\frac{1}{f} = 1$;
- in all other components $\mathcal{O}_{P_i}/\langle F,G\rangle$ for $i>0$ to 0 since $f = 0 \in \mathcal{O}_{P_i}/\langle F,G\rangle$.

By symmetry, we can find in the same way for all $i = 1,\dots,m$ a polynomial that is mapped by φ to 1 in the P_i-component and to 0 in all others. As the image of φ is a subring, it follows that φ is surjective.

(b) In view of (a) it remains to be shown that φ is injective. So let $f \in K[x,y]$ with $\varphi(f) = 0$, and consider the set $I := \{g \in K[x,y] : gf \in \langle F,G\rangle\}$. This is clearly an ideal containing $\langle F,G\rangle$ (usually called the *ideal quotient* $\langle F,G\rangle : \langle f\rangle$). By the Nullstellensatz of Fact 4.1 it suffices to prove that $V(I) = \emptyset$, since then $I = K[x,y]$, hence $1 \in I$, i. e. $f \in \langle F,G\rangle$, and thus $f = 0 \in K[x,y]/\langle F,G\rangle$.

So assume that there is a point $P \in V(I)$. As $F,G \in I$ we know that $P \in F \cap G$. Hence P is one of the points in the product in the target space of φ, and so $f = 0 \in \mathcal{O}_P/\langle F,G\rangle$ as $f \in \ker\varphi$. This means that $f = \frac{a}{g}F + \frac{b}{g}G$ for some polynomials $a,b,g \in K[x,y]$ with $g(P) \neq 0$. But then $gf = aF + bG$, hence $g \in I$, and as $P \in V(I)$ we arrive at the contradiction $g(P) = 0$. □

Remark 4.3. There are two ways to interpret the statement of Lemma 4.2:

(a) A case that often occurs in Lemma 4.2 is that F and G intersect transversely, i. e. that the intersection multiplicities $\mu_P(F,G)$ at all $P \in F \cap G$ are equal to 1. In this case every factor $\mathcal{O}_P/\langle F,G\rangle$ is isomorphic to K by Definition 2.3, and the morphism φ is just the combined evaluation map at all points of $F \cap G$. The assertion of Lemma 4.2 (a) is then simply the interpolation statement that we can always find a polynomial having prescribed values at these points – which is probably not surprising, and is in fact already achieved by a suitable linear combination of polynomials as in Step 1 in the proof. If the intersection is not transverse and $\mu_P(F,G) > 1$ at some point P, then the map φ remembers more information at P on the polynomial than just its value, such as the values of some of its partial derivatives at P.

(b) If you have some commutative algebra background then you probably know the statement of Lemma 4.2 already: As $V(F,G)$ is 0-dimensional, the ring $K[x,y]/\langle F,G\rangle$ is Artinian, and thus by the Structure Theorem on Artinian rings it is isomorphic to the product of its localizations at its various maximal ideals [G6, Proposition 7.20]. If K is algebraically closed then these maximal ideals all correspond to points in $\mathbb{A}^2$ [G6, Corollary 10.10], and so the map φ of the lemma is an isomorphism. If K is not necessarily algebraically closed then there are maximal ideals of $K[x,y]/\langle F,G\rangle$ that are not of this form and thus "missing" in the target space of φ, so that φ is only surjective.

Of course, our goal must now be to compute the dimension of the quotient $K[x,y]/\langle F,G\rangle$. In order to do this, we need a lemma first that tells us how polynomials in the ideal $\langle F,G\rangle$ of $K[x,y]$ can be represented.

Lemma 4.4. *Let F and G be two affine curves of degrees $m := \deg F$ and $n := \deg G$, respectively, such that their leading parts F_m and G_n (as in Notation 2.16) have no common component.*

Then every $f \in \langle F,G\rangle \subset K[x,y]$ of degree $d := \deg f$ can be written as $f = aF + bG$ for two polynomials a and b with $\deg a \leq d-m$ and $\deg b \leq d-n$.

Proof. As $f \in \langle F,G\rangle$ we can write $f = aF + bG$ for some $a,b \in K[x,y]$; choose such a representation with $\deg a$ minimal.

Assume for a contradiction that $\deg a > d-m$ or $\deg b > d-n$. Then aF or bG contains a term of degree bigger than d. As $f = aF + bG$ has degree d this means that the leading terms of aF and bG must cancel in f. Hence, if a_* and b_* denote the leading terms of a and b, respectively, we have $a_*F_m = -b_*G_n$. But F_m and G_n have no common component by assumption, and so we must have $a_* = cG_n$ and $b_* = -cF_m$ for some homogeneous polynomial c. This gives us a new representation

$$f = (a-cG)F + (b+cF)G$$

in which the leading term a_* of a cancels the leading term cG_n of cG in the first bracket. Hence $\deg(a-cG) < \deg a$, contradicting the minimality of $\deg a$. □

Lemma 4.5. *Let F and G be affine curves with no common component, of degrees $m := \deg F$ and $n := \deg G$.*

(a) $\dim K[x,y]/\langle F,G\rangle \le mn$.

(b) *If the leading parts F_m and G_n have no common component either then equality holds in (a).*

Proof. For all $d \ge m+n$ consider the sequence of vector space homomorphisms

$$K[x,y]_{\le d-m} \times K[x,y]_{\le d-n} \xrightarrow{\alpha} K[x,y]_{\le d} \xrightarrow{\pi} K[x,y]/\langle F,G\rangle$$
$$(a,b) \longmapsto aF+bG$$

where $K[x,y]_{\le d}$ denotes the vector subspace of $K[x,y]$ of all polynomials of degree at most d, which has dimension $\binom{d+2}{2}$, and π is the quotient map.

The kernel of α consists of all pairs (a,b) of polynomials of degrees at most $d-m$ and $d-n$, respectively, with $aF = -bG$. As F and G have no common component, this is equivalent to $a = cG$ and $b = -cF$ for some $c \in K[x,y]_{\le d-m-n}$, so that

$$\ker\alpha = K[x,y]_{\le d-m-n} \cdot (G,-F). \tag{1}$$

Moreover, it is obvious that

$$\operatorname{im}\alpha \subset \ker\pi. \tag{2}$$

So we conclude with the homomorphism theorem

$$\begin{aligned}
\dim\operatorname{im}\pi &= \binom{d+2}{2} - \dim\ker\pi \\
&\overset{(2)}{\le} \binom{d+2}{2} - \dim\operatorname{im}\alpha \\
&= \binom{d+2}{2} - \binom{d-m+2}{2} - \binom{d-n+2}{2} + \dim\ker\alpha \\
&\overset{(1)}{=} \binom{d+2}{2} - \binom{d-m+2}{2} - \binom{d-n+2}{2} + \binom{d-m-n+2}{2} \\
&= mn.
\end{aligned}$$

Note that this bound is independent of d (as long as $d \ge m+n$), and thus also holds for the projection map $\pi\colon K[x,y] \to K[x,y]/\langle F,G\rangle$ from the full polynomial ring, which is surjective. It follows that $\dim K[x,y]/\langle F,G\rangle \le mn$, which is (a).

For (b), it suffices to establish equality in (2) above, i. e. that $\ker\pi \subset \operatorname{im}\alpha$. But this is precisely the statement of Lemma 4.4. □

We can now switch back to the projective case and prove the main result of this chapter.

Corollary 4.6 (Bézout's Theorem). *Let F and G be projective curves without common component over an infinite field K. Then*

$$\sum_{P \in F\cap G} \mu_P(F,G) \le \deg F \cdot \deg G.$$

Moreover, equality holds if K is algebraically closed.

Proof. By Lemma 1.11 (b) there is a point Q in the affine part of $\mathbb{P}^2$ which does not lie on $F^{\mathrm{i}} \cup G^{\mathrm{i}}$, i. e. neither on F nor on G. Moreover, as K is infinite but $F \cap G$ finite by Proposition 1.12 (b), we can pick a line L through Q which does not intersect $F \cap G$. Now we make a projective coordinate transformation so that L becomes the line at infinity. Then neither F nor G contains the line at infinity as a component (so that $\deg F^{\mathrm{i}} = \deg F$ and $\deg G^{\mathrm{i}} = \deg G$), and all intersection points of F and G lie in the affine part (i. e. they are also intersection points of the affine curves F^{i} and G^{i}).

Applying Lemma 4.2 (a) and 4.5 (a) to F^{i} and G^{i} then yields

$$\sum_{P \in F \cap G} \mu_P(F,G) = \sum_{P \in F^{\mathrm{i}} \cap G^{\mathrm{i}}} \mu_P(F^{\mathrm{i}}, G^{\mathrm{i}}) \overset{4.2}{\leq} \dim K[x,y]/\langle F^{\mathrm{i}}, G^{\mathrm{i}} \rangle \overset{4.5}{\leq} \deg F^{\mathrm{i}} \cdot \deg G^{\mathrm{i}} = \deg F \cdot \deg G. \quad (*)$$

Now let K be algebraically closed. Then the first inequality is actually an equality by Lemma 4.2 (b). Moreover, the leading parts of F^{i} and G^{i} are homogeneous polynomials in two variables, and hence a product of linear factors by Exercise 3.11 (b). But these factors correspond exactly to the points at infinity of the two curves by Construction 3.15 (b). As there are no such common points by our choice of L, we conclude that the leading parts of F^{i} and G^{i} have no common component, and thus by Lemma 4.5 (b) that the second inequality in $(*)$ is actually an equality as well. □

Remark 4.7 (Bézout's Theorem over arbitrary ground fields)**.** It can be shown that (the inequality part of) Bézout's Theorem holds in fact over arbitrary fields. The assumption of an infinite ground field was only necessary for the strategy of our proof to choose coordinates so that all intersection points of the curves lie in the affine part – which would not be possible over finite fields, since the two curves might then intersect in every point of $\mathbb{P}^2$ (without having a common component).

06

Remark 4.8. Let F and G be two projective curves without common component (over an infinite ground field K).

(a) As the intersection multiplicity at each point of $F \cap G$ is at least 1, it follows from Bézout's Theorem that F and G intersect in at most $\deg F \cdot \deg G$ points (disregarding the multiplicities).

(b) If K is algebraically closed, Bézout's Theorem implies in particular that F and G intersect in at least one point. Note that already this statement is non-trivial – and clearly false for general ground fields, as then already $V(F)$ might be empty.

(c) Moreover, Bézout's Theorem shows that $\mu_P(F,G) \leq \deg F \cdot \deg G$ for all $P \in \mathbb{P}^2$. Of course, this then holds for affine curves as well and can be used to improve Algorithm 2.12 to compute $\mu_P(F,G)$ without having checked before whether F and G have a common component through P (see also Remark 2.14): If the contributions to $\mu_P(F,G)$ collected by the algorithm exceed $\deg F \cdot \deg G$ we can stop, knowing that F and G must have a common component through P. This additional rule will make the algorithm terminate for all F and G, and means that we can also use it to determine whether F and G have a common component through P.

Exercise 4.9. For the following complex affine curves F and G, determine the points at infinity of their projective closures, and use Bézout's Theorem to read off the intersection multiplicities at all points of $F \cap G$.

(a) $F = x + y^2$ and $G = x + y^2 - x^3$;

(b) $F = y^2 - x^2 + 1$ and $G = (y + x + 1)(y - x + 1)$.

Exercise 4.10. Deduce the following real version of Bézout's Theorem from the complex case: If F and G are two real projective curves without common components then

$$\sum_{P \in F \cap G} \mu_P(F,G) = \deg F \cdot \deg G \mod 2.$$

In particular, two real projective curves of odd degree always intersect in at least one point.

Exercise 4.11. Let F be a complex irreducible projective curve of degree d, and let $P \in \mathbb{P}^2$ be a point. We set $m := m_P(F) \in \mathbb{N}$.

Show that for all but finitely many lines L in $\mathbb{P}^2$ through P, the intersection $F \cap L$ consists of exactly $d - m$ points not equal to P.

We will discuss many examples and applications of Bézout's Theorem in the next chapter. Instead, at the end of this chapter let us prove another theorem that can be obtained by very similar methods and that will be useful later on. It considers a *smooth* projective curve F over an algebraically closed

field and states roughly that, for any two other curves G and H such that F intersects H everywhere with at least the same multiplicity as G, the "remaining multiplicities" $\mu_P(F,H) - \mu_P(F,G)$ can be obtained by intersecting F with another curve.

Corollary 4.12 (**Max Noether's Theorem**). *Let F be a smooth projective curve over an algebraically closed field. Moreover, let G and H be two projective curves that do not have a common component with F.*

If $\mu_P(F,G) \leq \mu_P(F,H)$ for all points $P \in F \cap G$ then there are homogeneous polynomials A and B (of degrees $\deg H - \deg F$ resp. $\deg H - \deg G$ if non-zero), such that

(a) $H = AF + BG$;

(b) $\mu_P(F,H) = \mu_P(F,G) + \mu_P(F,B)$ *for all* $P \in \mathbb{P}^2$.

Proof. As in the proof of Corollary 4.6 we may assume by a projective coordinate transformation that none of the curves contain the line at infinity as a component, and that all points of $F \cap G$ lie in the affine part of $\mathbb{P}^2$. We then have again $\deg F^{\mathrm{i}} = \deg F$, $\deg G^{\mathrm{i}} = \deg G$, $\deg H^{\mathrm{i}} = \deg H$, and the leading parts of F^{i} and G^{i} have no common component.

Now F is assumed to be smooth, and hence – working with the affine curves for a moment – the assumption $\mu_P(F^{\mathrm{i}},G^{\mathrm{i}}) \leq \mu_P(F^{\mathrm{i}},H^{\mathrm{i}})$ implies $\langle F^{\mathrm{i}},H^{\mathrm{i}}\rangle \subset \langle F^{\mathrm{i}},G^{\mathrm{i}}\rangle$ in $\mathscr{O}_P$ for all $P \in F \cap G$ by Proposition 2.26, and thus in particular $H^{\mathrm{i}} \in \langle F^{\mathrm{i}},G^{\mathrm{i}}\rangle$ in $\mathscr{O}_P$. By Lemma 4.2 (b) we then have $H^{\mathrm{i}} \in \langle F^{\mathrm{i}},G^{\mathrm{i}}\rangle$ in $K[x,y]$ as well. But as the leading parts of F^{i} and G^{i} have no common component, Lemma 4.4 gives us an equation

$$H^{\mathrm{i}} = aF^{\mathrm{i}} + bG^{\mathrm{i}}$$

for some polynomials a and b of degrees at most $\deg H - \deg F$ and $\deg H - \deg G$, respectively. Homogenizing this yields H, so a homogeneous polynomial of degree $\deg H$, and thus

$$H = \underbrace{z^{\deg H - \deg F - \deg a} a^{\mathrm{h}}}_{=:A} F + \underbrace{z^{\deg H - \deg G - \deg b} b^{\mathrm{h}}}_{=:B} G,$$

which proves (a). But this also implies part (b), since by (the projective version of) the properties of intersection multiplicities we have

$$\mu_P(F,H) = \mu_P(F,AF+BG) \overset{2.4\,(c)}{=} \mu_P(F,BG) \overset{2.10\,(b)}{=} \mu_P(F,B) + \mu_P(F,G). \qquad \square$$

Exercise 4.13 (*Cayley-Bacharach*). Let F and G be smooth projective cubics over an algebraically closed field that intersect in exactly 9 points $P_1,\ldots,P_9$. Moreover, let E be another cubic that also contains the first eight points $P_1,\ldots,P_8$. Prove that E then also contains P_9.

(Hint: Apply Max Noether's Theorem to a suitable curve H.)

Exercise 4.14. Show by example that Max Noether's Theorem is in general false ...

(a) if the ground field is not algebraically closed; or

(b) if the curve F is not assumed to be smooth.

5. Applications of Bézout's Theorem

Bézout's Theorem as in Corollary 4.6 is our first powerful result of algebraic geometry in these notes. Let us now take some time to study several of its applications, which are in fact of very different flavors.

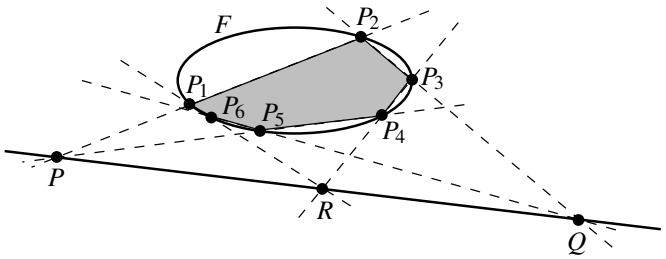
The first application is not much more than an immediate remark; it states that every smooth projective curve over an algebraically closed field is irreducible. As smoothness is easy to check using the Jacobi Criterion of Proposition 3.25 (a), this gives us a very useful sufficient criterion to determine whether a given curve is irreducible (which is usually hard to figure out).

Proposition 5.1 (Irreducibility criterion)**.** *Every smooth projective curve over an algebraically closed field is irreducible.*

Proof. Let $F = G \cdot H$ be a reducible projective curve. By Remark 4.8 (b) there is a point $P \in G \cap H$. Then $m_P(F) = m_P(G) + m_P(H) \geq 1 + 1 = 2$ by Remark 2.23, and so P is a singular point of F. □

Our next statement lies in the field of classical geometry. Over the real numbers it could in principle be proven using elementary methods (and was in fact shown in this way in the first place), but Bézout's Theorem makes the proof much simpler.

Proposition 5.2 (**Pascal's Theorem**)**.** *Let F be an irreducible projective conic with infinitely many points (e. g. over an algebraically closed field, or an ellipse over $\mathbb{R}$). Pick six distinct points $P_1, \dots, P_6$ on F (that can be thought of as the vertices of a hexagon inscribed in F). Then the intersection points of the opposite edges of the hexagon (i. e. $P = \overline{P_1P_2} \cap \overline{P_4P_5}$, $Q = \overline{P_2P_3} \cap \overline{P_5P_6}$, and $R = \overline{P_3P_4} \cap \overline{P_6P_1}$, where $\overline{P_iP_j}$ denotes the line through P_i and P_j) lie on a line.*



Proof. Consider the two (reducible) cubics $G_1 = \overline{P_1P_2} \cup \overline{P_3P_4} \cup \overline{P_5P_6}$ and $G_2 = \overline{P_2P_3} \cup \overline{P_4P_5} \cup \overline{P_6P_1}$. In accordance with Bézout's Theorem, they intersect in the 9 points $P_1, \dots, P_6, P, Q, R$.

Now pick any point $S \in F$ not equal to the previously chosen ones. Of course there are $\lambda_1, \lambda_2 \in K$, not both zero, such that the cubic $G := \lambda_1 G_1 + \lambda_2 G_2$ vanishes at S (since $G(S) = 0$ is one homogeneous linear equation in two variables λ_1, λ_2). Then F meets G in the 7 points $P_1, \dots, P_6, S$, and so by Bézout's Theorem these two curves must have a common component. As $\deg F = 2$, $\deg G = 3$, and F is irreducible, the only possibility for this is that G contains the factor F, so that $G = F \cdot L$ for a line L.

But P, Q, R lie on G (as they lie on G_1 and G_2) and not on F, so they must be on the line L. □

Exercise 5.3. Prove the following converse of Pascal's Theorem:

Let $P_1, \dots, P_6 \in \mathbb{P}^2$ be distinct points so that the six lines $\overline{P_1P_2}, \overline{P_2P_3}, \dots, \overline{P_5P_6}, \overline{P_6P_1}$ (which can be thought of as the sides of the hexagon with vertices $P_1, \dots, P_6$) are also distinct. Let $P = \overline{P_1P_2} \cap \overline{P_4P_5}$, $Q = \overline{P_2P_3} \cap \overline{P_5P_6}$, $R = \overline{P_3P_4} \cap \overline{P_6P_1}$ be the intersection points of opposite sides of the hexagon. If P, Q, R lie on a line, then $P_1, \dots, P_6$ lie on a conic.

Let us next address the question how many singular points we can have on a given projective curve. Exercise 2.30 (b) implies that, for an irreducible curve (over a field of characteristic 0), the number of singular points is always finite. Using Bézout's Theorem, we can now also give an upper bound for this number.

Example 5.4 (Singular points in low degrees)**.** Let F be an irreducible projective curve with infinitely many points (e. g. over an algebraically closed field).

(a) If $\deg F = 1$ then F is a line, which never has any singular points.

(b) If $\deg F = 2$ we claim that F has again no singular points. To show this, assume to the contrary that $P \in F$ is a singular point, and choose any other point $Q \in F$. Let G be the line through P and Q.

As P is a singular point of F, we know by Corollary 2.22 that $\mu_P(F,G) \geq 2$. Hence the total intersection multiplicity of F and G is at least

$$\mu_P(F,G) + \mu_Q(F,G) \geq 2+1 = 3,$$

which is bigger than $\deg F \cdot \deg G = 2$. So by Bézout's Theorem F and G must have a common component – which is impossible since F and G are irreducible.

(c) As for degree 3 we have already seen in Example 3.27 that the cubic $F = y^2z - x^2z - x^3$ has exactly one singular point; since F does not contain a line it is also irreducible. In fact, we will see below that an irreducible cubic can have at most one singular point.

The idea to prove such bounds on the number of singular points is very similar to (b) above: Find a suitable curve G through the assumed singular points and some other points of F, and compute the total intersection multiplicity of F and G, where each singular point of F can be counted with multiplicity at least 2. If this total number exceeds $\deg F \cdot \deg G$ we arrive at a contradiction, i. e. the assumed number of singular points was too high.

In order to make this idea into an exact proof, we need an auxiliary lemma first that tells us how we can find curves (such as G above) through a given number of points.

Lemma 5.5 (Curves through given points)**.** *Let $d \in \mathbb{N}_{>0}$. For any $n := \binom{d+2}{2} - 1$ given points in $\mathbb{P}^2$ there is a projective curve of degree d passing through them.*

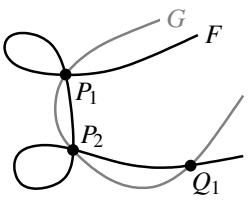
Proof. As in Remark 3.17, the vector space of all homogeneous polynomials of degree d in $K[x,y,z]$ has dimension $n+1$. Its elements are polynomials of the form $F = \sum_{i+j+k=d} a_{i,j,k} x^i y^j z^k$, where $a_{i,j,k}$ are the $n+1$ coordinates of F.

Now note that, for a given point $P = (x_0 : y_0 : z_0)$, the condition $F(P) = \sum_{i+j+k=d} a_{i,j,k} x_0^i y_0^j z_0^k \stackrel{!}{=} 0$ is just a linear equation in these coordinates. Hence, the condition that F vanishes at n given points is a system of n linear equations in $n+1$ variables. By linear algebra, such a system always has a non-trivial solution, which then is a curve of degree d passing through all the given points. □

Proposition 5.6. *Let F be an irreducible projective curve of degree d with infinitely many points (e. g. over an algebraically closed field). Then F has at most $\binom{d-1}{2}$ singular points.*

Proof. By Example 5.4 it suffices to prove the proposition for curves of degree $d \geq 3$. Assume for a contradiction that there are distinct singular points $P_1, \ldots, P_{\binom{d-1}{2}+1}$ of F. Moreover, pick $d-3$ arbitrary further distinct points $Q_1, \ldots, Q_{d-3}$ on F, so that the total number of points is

$$\binom{d-1}{2} + 1 + d - 3 = \binom{d}{2} - 1.$$



By Lemma 5.5, there is therefore a curve G of degree $d-2$ through all these points. As F is irreducible and of bigger degree than G, the curves F and G cannot have a common component. Hence Corollary 4.6 shows that F and G can intersect in at most $\deg F \cdot \deg G = d(d-2)$ points, counted with multiplicities. But the intersection multiplicity at all P_i is at least 2 by Corollary 2.22

since F is singular there. Hence the number of intersection points that we know already, counted with their respective multiplicities, is at least

$$2 \cdot \left(\binom{d-1}{2} + 1 \right) + (d-3) = d(d-2) + 1 > d(d-2),$$

which is a contradiction. □

Exercise 5.7.

(a) Show that a (not necessarily irreducible) reduced curve of degree d in $\mathbb{P}^2$ has at most $\binom{d}{2}$ singular points.

(b) Find an example for each d in which this maximal number of singular points is actually reached.

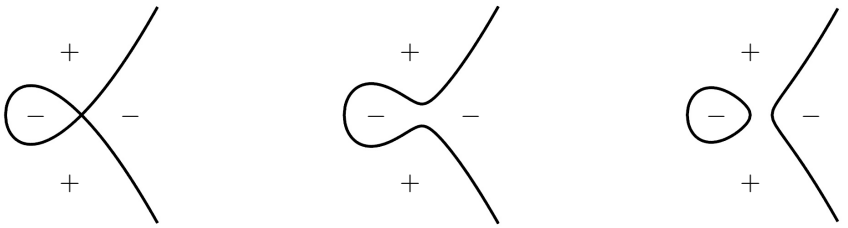
Let us now study smooth curves in more detail. An interesting topic that we have neglected entirely so far is the topology of such curves when we consider them over the real or complex numbers, e. g. their number of connected components in the usual topology. We will now see that Bézout's Theorem is able to answer such questions.

Of course, for these results we will need some techniques and statements from topology that have not been discussed in these notes. The following proofs in this chapter should therefore rather be considered as sketch proofs, which can be made into exact arguments with the necessary topological background. However, all topological results that we will need should be intuitively clear – although their exact proofs are often quite technical. Let us start with the real case, as real curves are topologically simpler than complex ones.

Remark 5.8 (Loops of real projective curves)**.** Let F be a smooth projective curve over $\mathbb{R}$. In the usual topology, its set of points $V(F)$ is then a compact 1-dimensional manifold (see Remark 2.28 (a)). This just means that $V(F)$ is a disjoint union of finitely many connected components, each of which is homeomorphic to a circle. We will refer to these components as *loops* of F. In the following pictures, we will often just draw the affine part of F; a point at infinity in such a loop will then show up as two unbounded ends of the curve. Note that the curve can consist of several loops even if it is irreducible (see Example 1.7 (c)).

07

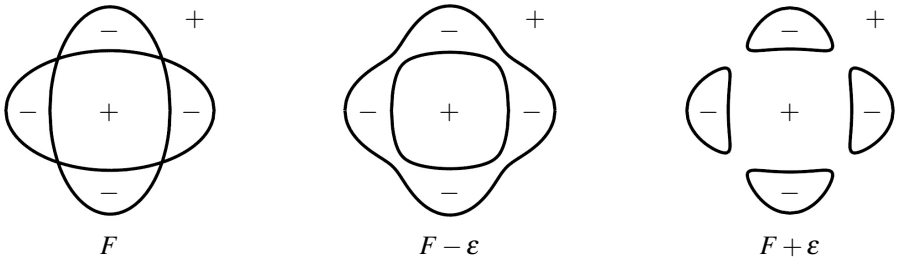
A convenient way to construct such curves is by deformations of singular curves. For example, consider (the projective closure of) the affine cubic $F = y^2 - x^2 - x^3$ with a node at the origin as in Example 2.21 (b) and the picture below on the left. In this picture, we have indicated in addition in which regions of $\mathbb{A}^2 \setminus V(F)$ the polynomial F is negative resp. positive. Together with its one point at infinity, the projective closure of F is homeomorphic to two circles glued together at a point.



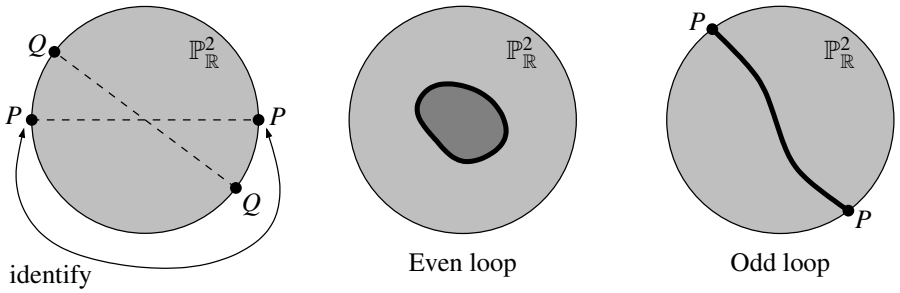
F $\quad$ $F-\varepsilon$ $\quad$ $F+\varepsilon$

Let us now perturb F and consider (the projective closures of) the curves $F \pm \varepsilon$ for a small number $\varepsilon \in \mathbb{R}_{>0}$ instead. Of course, this will only change the regions in which this polynomial is negative resp. positive by a little bit – but the origin, which was on the curve before, now lies in the negative (for $F - \varepsilon$) resp. positive (for $F + \varepsilon$) region. This leads to smooth cubics with one or two loops as in the picture above, depending on the sign of the perturbation.

The same technique applied to a singular quartic curve, e. g. the union of two ellipses given by $F = (x^2 + 2y^2 - 1)(y^2 + 2x^2 - 1)$, yields two or four loops as in the following picture.



Remark 5.9 (Even and odd loops)**.** Although all loops of real smooth curves are homeomorphic to a circle, there are two different kinds of them when we consider their embeddings in projective space. To understand this, recall from Remark 3.6 (a) that $\mathbb{P}^2_{\mathbb{R}}$ is obtained from the upper half sphere (which we will draw topologically as a disc below) by identifying opposite points on the boundary, as in the following picture on the left.



The consequence of this is that we have two different types of loops. An *even loop* is a loop such that its complement has two connected components, which we might call its "interior" (shown in dark in the picture above, homeomorphic to a disc) and "exterior" (homeomorphic to a Möbius strip), respectively. In contrast, an *odd loop* does not divide $\mathbb{P}^2_{\mathbb{R}}$ into two regions; its complement is a single component homeomorphic to a disc. Note that the distinction between even and odd is *not* whether the affine part of the curve is bounded: Whereas an odd loop always has to be unbounded, an even loop may well be unbounded, too. Instead, if you know some topology you will probably recognize that the statement being made here is just that the fundamental group $\pi_1(\mathbb{P}^2_{\mathbb{R}})$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$; the two types of loops simply correspond to the two elements of this group.

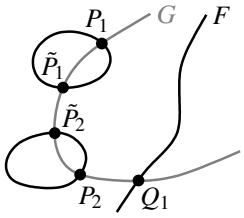
In principle, a real curve can have even as well as odd loops. There is one restriction however: As the complement of an odd loop is simply a disc, all other loops in this complement will have an interior and exterior, so that they are even. In other words, a real smooth curve can have at most one odd loop.

We are now ready to find a bound on the number of loops in an irreducible smooth curve in $\mathbb{P}^2_{\mathbb{R}}$ of a given degree. Interestingly, the idea in its proof is almost identical to that of Proposition 5.6, although the resulting statement is quite different.

Proposition 5.10 (Harnack's Theorem). *An irreducible smooth curve of degree d in $\mathbb{P}^2_{\mathbb{R}}$ has at most $\binom{d-1}{2}+1$ loops.*

Example 5.11. A line ($d=1$) has always exactly one loop. An irreducible smooth conic ($d=2$) is a hyperbola, parabola, or ellipse as in Example 3.16, so in every case the number of loops is again 1 (after adding the points at infinity). For $d=3$ Harnack's Theorem gives a maximum number of 2 loops, and for $d=4$ we get at most 4 loops. We have just seen examples of these numbers of loops in Remark 5.8. In fact, one can show that the bound given in Harnack's theorem is sharp, i. e. that for every d one can find real smooth curves of degree d with exactly $\binom{d-1}{2}+1$ loops.

Proof sketch of Proposition 5.10. Let F be a real irreducible smooth projective curve of degree d; by Example 5.11 it suffices to consider the case $d \geq 3$. Assume that the statement of the proposition is false, i. e. that there are at least $\binom{d-1}{2}+2$ loops. We have seen in Remark 5.9 that at least $\binom{d-1}{2}+1$ of these loops must be even. Hence we can pick points $P_1, \dots, P_{\binom{d-1}{2}+1}$ on distinct even loops of F, and $d-3$ more points $Q_1, \dots, Q_{d-3}$ on another loop (which might be even or odd). So, as in the proof of Proposition 5.6, we have a total of $\binom{d}{2}-1$ points.



Again as in the proof of Proposition 5.6, it now follows that there is a real curve G of degree at most $d-2$ passing through all these points. As F is irreducible and has bigger degree than G, these two curves cannot have a common component, so Bézout's Theorem as in Corollary 4.6 implies that they intersect in at most $d(d-2)$ points, counted with multiplicities. But recall from Remark 5.9 that the even loops of F containing the points P_i divide the real projective plane into two regions, hence if G enters the interior of such a loop it has to exit it again at another point $\tilde{P}_i$ of the same loop as in the picture above (it may also happen that G is singular or tangent to F at P_i, in which case $\mu_{P_i}(F,G) \geq 2$ by Corollary 2.22). So in any case the total number of intersection points, counted with their respective multiplicities, is at least

$$2 \cdot \left(\binom{d-1}{2} + 1 \right) + (d-3) = d(d-2) + 1 > d(d-2),$$

which is a contradiction. □

Let us now turn to the case of complex curves. Of course, their topology is entirely different, as they are 2-dimensional spaces and thus surfaces in the usual topology. In fact, we have seen such a case already in Example 0.2 of the introduction.

Remark 5.12 (Topology of complex curves)**.** Let F be a smooth projective curve over $\mathbb{C}$. Similarly to the real case, its set of points $V(F)$ is then a compact 1-dimensional complex manifold, and hence a compact 2-dimensional real manifold. Moreover, one can show:

(a) $V(F)$ is always an *oriented manifold*, i. e. a "two-sided surface", as opposed to e. g. a Möbius strip. To see this, note that all tangents $T_P F$ for $P \in F$ are 1-dimensional complex vector spaces after shifting P to the origin, and hence admit a well-defined multiplication with the imaginary unit i. Geometrically, this means that all tangent planes to the surface have a well-defined notion of a *positive* rotation by 90 degrees, varying continuously with P – which defines an orientation of the surface.

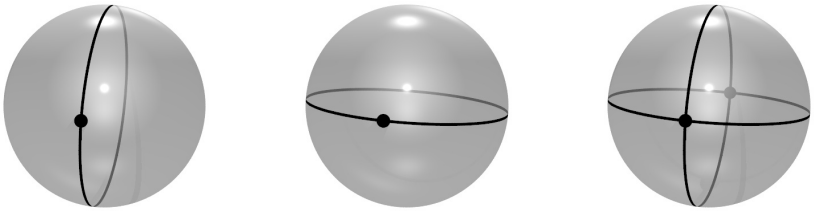
(b) In contrast to the real case that we have just studied, $V(F)$ is always *connected*. In short, the reason for this is that the notion of degree as well as Bézout's Theorem can be extended to compact oriented 2-dimensional submanifolds of $\mathbb{P}^2_{\mathbb{C}}$. Hence, if $V(F)$ had (at least) two connected components X_1 and X_2, each of them would be a compact oriented 2-dimensional manifold itself, and there would thus be well-defined degrees $\deg X_1, \deg X_2 \in \mathbb{N}_{>0}$. But then X_1 and X_2 would have to intersect in $\deg X_1 \cdot \deg X_2$ points (counted with multiplicities), which is obviously a contradiction.

Of course, the methods needed to prove Bézout's Theorem in the topological setting are entirely different from ours in Chapter 4. If you know some algebraic topology, the statement here is that the 2-dimensional homology group $H_2(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z})$ is isomorphic to $\mathbb{Z}$. With this isomorphism, the class of a compact oriented 2-dimensional submanifold in $H_2(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z})$ is a positive number, and the intersection product $H_2(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z}) \times H_2(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z}) \to H_0(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z}) \cong \mathbb{Z}$ (using Poincaré duality) is just the product of these numbers.

It is now a (non-trivial but intuitive) topological result that a connected compact orientable 2-dimensional manifold X is always homeomorphic to a sphere with some finite number of "handles". This number of handles is called the **(topological) genus** of X. Hence every curve in $\mathbb{P}^2_{\mathbb{C}}$ can be assigned a genus that describes its topological type. The picture on the right shows a complex curve of genus 2.

We will see in Definition 8.10 that there is also an algebraic way to assign a genus to a smooth projective curve. It is then applicable to any (algebraically closed) ground field, coincides with the topological genus over $\mathbb{C}$ and plays an important role in the study of functions on the curve. Our goal for the rest of this chapter however will just be to compute the topological genus of a smooth complex projective curve in terms of its degree. To do this, we will need the following technique from topology.

Construction 5.13 (Cell decompositions)**.** Let X be a compact 2-dimensional manifold. A *cell decomposition* of X is given by writing X topologically as a finite disjoint union of points, (open) lines, and (open) discs. This decomposition should be "nice" in a certain sense, e. g. the boundary points of every line in the decomposition must be points of the decomposition. We do not want to give a precise definition here (which would necessarily be technical), but only remark that every "reasonable" decomposition that one could think of will be allowed. For example, the following picture shows three valid decompositions of the complex curve $\mathbb{P}^1_{\mathbb{C}}$, which is topologically a sphere by Remark 3.6 (b).



In the left two pictures, we have 1 point, 1 line, and 2 discs (the two halves of the sphere), whereas in the picture on the right we have 2 points, 4 lines, and 4 discs.

Of course, there are many possibilities for cell decompositions of X. But there is an important number that does not depend on the chosen decomposition:

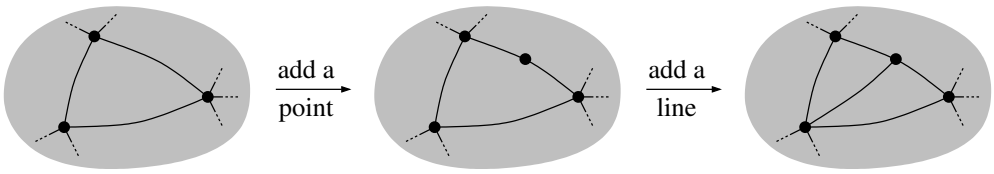
Lemma and Definition 5.14 (Euler characteristic)**.** *Let X be a compact 2-dimensional manifold. Consider a cell decomposition of X, consisting of σ_0 points, σ_1 lines, and σ_2 discs. Then the number*

$$\chi := \sigma_0 - \sigma_1 + \sigma_2$$

*depends only on X, and not on the chosen decomposition. We call it the (topological) **Euler characteristic** of X.*

Proof sketch. Let us first consider the case when we move from one decomposition to a finer one, i. e. if we add points or lines to the decomposition. Such a process is always obtained by performing the following steps a finite number of times:

- Adding another point on a line: In this case we raise σ_0 and σ_1 by 1 as in the picture below, hence the alternating sum $\chi = \sigma_0 - \sigma_1 + \sigma_2$ does not change.
- Adding another line in a disc: In this case we raise σ_1 and σ_2 by 1, so again χ remains invariant.



We conclude that the alternating sum $\sigma_0 - \sigma_1 + \sigma_2$ does not change under refinements. But any two decompositions have a common refinement – which is essentially given by taking all the points and lines in both decompositions, and maybe adding more points where two such lines intersect. For example, in Construction 5.13 the decomposition in the picture on the right is a common refinement of the other two. Hence the Euler characteristic is independent of the chosen decomposition. □

Example 5.15 (Euler characteristic $\leftrightarrow$ genus)**.** Let X be a connected compact orientable 2-dimensional manifold of genus g, and consider the cell decomposition of X as shown on the right. It has $\sigma_0 = 2g+2$ points, $\sigma_1 = 4g+4$ lines, and 4 discs, and hence we conclude that the Euler characteristic of X is

$$\chi = \sigma_0 - \sigma_1 + \sigma_2 = 2 - 2g.$$

In other words, the genus is given in terms of the Euler characteristic as $g = 1 - \frac{\chi}{2}$.

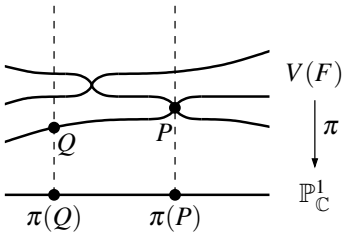
We are now ready to compute the genus of a smooth curve in $\mathbb{P}^2_{\mathbb{C}}$.

Proposition 5.16 (**Topological degree-genus formula**)**.** *A smooth curve of degree d in $\mathbb{P}^2_{\mathbb{C}}$ has topological genus $\binom{d-1}{2}$.*

Proof sketch. Let F be a smooth curve of degree d in $\mathbb{P}^2_{\mathbb{C}}$. By a projective coordinate transformation we can assume that $(0:1:0) \notin F$. Then

$$\pi\colon V(F) \to \mathbb{P}^1_{\mathbb{C}},\ (x:y:z) \mapsto (x:z)$$

is a well-defined map that can be interpreted as a projection, since in the affine part where $z = 1$ it is given by $(x,y) \mapsto x$ as in the picture on the right. Let us study its inverse images of a fixed point $(x:z) \in \mathbb{P}^1_{\mathbb{C}}$. Of course, they are given by the values of y such that $F(x,y,z) = 0$, so that there are exactly d such points – unless the polynomial $F(x,\cdot,z)$ has a multiple zero in y at a point in the inverse image, which happens if and only if F and $\frac{\partial F}{\partial y}$ are simultaneously zero there.



If we choose our original coordinate transformation general enough, exactly two of the zeros of $F(x,\cdot,z)$ will coincide at these points in the common zero locus of F and $\frac{\partial F}{\partial y}$, so that $\frac{\partial^2 F}{\partial y^2} \neq 0$ there and $\pi^{-1}(x:z)$ consists of $d-1$ instead of d points. These points, as e. g. P in the picture above, are usually called the *ramification points* of π. Note that the picture might be a bit misleading since it suggests that $V(F)$ is singular at P, which is not the case. The correct topological picture of the map is impossible to draw however since it would require the real 4-dimensional space $\mathbb{A}^2_{\mathbb{C}}$.

08

At such a ramification point P we have $\mu_P(F, \frac{\partial F}{\partial y}) = 1$ by Corollary 2.22, since in affine coordinates with $P = (x_0, y_0)$ the tangents to the two curves are by Proposition 2.24 (b)

$$T_P F = \frac{\partial F}{\partial x}(P)\cdot(x-x_0) + \underbrace{\frac{\partial F}{\partial y}(P)}_{=0}\cdot(y-y_0) \quad \text{and} \quad T_P \frac{\partial F}{\partial y} = \frac{\partial^2 F}{\partial x \partial y}(P)\cdot(x-x_0) + \underbrace{\frac{\partial^2 F}{\partial y^2}(P)}_{\neq 0}\cdot(y-y_0),$$

which are clearly distinct. Hence by Bézout's Theorem there are exactly $\deg F \cdot \deg \frac{\partial F}{\partial y} = d(d-1)$ ramification points.

Let us now pick a sufficiently fine cell decomposition of $\mathbb{P}^1_{\mathbb{C}}$, containing all images of the ramification points as points of the decomposition. If $\sigma_0, \sigma_1, \sigma_2$ denote the number of points, lines, and discs in this decomposition, respectively, we have $\sigma_0 - \sigma_1 + \sigma_2 = 2$ by Example 5.15 since $\mathbb{P}^1_{\mathbb{C}}$ is topologically a sphere, i. e. of genus 0. Now lift this cell decomposition to a decomposition of $V(F)$ by taking all inverse images of the cells of $\mathbb{P}^1_{\mathbb{C}}$. By our above argument, all cells will have exactly d inverse images – except for the images of the $d(d-1)$ ramification points, which have one inverse image less. So the resulting decomposition of $V(F)$ has $d\sigma_0 - d(d-1)$ points, $d\sigma_1$ lines, and $d\sigma_2$ discs. Hence by Lemma 5.14 the Euler characteristic of $V(F)$ is

$$\chi = d\sigma_0 - d(d-1) - d\sigma_1 + d\sigma_2 = 2d - d(d-1) = 3d - d^2,$$

which means by Example 5.15 that its genus is

$$g = 1 - \frac{\chi}{2} = \frac{1}{2}(d^2 - 3d + 2) = \binom{d-1}{2}. \qquad \square$$

Example 5.17.

(a) A smooth curve of degree 1 or 2 in $\mathbb{P}^2_{\mathbb{C}}$ has topological genus 0, i. e. it is homeomorphic to a sphere. A smooth cubic has genus 1, so it is topologically a torus. We will study such cubic curves in detail in Chapter 7.

(b) Not every natural number can occur as the topological genus of a smooth complex plane curve: For example, there is no smooth complex plane curve of genus 2 since there is no $d \in \mathbb{N}$ with $\binom{d-1}{2} = 2$.

6. Functions and Divisors

Up to now we have essentially studied curves for themselves, i. e. no functions on them or maps between them. In fact, as we restrict ourselves to *plane* curves in these notes it does not make too much sense to consider maps between them as these maps would then somehow have to be compatible with the embeddings in the plane, which is quite restrictive and not very natural. But it is still very fruitful to consider functions on plane curves, i. e. maps to the ground field K, as we will see in the following chapters.

It turns out that the theory of such functions on curves is significantly easier from an algebraic point of view if we restrict to smooth and irreducible curves over an algebraically closed field (where irreducibility is automatic for projective curves by Proposition 5.1). So let us make the convention:

From now on, the ground field K is always assumed to be algebraically closed.
Curves are always assumed to be smooth and irreducible.

In particular, by Remarks 1.14 and 3.19 we can then think of a curve as a subset of $\mathbb{A}^2$ resp. $\mathbb{P}^2$.

Let us start by studying polynomial functions on affine curves.

Definition 6.1 (Affine coordinate rings)**.** Let F be a (smooth and irreducible) affine curve (over an algebraically closed field K). We call

$$A(F) := K[x,y]/\langle F \rangle$$

the **coordinate ring** of F.

In order to avoid overly complicated notations, we will not use any special symbols to denote the equivalence classes in $A(F)$, but rather write e. g. $f \in K[x,y]$ or $f \in A(F)$ for a polynomial resp. its equivalence class modulo F.

Remark 6.2 ($A(F)$ as ring of polynomial functions)**.** Clearly, the elements of $A(F)$ determine well-defined polynomial functions on $V(F) \subset \mathbb{A}^2$ to K by evaluation. Conversely, two polynomials f,g in $K[x,y]$ determine the same polynomial function on $V(F)$ if and only if $f-g$ is identically zero on $V(F)$, i. e. $V(F) \subset V(f-g)$. But as F is irreducible (and the ground field is algebraically closed) this is equivalent to $F \mid f-g$ by Corollary 1.13, and thus to $f = g \in A(F)$. In other words, we see that $A(F)$ is exactly the ring of polynomial functions on the curve.

Remark 6.3 (Algebraic properties of $A(F)$)**.**

(a) As the curve F is assumed to be irreducible, the coordinate ring $A(F)$ is an integral domain: If $fg = 0 \in A(F)$ this means that $F \mid fg$, hence $F \mid f$ or $F \mid g$, which means that $f = 0$ or $g = 0$ in $A(F)$.

(b) In contrast to the polynomial ring $K[x,y]$, the coordinate ring $A(F)$ of an affine curve is in general *not* a unique factorization domain as in Fact 1.2. Actually, determining whether a given coordinate ring $A(F)$ is factorial or not is in general a difficult problem. In these notes we will not study this question in detail; we just have to remember that it does not make sense to talk about irreducible decompositions of elements of $A(F)$.

As $A(F)$ is an integral domain we can also construct its quotient field, corresponding to functions on the curve that are given by quotients of polynomials. Just as in Definition 2.1 this gives rise to local rings describing such functions that have a well-defined value at a given point, and thus also on a neighborhood of this point.

Definition 6.4 (Rational functions and local rings)**.** Let F be an affine curve.

(a) The quotient field (see Construction 1.10)

$$K(F) := \operatorname{Quot} A(F) = \left\{ \frac{f}{g} : f,g \in A(F) \text{ with } g \neq 0 \right\}$$

of the coordinate ring is called the field of **rational functions** on F.

(b) A rational function $\varphi \in K(F)$ is called **regular** at a point $P \in F$ if it can be written as $\varphi = \frac{f}{g}$ with $f,g \in A(F)$ and $g(P) \neq 0$. The regular functions at P form a subring of $K(F)$ containing $A(F)$ denoted by

$$\mathcal{O}_{F,P} := \left\{ \frac{f}{g} : f,g \in A(F) \text{ with } g(P) \neq 0 \right\} \quad \subset K(F).$$

This ring of regular functions at P is called the **local ring** of F at P.

(c) There is a well-defined **evaluation map**

$$\mathcal{O}_{F,P} \to K,\ \frac{f}{g} \mapsto \frac{f(P)}{g(P)}$$

which we will simply write as $\varphi \mapsto \varphi(P)$ for $\varphi \in \mathcal{O}_{F,P}$, and whose kernel is

$$I_{F,P} := \left\{ \frac{f}{g} : f,g \in A(F) \text{ with } f(P) = 0 \text{ and } g(P) \neq 0 \right\}.$$

Remark 6.5 (Algebraic interpretation of local rings)**.** As in the case of the ring $\mathcal{O}_{\mathbb{A}^2,P}$ in Remark 2.2, the rings $\mathcal{O}_{F,P}$ are also local rings in the algebraic sense that they contain exactly one maximal ideal, namely $I_{F,P}$. The proof of this statement is the same as before: If I is an ideal in $\mathcal{O}_{F,P}$ which is not a subset of $I_{F,P}$ then there is an element $\frac{f}{g} \in I$ with $f(P) \neq 0$ and $g(P) \neq 0$. But this is then a unit in $\mathcal{O}_{F,P}$, so that $I = \mathcal{O}_{F,P}$.

Alternatively, just as in Remark 2.2 the ring $\mathcal{O}_{F,P}$ is the localization of $A(F)$ at the maximal ideal $\langle x - x_0, y - y_0 \rangle$ with $P = (x_0, y_0)$, and thus a local ring.

It is straightforward to transfer our notion of intersection multiplicity of two curves to a definition of multiplicity of a polynomial or rational function (and hence also of elements of local rings) on a curve. It should be thought of as the order of a zero or pole of such a function as in the introduction to Chapter 2 – an interpretation that will become even more natural in Proposition 6.10 and Remark 6.11.

Construction 6.6 (Multiplicities of rational functions)**.** Let P be a point on an affine curve F.

(a) For a polynomial function $f \in A(F)$ we define its **multiplicity** at P to be

$$\mu_P(f) := \mu_P(F,f) \overset{2.3}{=} \dim \mathcal{O}_{\mathbb{A}^2,P}/\langle F,f \rangle \quad \in \mathbb{N} \cup \{\infty\}.$$

Note that this is well-defined since $f = g \in A(F)$ implies $g = f + hF$ for some polynomial h, and thus $\mu_P(F,f) = \mu_P(F,g)$ by Remark 2.4 (c). By Exercises 2.7 and 2.8 this multiplicity is infinite if and only if f and F have a common component through P, i. e. (since F is irreducible) if and only if $f = 0 \in A(F)$.

The most important property of this multiplicity is that it is additive: By Proposition 2.10 (b) we have

$$\mu_P(fg) = \mu_P(f) + \mu_P(g)$$

for any $f,g \in A(F)$.

(b) For a rational function $\varphi = \frac{f}{g} \in K(F)$ the **multiplicity** at P is defined by

$$\mu_P(\varphi) := \mu_P(f) - \mu_P(g) \quad \in \mathbb{Z} \cup \{\infty\}.$$

Again this is well-defined: As $g \neq 0 \in A(F)$ we have $\mu_P(g) < \infty$ by (a), and if $\frac{f}{g} = \frac{f'}{g'} \in K(F)$ then $fg' = gf' \in A(F)$, so that

$$\mu_P(f) - \mu_P(g) = \mu_P(f') - \mu_P(g')$$

by the additivity of multiplicities of polynomial functions. Moreover, the multiplicity $\mu_P(\varphi)$ is infinite if and only if $\mu_P(f)$ is infinite, i. e. if and only if $f = 0$, and thus $\varphi = 0$.

The additivity of multiplicities immediately extends to rational functions as well: For $\varphi = \frac{f}{g}$ and $\psi = \frac{f'}{g'}$ in $K(F)$ we have

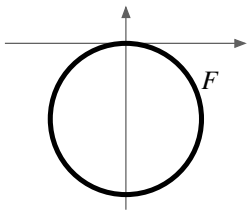
$$\begin{aligned}\mu_P(\varphi\psi) &= \mu_P\left(\frac{ff'}{gg'}\right) = \mu_P(ff') - \mu_P(gg') = \mu_P(f) + \mu_P(f') - \mu_P(g) - \mu_P(g') \\ &= \mu_P(\varphi) + \mu_P(\psi).\end{aligned}$$

In particular, as the multiplicity is finite for elements of $K(F)^*$ this means that μ_P is a group homomorphism from $K(F)^*$ to $\mathbb{Z}$.

If a (polynomial or rational) function has multiplicity $n > 0$ at P we say that it has a **zero** of **order** n at P; if $n < 0$ we say that it has a **pole** of **order** $-n$ at P.

Example 6.7. Consider the rational function $\varphi = \frac{y}{x}$ on the (complex) affine curve $F = y^2 + y + x^2$. A picture of the real points of F is shown in the picture on the right. Using the rules of Chapter 2 for computing intersection multiplicities we obtain at the origin

$$\mu_0(x) = \mu_0(x, y^2 - y - x^2) = \mu_0(x, y^2 - y) = 1$$
$$\text{and} \quad \mu_0(y) = \mu_0(y, y^2 - y - x^2) = \mu_0(y, x^2) = 2,$$



which is also easy to interpret geometrically by Corollary 2.22 as y is the tangent to F there. We conclude that $\mu_0(\varphi) = 2 - 1 = 1$, i. e. that φ has a zero of order 1 at the origin.

Exercise 6.8. Let P be a point on an affine curve F. Check that the local rings of $\mathbb{A}^2$ and F at P are related by $\mathcal{O}_{F,P} \cong \mathcal{O}_{\mathbb{A}^2,P}/\langle F \rangle$, and hence that $\mu_P(f) = \dim \mathcal{O}_{F,P}/\langle f \rangle$ for all $f \in A(F)$.

Remark 6.9 (Multiplicities of regular functions)**.** If $\varphi \in \mathcal{O}_{F,P}$ is an element of the local ring we can write it as $\varphi = \frac{f}{g}$, where $f, g \in A(F)$ with $g(P) \neq 0$. As this means that $\mu_P(g) = 0$, we see that $\mu_P(\varphi) = \mu_P(f) \geq 0$: Elements of the local ring cannot have a pole there. In particular, if φ is a unit then $\mu_P(\varphi^{-1}) = -\mu_P(\varphi) \geq 0$ as well, and we must have $\mu_P(\varphi) = 0$.

One would probably expect that the converse holds as well, i. e. that a rational function without a pole at P is regular at P. Note however that this is not obvious from the definitions, as it might happen just as in Example 6.7 that $\varphi = \frac{f}{g}$ with $\mu_P(f) \geq \mu_P(g) > 0$: In this case we have $\mu_P(\varphi) \geq 0$ but φ is not given as a quotient with non-vanishing denominator, so that it is not visibly regular. Nevertheless this statement turns out to be true as we will show in the next proposition: It is a consequence of the fact that $\mathcal{O}_{F,P}$ is what is called a *discrete valuation ring* in commutative algebra (see also Remark 2.28 (c)).

Proposition 6.10 ($\mathcal{O}_{F,P}$ is a discrete valuation ring)**.** *Let P be a point on an affine curve F.*

(a) *The ideal $I_{F,P}$ is principal, i. e. it can be written as $I_{F,P} = \langle t \rangle$ for some $t \in \mathcal{O}_{F,P}$ (which is unique up to units). We call t a **local coordinate** for F at P.*

(b) *Given a local coordinate t for F at P, every non-zero rational function $\varphi \in K(F)^*$ can be written uniquely as $\varphi = c\,t^n$ for a unit $c \in \mathcal{O}_{F,P}$ and $n \in \mathbb{Z}$, namely for $n = \mu_P(\varphi)$.*

In particular, we have $\varphi \in \mathcal{O}_{F,P}$ if and only if $\mu_P(\varphi) \geq 0$, i. e. if and only if φ does not have a pole at P.

Proof.

(a) Let t be (the class of) a line through P which is not the tangent $T_P F$, so that $\mu_P(t) = 1$ by Corollary 2.22. As t vanishes at P we have $t \in I_{F,P}$, and thus

$$1 = \mu_P(t) \overset{6.8}{=} \dim \mathscr{O}_{F,P}/\langle t \rangle \geq \dim \mathscr{O}_{F,P}/I_{F,P} \geq 1,$$

where the last inequality holds as the constant function 1 is a non-zero element of $\mathscr{O}_{F,P}/I_{F,P}$. We conclude that we must have equality, and thus $I_{F,P} = \langle t \rangle$.

(b) We can write $\varphi = \frac{f}{g}$ with $f, g \in A(F)\backslash\{0\}$. For $m = \mu_P(f)$, i. e. $\mu_P(f) = \mu_P(t^m)$, we then have $\langle F, f \rangle = \langle F, t^m \rangle$ in $\mathscr{O}_{\mathbb{A}^2,P}$ by Proposition 2.26. This means that $\langle f \rangle = \langle t^m \rangle$ in $\mathscr{O}_{F,P}$ by Exercise 6.8, and hence that $f = d\,t^m \in \mathscr{O}_{F,P}$ for a unit d. In the same way we can write $g = e\,t^r$ for a unit e and $r = \mu_P(g)$, and hence $\varphi = c\,t^n$ with $c = \frac{d}{e}$ and $n = m - r$ as desired.

The number n in such a representation is clearly unique: As units have multiplicity 0 by Remark 6.9 we must have $n = \mu_P(c\,t^n) = \mu_P(\varphi)$.

In particular, if $n \geq 0$ then clearly $\varphi = c\,t^n \in \mathscr{O}_{F,P}$; the converse follows again from Remark 6.9. □

Remark 6.11 (Discrete valuation rings in commutative algebra)**.** Over the ground field $K = \mathbb{C}$, the local coordinate t in Proposition 6.10 can be thought of as an analytic local coordinate around P on the 1-dimensional complex manifold $V(F)$ as in Remark 2.28 (b). Consequently, the multiplicity of a function at P just specifies how often this coordinate t can be split off as a linear factor.

As we have mentioned already, in commutative algebra a local ring with the properties of Proposition 6.10 is called a *discrete valuation ring*; the multiplicity $\mu_P(\varphi)$ is therefore also often called the *valuation* of φ. Moreover, Proposition 6.10 (b) means that (in contrast to the ring $A(F)$, see Remark 6.3 (b)), $\mathscr{O}_{F,P}$ is a factorial ring again, with t as the only irreducible element. Despite its more complicated construction, the local ring $\mathscr{O}_{F,P}$ is therefore much simpler than $A(F)$ from an algebraic point of view, and we will often prefer to work with it rather than with polynomials.

The structure of a discrete valuation ring also allows to compute the multiplicity of the sum of two rational functions.

09

Corollary 6.12. *Let P be a point on a curve F. For any two rational functions $\varphi, \psi \in K(F)$ we have*

$$\mu_P(\varphi + \psi) \geq \min(\mu_P(\varphi), \mu_P(\psi)),$$

with equality holding if $\mu_P(\varphi) \neq \mu_P(\psi)$.

Proof. We may restrict to the case when φ and ψ are non-zero, as the statement is trivial otherwise. By symmetry we may also assume that $n := \mu_P(\varphi) \leq m := \mu_P(\psi)$. Proposition 6.10 then tells us that we can write $\varphi = c\,t^n$ and $\psi = d\,t^m$ for some units c and d and a local coordinate t, and thus

$$\mu_P(\varphi + \psi) = \mu_P\Big(c\,t^n\Big(1 + \frac{d}{c}\,t^{m-n}\Big)\Big) = \mu_P(c\,t^n) + \mu_P\Big(\underbrace{1 + \frac{d}{c}\,t^{m-n}}_{\in \mathscr{O}_{F,P}}\Big) \overset{6.9}{\geq} \mu_P(c\,t^n) = n. \qquad (*)$$

Moreover, if $n \neq m$ then $1 + \frac{d}{c}\,t^{m-n}$ has value 1 at P and hence is a unit in $\mathscr{O}_{F,P}$, which means by Remark 6.9 again that we have equality in $(*)$. □

As a final result on affine curves, we can now show that rational functions that are required to be regular at every point of the curve are exactly the polynomial functions, i. e. the elements of the coordinate ring.

Proposition 6.13 (Global regular functions on affine curves)**.** *Let F be an affine curve. Then*

$$\bigcap_{P \in F} \mathscr{O}_{F,P} = A(F) \quad \subset K(F).$$

Proof. Clearly, all polynomial functions in $A(F)$ are everywhere regular, so it remains to prove the converse. For $\varphi \in \bigcap_{P \in F} \mathcal{O}_{F,P} \subset K(F)$ consider the ideal $I := \{g \in K[x,y] : g\varphi \in A(F)\}$. Then $V(I) = \emptyset$: If we had a point $P \in V(I)$, it would follow first of all that $P \in F$ since $F \in I$. Hence we have $\varphi \in \mathcal{O}_{F,P}$, i. e. we can write $\varphi = \frac{f}{g}$ for polynomials f and g with $g(P) \neq 0$. As $g\varphi = f \in A(F)$ this means that $g \in I$, leading to the contradiction $g(P) = 0$ since $P \in V(I)$.

We conclude that $V(I) = \emptyset$, and hence by the Nullstellensatz of Fact 4.1 that $I = K[x,y]$, which means that $1 \in I$, i. e. $\varphi \in A(F)$. □

Let us now pass on to projective curves, which will be our main objects of interest for the rest of these notes. The constructions of rational functions, local rings, and multiplicities in this case are essentially analogous to the ones considered above, taking care of the fact as in Remark 3.7 that we need to consider homogeneous polynomials resp. quotients of homogeneous polynomials of the same degree.

Definition 6.14 (Homogeneous coordinate rings)**.** Let F be a projective curve.

(a) We call
$$S(F) := K[x,y,z]/\langle F \rangle$$
the **(homogeneous) coordinate ring** of F. As in Remark 6.3 in the affine case, it is an integral domain as F is still assumed to be irreducible.

(b) A non-zero element $f \in S(F)$ is called **homogeneous** of degree d if it can be represented by a homogeneous polynomial of degree d in $K[x,y,z]$. The vector space of these elements, together with 0, will be denoted $S_d(F)$.

Remark 6.15 (Direct sum decomposition of $S(F)$)**.** Even if the representative modulo F of an element in $S(F)$ is not unique, we claim that we still have a direct sum decomposition
$$S(F) = \bigoplus_{d \in \mathbb{N}} S_d(F).$$
In fact, it is obvious that $S(F)$ is the sum of all $S_d(F)$, so let us show that this sum is direct. To do this, assume that $f_0 + \cdots + f_n = 0 \in S(F)$ for some polynomials $f_0, \ldots, f_n$ such that f_d is zero or homogeneous of degree d for all $d = 0, \ldots, n$. This means that $f_0 + \cdots + f_n = gF$ for a polynomial g. Taking the degree-d part of this equation then tells us that $f_d = g_{d-\deg F} F$ (where g_k denotes the degree-k part of g as in Notation 2.16), and thus $f_d = 0 \in S(F)$ for all d.

Construction 6.16 (Rational functions and local rings)**.** Let F be a projective curve.

(a) The field of **rational functions** on F is defined as
$$K(F) := \left\{ \frac{f}{g} : f, g \in S_d(F) \text{ for some } d \in \mathbb{N}, g \neq 0 \right\} \subset \operatorname{Quot} S(F).$$

(b) Analogously to Definition 6.4, we call a rational function $\varphi \in K(F)$ **regular** at a point $P \in F$ if it can be written as $\varphi = \frac{f}{g}$ with $f, g \in S(F)$ homogeneous of the same degree and $g(P) \neq 0$. The regular functions at P form a subring
$$\mathcal{O}_{F,P} := \left\{ \frac{f}{g} \in K(F) : g(P) \neq 0 \right\}$$
of $K(F)$ called the **local ring** of F at P.

(c) The ring of regular functions admits an **evaluation map** $\mathcal{O}_{F,P} \to K$, $\varphi \mapsto \varphi(P)$ with kernel $I_{F,P} := \{\varphi \in \mathcal{O}_{F,P} : \varphi(P) = 0\}$.

Construction 6.17 (Multiplicities of rational functions)**.** Let P be a point on a projective curve F.

(a) For a homogeneous element $f \in S(F)$ we define the **multiplicity** at P as
$$\mu_P(f) := \mu_P(F,f) \overset{3.21}{=} \dim \mathcal{O}_{\mathbb{P}^2,P}/\langle F, f \rangle \in \mathbb{N} \cup \{\infty\}.$$

(b) The **multiplicity** of a rational function $\varphi = \frac{f}{g} \in K(F)$ at P is defined as

$$\mu_P(\varphi) := \mu_P(f) - \mu_P(g).$$

It follows in the same way as in the affine case in Construction 6.6 that these multiplicities are well-defined, infinite exactly for the zero element of $S(F)$ resp. $K(F)$, and additive. The notions of (orders of) zeros and poles are also carried over directly.

Remark 6.18 (Affine and projective local rings)**.** As in Construction 3.20, for a point $P = (x_0 : y_0 : 1)$ on a projective curve F one can check that there is an isomorphism

$$\mathscr{O}_{F,(x_0:y_0:1)} \to \mathscr{O}_{F^{\mathrm{i}},(x_0,y_0)}, \frac{f}{g} \mapsto \frac{f^{\mathrm{i}}}{g^{\mathrm{i}}}$$

sending $I_{F,(x_0:y_0:1)}$ to $I_{F^{\mathrm{i}},(x_0,y_0)}$. Hence the algebraic properties of the local ring as e. g. in Proposition 6.10, Remark 6.11, and Corollary 6.12 carry over directly from the affine to the projective case.

Exercise 6.19. Consider the rational function $\varphi = \frac{x^2}{y^2+yz}$ on the projective curve $F = y^2z + x^3 - xz^2$. Moreover, let $P = (0:0:1) \in F$.

(a) Compute the order $n = \mu_P(\varphi)$.

(b) Determine a local coordinate $t \in \mathscr{O}_{F,P}$.

(c) Give an explicit description of φ in the form $\varphi = ct^n$ for a unit $c \in \mathscr{O}_{F,P}$, where c should be written as $\frac{f}{g}$ for some homogeneous $f, g \in S(F)$ of the same degree with $f(P) \neq 0$ and $g(P) \neq 0$.

Exercise 6.20.

(a) Let P be a point on an affine curve F. Show that there is a rational function $\varphi \in K(F)$ which has exactly one pole which is of order 1 and at P, i. e. such that $\mu_P(\varphi) = -1$ and $\mu_Q(\varphi) \geq 0$ for all $Q \neq P$.

(b) Let P_1 and P_2 be distinct points on a projective conic F. Show that there is a rational function $\varphi \in K(F)$ with $\mu_{P_1}(\varphi) = 1$, $\mu_{P_2}(\varphi) = -1$, and $\mu_P(\varphi) = 0$ at all other points P of F.

Exercise 6.21. Let F be an affine curve. Prove that the affine field of rational functions $K(F)$ is isomorphic to the projective one $K(F^{\mathrm{h}})$.

Before we continue our study of multiplicities of rational functions on projective curves let us introduce the so-called *divisors*, a very convenient piece of notation that allows us to consider the multiplicities at all points of a curve at once. We could have done this already in the affine case, but have chosen not to do so as we will only consider projective curves from now on.

Definition 6.22 (Divisors)**.** Let F be a projective curve.

(a) A **divisor** on F is a formal finite linear combination $a_1P_1 + \cdots + a_nP_n$ of distinct points $P_1, \ldots, P_n \in F$ with integer coefficients $a_1, \ldots, a_n \in \mathbb{Z}$ for some $n \in \mathbb{N}$. Obviously, the divisors on F form an Abelian group under pointwise addition of the coefficients. We will denote it by $\operatorname{Div} F$.

Equivalently, in algebraic terms $\operatorname{Div} F$ is just the *free Abelian group* generated by the points of F (i. e. the group of maps $V(F) \to \mathbb{Z}$ being non-zero at only finitely many points; with a point mapping to its coefficient in the sense above).

(b) A divisor $D = a_1P_1 + \cdots + a_nP_n$ as above is called **effective**, written $D \geq 0$, if $a_i \geq 0$ for all $i = 1, \ldots, n$. If D_1, D_2 are two divisors with $D_2 - D_1$ effective, we also write this as $D_2 \geq D_1$ or $D_1 \leq D_2$. In other words, we have $D_2 \geq D_1$ if and only if the coefficient of any point in D_2 is greater than or equal to the coefficient of this point in D_1. Note that this defines a partial order on $\operatorname{Div} F$.

(c) The **degree** of a divisor $D = a_1P_1 + \cdots + a_nP_n$ is the number $\deg D := a_1 + \cdots + a_n \in \mathbb{Z}$. Obviously, the degree is a group homomorphism $\deg\colon \operatorname{Div} F \to \mathbb{Z}$. Its kernel is denoted by

$$\operatorname{Div}^0 F = \{D \in \operatorname{Div} F : \deg D = 0\}.$$

Note that the name "divisor" in this context is entirely unrelated to the idea of elements of rings dividing one another. Instead, divisors are just given by multiplicities attached to all points on a curve, as appearing naturally in the following situations.

Construction 6.23 (Divisors from polynomials and rational functions)**.** Again, let F be a projective curve. The multiplicities of polynomials and rational functions of Construction 6.17 allow us to define divisors on F as follows.

(a) For a non-zero homogeneous polynomial $f \in S(F)\backslash\{0\}$ the *divisor of* f is defined to be

$$\operatorname{div} f := \sum_{P \in F} \mu_P(f) \cdot P \quad \in \operatorname{Div} F.$$

Hence, the effective divisor $\operatorname{div} f$ contains the data of the zeros of f together with their multiplicities. Note that the sum runs formally over all points of F – but as the number of zeros of f is finite by Remark 3.18, there are only finitely many points in this sum with a non-zero multiplicity, so that we obtain a well-defined divisor.

(b) Similarly, for a non-zero rational function $\varphi \in K(F)^*$ we set

$$\operatorname{div} \varphi := \sum_{P \in F} \mu_P(\varphi) \cdot P \quad \in \operatorname{Div} F.$$

This divisor is not effective; it encodes the zeros and poles of φ together with their multiplicities. By definition, if we write $\varphi = \frac{f}{g}$ as a quotient of two non-zero homogeneous polynomials $f, g \in S(F)\backslash\{0\}$ of the same degree then $\operatorname{div} \varphi = \operatorname{div} f - \operatorname{div} g$.

Example 6.24. Consider the rational function $\varphi = \frac{y}{x}$ on the projective curve $F = y^2 + yz + x^2$ over $\mathbb{C}$, i. e. on the projective closure of the affine curve in Example 6.7. We have seen in this example that φ has a zero of order 1 at $(0:0:1)$. Apart from this point, it is easy to check that the only other point at which y or x vanishes is $(0:-1:1)$, where

$$\mu_{(0:-1:1)}(\varphi) = \mu_{(0:-1:1)}(y) - \mu_{(0:-1:1)}(x) = 0 - 1 = -1.$$

Hence the divisor of φ is

$$\operatorname{div} \varphi = 1 \cdot (0:0:1) - 1 \cdot (0:-1:1).$$

Exercise 6.25. Let $F = y^2z - x^3 + xz^2$. Compute the divisor $\operatorname{div} \frac{y}{z}$ on F.

Remark 6.26 (Additivity of multiplicities for divisors)**.** Let F be a projective curve. The additivity of multiplicities as in Constructions 6.6 and 6.17 translates immediately into the following statements for divisors:

(a) For two homogeneous polynomials $f, g \in S(F)\backslash\{0\}$ we get

$$\begin{aligned}\operatorname{div}(fg) &= \sum_{P \in F} \mu_P(fg) \cdot P = \sum_{P \in F} \mu_P(f) \cdot P + \sum_{P \in F} \mu_P(g) \cdot P \\ &= \operatorname{div} f + \operatorname{div} g.\end{aligned}$$

(b) In the same way we obtain

$$\operatorname{div}(\varphi\psi) = \operatorname{div} \varphi + \operatorname{div} \psi$$

for any two non-zero rational functions $\varphi, \psi \in K(F)^*$. In particular, this means that the map $\operatorname{div}\colon K(F)^* \to \operatorname{Div} F$ is a group homomorphism.

It is also very useful to translate the important theorems of Bézout and Max Noether of Chapter 4 into the language of divisors.

Remark 6.27 (Bézout's Theorem for divisors)**.** For a projective curve F, Bézout's Theorem of Corollary 4.6 implies for the degrees of the divisors of Construction 6.23:

(a) for a non-zero homogeneous polynomial $f \in S(F)\backslash\{0\}$

$$\deg \operatorname{div} f = \sum_{P \in F} \mu_P(f) = \sum_{P \in F} \mu_P(F, f) = \deg F \cdot \deg f;$$

(b) for a non-zero rational function $\varphi \in K(F)^*$ (which we can write as $\varphi = \frac{f}{g}$ with f and g non-zero and homogeneous of the same degree)

$$\deg \operatorname{div} \varphi = \deg \operatorname{div} f - \deg \operatorname{div} g \overset{\text{(a)}}{=} \deg F \cdot \deg f - \deg F \cdot \deg g = 0,$$

i. e. that "a rational function on a projective curve has equally many zeros as poles". In particular, the image of the group homomorphism $\operatorname{div}\colon K(F)^* \to \operatorname{Div} F$ of Remark 6.26 (b) lies in $\operatorname{Div}^0 F$.

Proposition 6.28 (Max Noether's Theorem for divisors)**.** *Let F be a projective curve. Moreover, let $g, h \in S(F)$ be non-zero homogeneous polynomials with $\operatorname{div} g \leq \operatorname{div} h$.*

Then there is a homogeneous polynomial $b \in S(F)$ (of degree $\deg h - \deg g$) with $h = bg$ in $S(F)$, and thus with $\operatorname{div} h = \operatorname{div} b + \operatorname{div} g$.

Proof. As $\operatorname{div} g \leq \operatorname{div} h$ means $\mu_P(g) \leq \mu_P(h)$ for all $P \in F$, Max Noether's Theorem as in Corollary 4.12 (a) implies that there are homogeneous polynomials a and b (of degrees $\deg h - \deg F$ and $\deg h - \deg g$, respectively) such that $h = aF + bg$ in $K[x,y,z]$, and hence $h = bg$ in $S(F)$. The equation $\operatorname{div} h = \operatorname{div} b + \operatorname{div} g$ now follows directly from Remark 6.26 (a) (or Corollary 4.12 (b)). □

As a first consequence of these statements we can identify the rational functions that are regular at every point of the curve. Analogously to Proposition 6.13 we expect such functions to be polynomials – but in the projective case polynomials are only well-defined functions if they are constants:

Corollary 6.29 (Global regular functions on projective curves)**.** *Let F be a projective curve. Then*

$$\bigcap_{P \in F} \mathcal{O}_{F,P} = K \quad \subset K(F),$$

i. e. the only rational functions that are everywhere regular on F are constants.

Proof. Let $\varphi = \frac{f}{g} \in K(F)$ be regular at all points $P \in F$. This means that $0 \leq \mu_P(\varphi) = \mu_P(f) - \mu_P(g)$ for all P, and hence that $\operatorname{div} g \leq \operatorname{div} f$. As f and g have the same degree, Proposition 6.28 then implies that $f = cg$ for a constant c, and hence that $\varphi = \frac{f}{g} = c$ is a constant. □

Remark 6.30 (Recovering rational functions from their divisors)**.** Corollary 6.29 implies that a rational function $\varphi \in K(F)^*$ on a projective curve F is determined up to scalars by its divisor $\operatorname{div} \varphi$: If ψ is another rational function with $\operatorname{div} \psi = \operatorname{div} \varphi$ then $\operatorname{div} \frac{\psi}{\varphi} = 0$ by Remark 6.26 (b), hence $\frac{\psi}{\varphi}$ is some constant $c \in K^*$ by Corollary 6.29, and thus $\psi = c\varphi$.

By definition, the group $\operatorname{Div} F$ of divisors on a projective curve F is a very large free Abelian group. As such, it is not very interesting from a group-theoretic point of view. It turns out that we can get a much smaller and more interesting group by considering a certain quotient of $\operatorname{Div} F$ as follows.

Definition 6.31 (Divisor classes and Picard groups)**.** Let F be a projective curve.

(a) A divisor on F is called **principal** if it is the divisor of a non-zero rational function as in Construction 6.23 (b). The set of all principal divisors will be denoted by

$$\operatorname{Prin} F := \{\operatorname{div} \varphi : \varphi \in K(F)^*\}.$$

As the image of the group homomorphism $\operatorname{div}\colon K(F)^* \to \operatorname{Div} F$ of Remark 6.26 (b) it is clearly a subgroup of $\operatorname{Div} F$, and by Remark 6.27 (b) also of $\operatorname{Div}^0 F$.

(b) The quotient group

$$\operatorname{Pic} F := \operatorname{Div} F / \operatorname{Prin} F$$

is called the **Picard group** or group of **divisor classes** on F. Two divisors D_1 and D_2 defining the same element in $\operatorname{Pic} F$, i. e. with $D_1 - D_2 = \operatorname{div} \varphi$ for a rational function $\varphi \in K(F)^*$, are

said to be **linearly equivalent**, written $D_1 \sim D_2$. Restricting to divisors of degree 0, we also set

$$\operatorname{Pic}^0 F := \operatorname{Div}^0 F / \operatorname{Prin} F,$$

which is a subgroup of $\operatorname{Pic} F$.

10

Remark 6.32. By the homomorphism theorem, the degree of divisors induces isomorphisms $\operatorname{Div} F / \operatorname{Div}^0 F \cong \mathbb{Z}$ and $\operatorname{Pic} F / \operatorname{Pic}^0 F \cong \mathbb{Z}$. This means that the Picard group $\operatorname{Pic} F$ and its degree-0 part $\operatorname{Pic}^0 F$ carry essentially the same information. It just depends on the specific application in mind whether it is more convenient to work with $\operatorname{Pic} F$ or $\operatorname{Pic}^0 F$.

Example 6.33 (Picard groups for curves of degree at most 2)**.**

(a) Let F be a projective line. For any point $P \in F$ let l_P be a line through P different from F, so that P is the only intersection point of F and l_P (with multiplicity 1), and hence $\operatorname{div} l_P = P$ on F. For another point $Q \in F$ we then obtain a rational function $\frac{l_P}{l_Q}$ whose divisor is $P - Q$, so that $P - Q \sim 0$ by definition of linear equivalence.

Now any divisor D of degree 0 can be written as $D = P_1 + \cdots + P_n - Q_1 - \cdots - Q_n$ for some points $P_1, \ldots, P_n, Q_1, \ldots, Q_n$ on F, and hence we conclude that

$$D = (P_1 - Q_1) + \cdots + (P_n - Q_n) \sim 0,$$

so that $\operatorname{Pic}^0 F = \{0\}$ is the trivial group.

(b) If F is a projective conic we have seen in Exercise 6.20 (b) that for any two points P and Q on F there is again a rational function with divisor $P - Q$, so that $P \sim Q$. So we conclude again that $\operatorname{Pic}^0 F = \{0\}$ in the same way as in (a).

For curves of bigger degree however, the Picard group is never trivial:

Proposition 6.34. *Let F be a curve of degree $d \geq 3$. Then $P \not\sim Q$ for any two distinct points P and Q on F. In particular, $\operatorname{Pic}^0 F$ is non-trivial.*

Proof. Assume that $P \sim Q$, i. e. that $P - Q = \operatorname{div} \frac{f}{g}$ for some homogeneous polynomials f and g of the same degree. Pick any line l through Q, so that $\operatorname{div} l = E + Q$ for an effective divisor E of degree $\deg E = d - 1 \geq 2$. As

$$\operatorname{div}(fl) = \operatorname{div} g + \operatorname{div} \frac{f}{g} + \operatorname{div} l = \operatorname{div} g + P - Q + E + Q = \operatorname{div} g + E + P \geq \operatorname{div} g$$

it follows from Max Noether's Theorem in Proposition 6.28 that there is a line l' with $\operatorname{div} l' = E + P$.

But $\deg E \geq 2$ means that E contains at least two points (or one point with multiplicity at least 2). Hence l and l' have to pass through them (resp. be tangent to F at the one point with multiplicity at least 2). As this fixes the line uniquely, it follows that $l = l'$, and thus that $P = Q$.

We conclude that $P \not\sim Q$ for $P \neq Q$, and thus that $P - Q \neq 0 \in \operatorname{Pic}^0 F$. □

Corollary 6.35 (Embedding of a curve in its Picard group)**.** *Let P_0 be a fixed base point on a projective curve F of degree at least 3. Then the map*

$$\Phi \colon V(F) \to \operatorname{Pic}^0 F,\ P \mapsto P - P_0$$

is injective.

Proof. If $\Phi(P) = \Phi(Q)$ then $P - P_0 \sim Q - P_0$, hence $P \sim Q$, and thus $P = Q$ as points in $V(F)$ by Proposition 6.34. □

Remark 6.36. For a projective curve F of degree $\deg F \geq 3$, Corollary 6.35 gives us a natural embedding (after choosing a base point) of the curve F into its degree-0 Picard group $\operatorname{Pic}^0 F$. This is a very interesting statement, as it gives us a natural map between mathematical objects of totally different types (namely a variety and a group).

In the next chapter we will see that this map is even a bijection if $\deg F = 3$, making this correspondence between varieties and groups even more surprising and useful.

Exercise 6.37.

(a) Let F be a projective curve, and let f be a homogeneous polynomial with $\operatorname{div} f = D + E$ for two divisors D and E on F. Show: If D' is linearly equivalent to D and $D' + E$ is effective then there is a homogeneous polynomial g with $\operatorname{div} g = D' + E$.

(b) Let P, Q, R, S be four distinct points on a cubic curve F with $P + Q \sim R + S$. Show that the intersection point of the lines $\overline{PQ}$ and $\overline{RS}$ lies on F.

7. Elliptic Curves

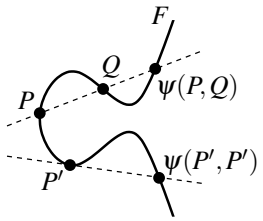
In this chapter, we will interrupt our general discussion of plane curves for a moment to study (projective) curves of degree 3 in detail. We have seen already that this is the first interesting case of curves in many respects: It is, for example, the lowest degree for which real or complex curves are topologically interesting (see Propositions 5.10 and 5.16), and for which the Picard group is non-trivial (see Example 6.33 and Proposition 6.34). We will show now that curves of degree 3 have in fact a very rich structure, both from an algebraic and – over $\mathbb{C}$ – from an analytic point of view. In the literature, they are usually called *elliptic curves*.

Definition 7.1 (Elliptic curves)**.** An **elliptic curve** is simply a projective cubic curve (which is smooth and defined over an algebraically closed field, in accordance with our convention at the beginning of Chapter 6).

The term "elliptic curve" might sound confusing at first, because the shape of a plane cubic curve has no similarities with an ellipse, not even over the real numbers (see e. g. Remark 5.8). The historical reason for this name is that the formula for the circumference of an ellipse can be expressed in terms of an integral over a plane cubic curve.

Probably the single most important (and surprising) result about elliptic curves is that they carry a natural group structure. The easiest, or at least the most conceptual way to prove this is by showing that an elliptic curve admits a natural bijection to its degree-0 Picard group. To establish this, we need the following construction.

Construction 7.2. Let P and Q be two (not necessarily distinct) points on an elliptic curve F. Then there is a unique line l with $P+Q \leq \operatorname{div} l$ on F, namely the line through P and Q if these points are distinct, and the tangent line to F at $P=Q$ otherwise. But $\operatorname{div} l$ is an effective divisor of degree 3 by Remark 6.27 (a), and hence there is a unique point $R \in F$ (which need not be distinct from P and Q) with $\operatorname{div} l = P+Q+R$. In the following, we will denote this point R by $\psi(P,Q)$. In short, it is just "the third point of intersection of the line through P and Q with F".



Lemma 7.3. *For any three points P,Q,R on an elliptic curve F there is a point S on F such that $P+Q \sim R+S$, namely*

$$S = \psi(\psi(P,Q),R).$$

Proof. Applying Construction 7.2 to the points P and Q we find a line l with $\operatorname{div} l = P+Q+\psi(P,Q)$ on F. Similarly, for $\psi(P,Q)$ and R we find a line l' with $\operatorname{div} l' = \psi(P,Q)+R+\psi(\psi(P,Q),R)$. The quotient of these lines is then a rational function on F, whose divisor is therefore linearly equivalent to zero: We have

$$0 \sim \operatorname{div} \frac{l}{l'} = P+Q+\psi(P,Q) - (\psi(P,Q)+R+\psi(\psi(P,Q),R)),$$

and hence, as claimed, $P+Q \sim R+S$ with $S = \psi(\psi(P,Q),R)$. □

Proposition 7.4. *Let P_0 be a fixed point on an elliptic curve F. Then the map*

$$\Phi\colon V(F) \to \operatorname{Pic}^0 F,\ P \mapsto P - P_0$$

of Corollary 6.35 is a bijection.

Proof. As we already know by Corollary 6.35 that Φ is injective, it remains to prove surjectivity. So let D be an arbitrary element of $\operatorname{Pic}^0 F$, which we can write as

$$D = P_1 + \cdots + P_m - Q_1 - \cdots - Q_m$$

for some $m \in \mathbb{N}_{>0}$ and not necessarily distinct points $P_1, \dots, P_m, Q_1, \dots, Q_m \in F$. Assume first that $m \geq 2$. By Lemma 7.3 there is then a point $S \in F$ with $P_1 + P_2 \sim Q_1 + S$, and hence

$$D \sim S + P_3 + \cdots + P_m - Q_2 - \cdots - Q_m.$$

Up to linear equivalence, we have thus reduced the number m of (positive and negative) points in D by 1. Continuing this process as long as $m \geq 2$, we see that $D \sim P - Q$ for some $P, Q \in F$. In the same way, Lemma 7.3 now gives us a point T with $P + P_0 \sim Q + T$, i. e. with $D \sim P - Q \sim T - P_0$. But this means that $\Phi(T) = D$, i. e. that Φ is surjective. □

Remark 7.5. Let F be an elliptic curve. After choosing a base point $P_0 \in F$, Proposition 7.4 gives us a canonical bijection between the variety $V(F)$ and the Abelian group $\mathrm{Pic}^0 F$, i. e. between two very different types of mathematical objects. We can use it to give $V(F)$ the structure of an Abelian group, and $\mathrm{Pic}^0 F$ the structure of a smooth projective variety.

In fact, it can be shown that $\mathrm{Pic}^0 F$ can be made into a variety (the so-called *Picard variety*) for every smooth projective curve F. In contrast, the statement that $V(F)$ has a natural structure of an Abelian group is very special to elliptic curves. Let us explore this group structure in more detail.

Construction 7.6 (The group structure on an elliptic curve)**.** Let P_0 be a fixed base point on an elliptic curve F. As in Remark 7.5, we can use Proposition 7.4 to define a group structure on $V(F)$ in such a way that the map Φ becomes an isomorphism of groups. More precisely, if we denote this group operation on $V(F)$ by the symbol $\oplus$ (to distinguish it from the addition of points in $\mathrm{Div} F$ or $\mathrm{Pic} F$), then $P \oplus Q$ for $P, Q \in F$ is the unique point of F satisfying

$$\Phi(P \oplus Q) = \Phi(P) + \Phi(Q),$$

where "+" denotes the addition of divisors in $\mathrm{Pic}^0 F$. We can use Lemma 7.3 to solve this for $P \oplus Q$:

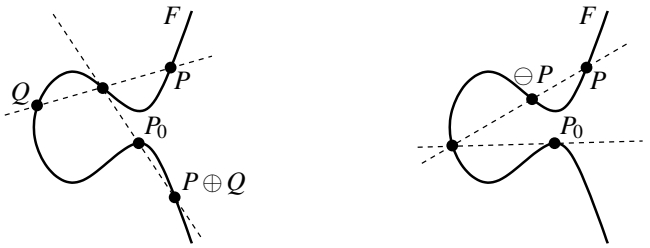
$$\begin{aligned} P \oplus Q &= \Phi^{-1}(\Phi(P) + \Phi(Q)) \\ &= \Phi^{-1}(P - P_0 + Q - P_0) \\ &= \Phi^{-1}(P + Q - 2P_0) \\ &\overset{7.3}{=} \Phi^{-1}(P_0 + \psi(\psi(P,Q), P_0) - 2P_0) \\ &= \Phi^{-1}(\psi(\psi(P,Q), P_0) - P_0) \\ &= \psi(\psi(P,Q), P_0). \end{aligned}$$

In other words, to construct the point $P \oplus Q$ we draw a line through P and Q. Then we draw another line through the third intersection point $\psi(P,Q)$ of this line with F and the point P_0. The third intersection point of this second line with F is then $P \oplus Q$, as in the picture below on the left.

Similarly, for the inverse $\ominus P$ of P in the above group structure we obtain

$$\begin{aligned} \ominus P &= \Phi^{-1}(-\Phi(P)) \\ &= \Phi^{-1}(P_0 - P) \\ &= \Phi^{-1}(P_0 + P_0 - P - P_0) \\ &\overset{7.3}{=} \Phi^{-1}(P + \psi(\psi(P_0, P_0), P) - P - P_0) \\ &= \psi(\psi(P_0, P_0), P). \end{aligned}$$

So to construct the inverse $\ominus P$ we draw the tangent to F through P_0. Then we draw another line through the other intersection point $\psi(P_0, P_0)$ of this tangent with F and the point P. The third intersection point of this second line with F is $\ominus P$, as in the following picture.



Note that, using this geometric description, the operation $\oplus$ could also be defined in a completely elementary way, without referring to the theory of divisors. However, it would then be very difficult to show that we obtain a group structure in this way, in particular to prove associativity.

Remark 7.7 (Non-algebraically closed fields)**.** Let K' be a subfield of K which is not necessarily algebraically closed, such as $\mathbb{R}$ in $\mathbb{C}$ or a finite field in its algebraic closure. Assume that $F \in K'[x,y,z]$ is defined over K'. Note that for two points $P,Q \in V(F) \cap \mathbb{P}^2_{K'}$ on F with coordinates in K' the point $\psi(P,Q)$ then lies in $V(F) \cap \mathbb{P}^2_{K'}$ as well: The polynomial F restricted to the line through P and Q is a cubic homogeneous polynomial over K' that splits off two linear factors over K' corresponding to its zeros P and Q. Hence the remaining linear factor corresponding to $\psi(P,Q)$ is also defined over K', which means that $\psi(P,Q) \in V(F) \cap \mathbb{P}^2_{K'}$.

Choosing the base point P_0 in $V(F) \cap \mathbb{P}^2_{K'}$, we can therefore restrict the group structure on $V(F)$ to $V(F) \cap \mathbb{P}^2_{K'}$, obtaining a subgroup of $V(F)$.

Exercise 7.8. Let F be an elliptic curve of the form

$$F = y^2z - x^3 - \lambda xz^2 - \mu z^3$$

for some given $\lambda, \mu \in K$ (it can be shown that every elliptic curve can be brought into this form by a change of coordinates if the characteristic of K is not 2 or 3). Pick the point $P_0 = (0:1:0) \in F$ as the base point for the group structure on $V(F)$.

For given points P and Q on F compute explicitly the coordinates of the sum $P \oplus Q$ and the inverse $\ominus P$ in terms of the coordinates of P and Q.

Example 7.9 (Elliptic Curve Cryptography)**.** There is an interesting application of the group structure on an elliptic curve to cryptography. The key observation is that "multiplication is easy, but division is hard". More precisely, assume that we are given a specific elliptic curve F, and that we choose a base point $P_0 \in F$ for the group structure as well as an additional point $P \in F$. In view of Remark 7.7, the ground field for the curve does not have to be algebraically closed; in fact, for practical computations one will usually choose a finite field so that its elements can be stored in a chunk of computer memory of fixed size without rounding errors. Then we observe the following:

(a) Given $n \in \mathbb{N}$, the n-fold addition $n \odot P := P \oplus \cdots \oplus P$ can be computed very quickly using Exercise 7.8, even for very large n (think of numbers with hundreds of digits):
- By repeatedly applying the operation $P \mapsto P \oplus P$, we can compute all points $2^k \odot P$ for all k such that $2^k \leq n$.
- Now we just have to add these points $2^k \odot P$ for all k such that the k-th digit in the binary representation of n is 1.

This computes the point $n \odot P$ in a time proportional to $\log n$ (i. e. in a very short time).

(b) On the other hand, given a sufficiently general point $Q \in V(F)$ it is essentially impossible to compute an integer $n \in \mathbb{N}$ such that $n \odot P = Q$ (in case such a number exists). Note that this is not a mathematically precise statement – there is just no known algorithm that can perform the "inverse" of the multiplication of (a) in shorter time than a simple trial-and-error approach (which would be impractical for large n).

Let us now assume that Alice and Bob want to establish an encrypted communication over an insecure channel, but that they have not met in person before, so that they could not secretly agree on

a key for the encryption. Using the above idea, they can then agree (publicly) on a ground field K, a specific elliptic curve F over K, a base point $P_0 \in V(F)$, and another point $P \in V(F)$. Now Alice picks a secret (very large) integer n, computes $n \odot P$ as in (a), and sends (the coordinates of) this point to Bob. In the same way, Bob chooses a secret number m, computes $m \odot P$, and sends this point to Alice.

As Alice knows her secret number n and the point $m \odot P$ from Bob, she can then compute the point $mn \odot P = n \odot (m \odot P)$. In the same way, Bob can compute this point as $mn \odot P = m \odot (n \odot P)$ as well. But except for the data of the chosen curve the only information they have exchanged publicly was P, $n \odot P$, and $m \odot P$, and by (b) it is not possible in practice to recover n or m, and hence $mn \odot P$, from these data. Hence Alice and Bob can use (the coordinates of) $mn \odot P$ as a secret key for their encrypted communication.

This method is actually used by many modern computer applications that need encryption, such as popular instant messengers for secure communication and file encryption software. The most common choice for the parameters is called Curve25519 in the literature, and uses the ground field $\mathbb{Z}/p\mathbb{Z}$ with the prime number $p = 2^{255} - 19$, the curve $F = y^2z - x^3 - 486662x^2z - xz^2$, the base point $P_0 = (0:1:0)$, and a point $P \in F$ with $x = 9$ and $z = 1$ [W].

Exercise 7.10. Let $F = y^2z - x^3 - \lambda xz^2$ be an elliptic curve as in Exercise 7.8 with $\mu = 0$, defined over a field of characteristic p (so that $\mathbb{Z}/p\mathbb{Z}$ is a subfield of K), and with $\lambda \in \mathbb{Z}/p\mathbb{Z}$. Show:

(a) If $p = 3 \bmod 4$ then $V(F) \cap \mathbb{P}^2_{\mathbb{Z}/p\mathbb{Z}}$ contains exactly $p+1$ points.

(b) If $p = 1 \bmod 4$ then the number of points of $V(F) \cap \mathbb{P}^2_{\mathbb{Z}/p\mathbb{Z}}$ may also be smaller or bigger than $p+1$, but is always even.

Exercise 7.11. Let $F = y^2z - x^3 - \lambda xz^2 - \mu z^3$ be an elliptic curve as in Exercise 7.8.

Show that the subgroup $\{D \in \operatorname{Pic} F : 2D \sim 0\}$ of $\operatorname{Pic} F$ has exactly 4 elements and is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

(Hint: Translate the problem to the group structure on $V(F)$.)

11

Let us now restrict our attention to the ground field $\mathbb{C}$, so that an elliptic curve is topologically a torus by Example 5.17 (a). In the remaining part of this chapter we want to see how these tori and elliptic curves arise in complex analysis in a totally different way. As we have not developed any analytic techniques in these notes we will only sketch most arguments; more details can be found e. g. in [Ki, Section 5.1]. Let us start by giving a quick review of what we will need from standard complex analysis. As usual, we will denote a complex variable in $\mathbb{C}$ by z. In contrast, for the rest of this chapter the homogeneous coordinates of $\mathbb{P}^2_{\mathbb{C}}$ will be called x_0, x_1, x_2 instead of x, y, z to avoid confusion.

Remark 7.12 (Holomorphic and meromorphic functions)**.** Let $U \subset \mathbb{C}$ be an open subset. Recall that a function $f\colon U \to \mathbb{C}$ is called *holomorphic* if it is complex differentiable at all points $z_0 \in U$, i. e. if the limit

$$f'(z_0) := \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. A function $f\colon U \to \mathbb{C} \cup \{\infty\}$ is called *meromorphic* if it is holomorphic except for some isolated singularities which are all poles, i. e. if for all $z_0 \in U$ there is a number $n \in \mathbb{Z}$ and a holomorphic function $\tilde{f}$ in a neighborhood of z_0 in U in which

$$f(z) = (z - z_0)^n \cdot \tilde{f}(z).$$

If f does not vanish identically in a neighborhood of z_0 we can moreover assume that $\tilde{f}(z_0) \neq 0$ in this representation; the number n is then uniquely determined. We will call it the *multiplicity* of f at z_0 and denote it by $\mu_{z_0}(f)$. It is obviously the analogue of the multiplicity of a rational function as in Construction 6.6 and Proposition 6.10 (b). The notions of (orders of) *zeros* and *poles* are used for meromorphic functions in the same way as for rational functions. Note that every rational function (i. e. every quotient of polynomials) in z is clearly meromorphic; there are however many more meromorphic than rational functions as e. g. the exponential function $z \mapsto e^z$.

Remark 7.13 (Properties of holomorphic and meromorphic functions)**.** Although the definition of holomorphic, i. e. *complex* differentiable functions is formally exactly the same as that of *real* differentiable functions, the behavior of the complex and real cases is totally different. The most notable differences that we will need are:

(a) Every holomorphic function f is analytic, i. e. it can be represented locally around every point z_0 by its Taylor series. Consequently, a meromorphic function f of order n at z_0 can locally be expanded in a *Laurent series* as $f(z) = \sum_{k=n}^{\infty} c_k (z-z_0)^k$, with $n = \mu_{z_0}(f)$ [G4, Proposition 9.8]. The coefficient c_{-1} of this series is called the *residue* of f at z_0 and denoted by $\operatorname{res}_{z_0} f$.

(b) (*Residue Theorem*) If γ is a closed (positively oriented) path in $\mathbb{C}$ and f is a meromorphic function in a neighborhood of γ and its interior, without poles on γ itself, then

$$\int_\gamma f(z)\,dz = 2\pi \mathrm{i} \sum_{z_0} \operatorname{res}_{z_0} f,$$

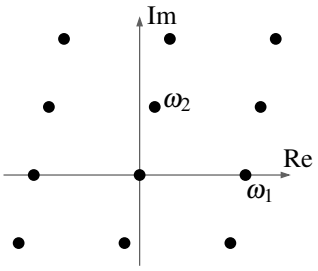
with the sum taken over all z_0 in the interior of γ (at which f has poles) [G4, Proposition 11.14]. In particular, if f is holomorphic in the interior of γ then this integral vanishes.

(c) (*Liouville's Theorem*) Every function that is holomorphic and bounded on the whole complex plane $\mathbb{C}$ is constant [G4, Proposition 8.2].

Construction 7.14 (Tori from lattices)**.** As mentioned above, for our applications to elliptic curves we have to construct a torus. To do this, fix two complex numbers $\omega_1, \omega_2 \in \mathbb{C}$ that are linearly independent over $\mathbb{R}$, i. e. that do not lie on the same real line in $\mathbb{C}$ through the origin. Then

$$\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{m\omega_1 + n\omega_2 : m,n \in \mathbb{Z}\} \quad \subset \mathbb{C}$$

is called a *lattice* in $\mathbb{C}$, as indicated by the points in the picture on the right. It is an additive subgroup of $\mathbb{C}$, and the quotient $\mathbb{C}/\Lambda$ is topologically a torus.



For the rest of this chapter, Λ will always be a fixed lattice in $\mathbb{C}$. Note that functions on the torus $\mathbb{C}/\Lambda$ correspond exactly to Λ-*periodic* functions on $\mathbb{C}$, i. e. to functions f on $\mathbb{C}$ with $f(z+\omega) = f(z)$ for all $z \in \mathbb{C}$ and $\omega \in \Lambda$. In the following, we will use the concepts of functions on $\mathbb{C}/\Lambda$ and Λ-periodic functions on $\mathbb{C}$ interchangeably.

It is our goal to show that the torus $\mathbb{C}/\Lambda$ can be identified with an elliptic curve in a natural way. Let us start with a first auxiliary result that already indicates the similarities between the algebraic and analytic setting: We will show the analytic analogue of Remark 6.27 (b), namely that a meromorphic function on the torus $\mathbb{C}/\Lambda$ has equally many zeros as poles.

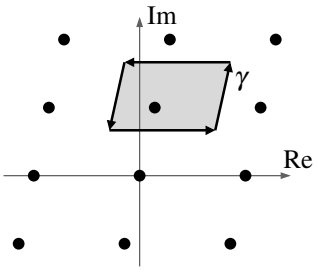
Lemma 7.15. *Let $f\colon \mathbb{C}/\Lambda \to \mathbb{C} \cup \{\infty\}$ be a non-zero meromorphic function. Then*

$$\sum_{z_0 \in \mathbb{C}/\Lambda} \mu_{z_0}(f) = 0.$$

Proof sketch. Let γ be the path around a "parallelogram of periodicity" as in the picture on the right, i. e. a parallelogram with side vectors spanning Λ. We choose it so that the zeros and poles of f do not lie on γ, and hence have a unique representative inside this parallelogram. It follows that

$$\int_\gamma \frac{f'(z)}{f(z)}\,dz = 0 \qquad (*)$$

since the integrals along opposite sides of the parallelogram cancel each other due to the periodicity of f.



On the other hand, we can compute this integral using the Residue Theorem of Remark 7.13 (b): At a point z_0 with $\mu_{z_0}(f) = n$ so that we can write $f(z) = (z-z_0)^n \tilde{f}(z)$ with $\tilde{f}$ holomorphic and non-zero around z_0 as in Remark 7.12 we have

$$\operatorname{res}_{z_0} \frac{f'}{f} = \operatorname{res}_{z_0} \frac{n(z-z_0)^{n-1}\tilde{f} + (z-z_0)^n \tilde{f}'}{(z-z_0)^n \tilde{f}} = \operatorname{res}_{z_0} \left(\frac{n}{z-z_0} + \frac{\tilde{f}'}{\tilde{f}} \right) = n = \mu_{z_0}(f),$$

and hence we obtain by the Residue Theorem

$$\int_\gamma \frac{f'(z)}{f(z)}\, dz = 2\pi \mathrm{i} \sum_{z_0 \in \mathbb{C}/\Lambda} \operatorname{res}_{z_0} \frac{f'}{f} = 2\pi \mathrm{i} \sum_{z_0 \in \mathbb{C}/\Lambda} \mu_{z_0}(f).$$

Comparing this with $(*)$ then gives the desired result. □

Remark 7.16 (Residue Theorem on manifolds)**.** In the same way as Remark 6.27 (b), Lemma 7.15 does not only hold for a torus $\mathbb{C}/\Lambda$, but also for an arbitrary compact 1-dimensional complex manifold X, and thus for any (smooth) complex projective curve. Let us briefly explain how to adapt the proof of Lemma 7.15 to this more general case.

The main step in this generalization is to extend the concepts of path integrals and the Residue Theorem from the complex plane to manifolds. This is not entirely straightforward, since the differential dz in the integral depends on the choice of a local coordinate z on X. As a consequence, there is no well-defined integral over a *function* on X since we would have to combine it with the coordinate-dependent dz to integrate it. Instead, we have to combine a function with a differential to obtain expressions of the form $\alpha = f\,dg$ for (meromorphic) functions f and g that satisfy the usual rules of differentiation. Such objects are called *differential forms* on X.

In these notes we will use differential forms only in a few side remarks that will not be needed later on, and hence we will not introduce them rigorously. Let us just mention that integrals and the Residue Theorem then behave as expected: For a closed path γ and a differential form α on X not having any poles on γ itself, we can define an integral $\int_\gamma \alpha$ whose value can be computed by the Residue Theorem

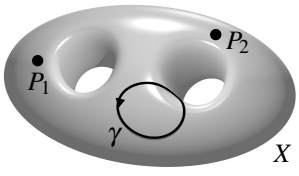
$$\int_\gamma \alpha = 2\pi \mathrm{i} \sum_P \operatorname{res}_P \alpha$$

as in Remark 7.13 (b), where the sum is taken over all points P in the interior of γ, and the residue of α at a point P is defined similarly to Remark 7.13 (a).

An additional benefit of this version of the Residue Theorem on manifolds is that we can exchange the roles of the interior and exterior of γ: Consider a differential form α on X with poles at some points (marked P_1 and P_2 in the picture below on the right). If we form the integral $\int_\gamma \alpha$ over a small loop γ that contains none of these points, the result will be 0 by the Residue Theorem. But we can also swap the roles of the interior and exterior of γ (without changing the value of the integral), so that now *all* poles lie in the interior of γ, and the Residue Theorem gives us the sum over all residues of α. Comparing these two results we see that

$$\sum_{P \in X} \operatorname{res}_P \alpha = 0,$$

which is also sometimes called the Residue Theorem (for manifolds) in the literature.



Applying this now to the differential form

$$\alpha = d(\log f) = \frac{f'(z)}{f(z)}\, dz \quad \text{gives us} \quad \sum_{P \in X} \operatorname{res}_P \frac{f'(z)}{f(z)}\, dz = 0,$$

and thus with the same (local) computation $\operatorname{res}_P \frac{f'(z)}{f(z)}\, dz = \mu_P(f)$ as in the proof of Lemma 7.15

$$\sum_{P \in X} \mu_P(f) = 0,$$

i. e. that f has equally many zeros as poles.

But let us now return to our study of the torus $\mathbb{C}/\Lambda$. The key ingredient to identify it with the points of an elliptic curve is the following meromorphic function.

Proposition and Definition 7.17 (The Weierstraß $\wp$-function)**.** *There is a meromorphic function $\wp$ on $\mathbb{C}$, called the **Weierstraß $\wp$-function** (pronounced like the letter "p"), defined by*

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right).$$

It has poles of order 2 exactly at the lattice points.

Proof sketch. It is a standard fact that an (infinite) sum of holomorphic functions is holomorphic at z_0 provided that the sum converges uniformly in a neighborhood of z_0. We will only sketch the proof of this convergence: Let $z_0 \in \mathbb{C}\setminus\Lambda$ be a fixed point that is not in the lattice. Then every summand is a holomorphic function in a neighborhood of z_0. The expansions of these summands for large ω are

$$\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} = \frac{1}{\omega^2}\left(\frac{1}{(1-\frac{z}{\omega})^2} - 1\right) = \frac{2z}{\omega^3} + \left(\text{terms of order at least } \frac{1}{\omega^4}\right),$$

so the summands grow like ω^3. Let us add up these values according to the absolute value of ω. Note that the number of lattice points with a given absolute value approximately equal to $n \in \mathbb{N}$ is roughly proportional to the area of the annulus with inner radius $n-\frac{1}{2}$ and outer radius $n+\frac{1}{2}$, which grows linearly with n. Hence the final sum is of the order $\sum_{n=1}^{\infty} n \cdot \frac{1}{n^3} = \sum_{n=1}^{\infty} \frac{1}{n^2}$, which is convergent.

Note that the sum would not have been convergent without the subtraction of the constant $\frac{1}{\omega^2}$ in each summand, as then the individual terms would grow like $\frac{1}{\omega^2}$, and therefore the final sum would be of the type $\sum_{n=1}^{\infty} \frac{1}{n}$, which is divergent.

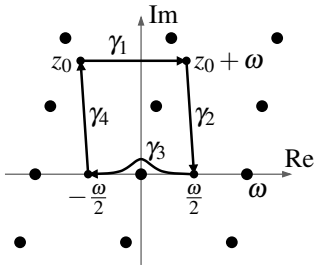
Finally, the poles of order 2 at the points of Λ are clearly visible. □

Remark 7.18 (Properties of the $\wp$-function)**.** One can show that in an absolutely convergent series as above all manipulations (reordering of the summands, term-wise differentiation) can be performed as expected. In particular, the following properties of the $\wp$-function are obvious:

(a) The $\wp$-function is an even function, i. e. $\wp(z) = \wp(-z)$ for all $z \in \mathbb{C}$. Hence its Laurent series at 0 as in Remark 7.13 (a) contains only even exponents.

(b) Its derivative is $\wp'(z) = \sum_{\omega\in\Lambda} \frac{-2}{(z-\omega)^3}$. It is an odd function, i. e. $\wp'(z) = -\wp'(-z)$ for all z. In other words, its Laurent series at 0 contains only odd exponents. It has poles of order 3 exactly at the lattice points.

(c) The $\wp$-function is Λ-periodic, and hence gives a meromorphic function $\wp\colon \mathbb{C}/\Lambda \to \mathbb{C}\cup\{\infty\}$. To show this note first that $\wp'$ is Λ-periodic by (b). Now, for given $z_0 \in \mathbb{C}$ and $\omega \in \Lambda$ we integrate $\wp'$ along the path $\gamma = \gamma_1+\gamma_2+\gamma_3+\gamma_4$ shown in the picture below on the right.



Of course, the result is 0, since $\wp$ is an integral of $\wp'$. But also the integral along γ_2 cancels the integral along γ_4 as $\wp'(z)$ is periodic. The integral along γ_3 is equal to $\wp(-\frac{\omega}{2}) - \wp(\frac{\omega}{2})$, so it vanishes as well since $\wp$ is an even function. So we conclude that

$$0 = \int_{\gamma_1} \wp'(z)\,dz = \wp(z_0+\omega) - \wp(z_0),$$

i. e. that $\wp$ is Λ-periodic.

Lemma 7.19 (Differential equation of the $\wp$-function)**.** *The $\wp$-function satisfies a differential equation*

$$\wp'(z)^2 = c_3\wp(z)^3 + c_2\wp(z)^2 + c_1\wp(z) + c_0 \quad \text{for all } z \in \mathbb{C}$$

for some constants $c_0, c_1, c_2, c_3 \in \mathbb{C}$ (depending on Λ).

Proof sketch. By Remark 7.18 (b) we know that $(\wp')^2$ is an even function with a pole of order 6 at the origin. Hence its Laurent series around 0 is of the form

$$\wp'(z)^2 = \frac{a_{-6}}{z^6} + \frac{a_{-4}}{z^4} + \frac{a_{-2}}{z^2} + a_0 + (\text{terms of positive multiplicity at } 0)$$

for some constants $a_{-6}, a_{-4}, a_{-2}, a_0 \in \mathbb{C}$. The functions $\wp^3$, $\wp^2$, $\wp$, and 1 are also even and have poles at the origin of order 6, 4, 2, and 0, respectively. Hence there are constants $c_3, c_2, c_1, c_0 \in \mathbb{C}$ such that the Laurent series of the linear combination

$$f(z) := \wp'(z)^2 - c_3\wp(z)^3 - c_2\wp(z)^2 - c_1\wp(z) - c_0$$

has only positive powers of z. This means that f is holomorphic around the origin and vanishes at 0.

But $\wp$ and $\wp'$, and hence also f, are Λ-periodic by Remark 7.18 (c). Hence f is holomorphic around all lattice points. Moreover, f is holomorphic around all other points as well, as $\wp$ and $\wp'$ are. Hence f is holomorphic on all of $\mathbb{C}$.

The Λ-periodicity means that every value taken on by f is already assumed on the parallelogram $\{x\omega_1 + y\omega_2 : x, y \in [0,1]\}$. As f is continuous, its image on this compact parallelogram, and hence on all of $\mathbb{C}$, is bounded. So we see by Liouville's Theorem of Remark 7.13 (c) that f must be constant. But as we have already shown that $f(0) = 0$, it follows that f is the zero function, which is exactly the statement of the lemma. □

Remark 7.20. By an explicit computation one can show that the coefficients c_3, c_2, c_1, c_0 in Lemma 7.19 are given by

$$c_3 = 4, \quad c_2 = 0, \quad c_1 = -60 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^4}, \quad \text{and} \quad c_0 = -140 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^6}.$$

The proof of Lemma 7.19 shows impressively the powerful methods of complex analysis: To prove our differential equation, i. e. the equality of the two functions $(\wp')^2$ and $c_3\wp^3 + c_2\wp^2 + c_1\wp + c_0$, it was sufficient to just compare four coefficients of their Laurent expansions at the origin – the rest then follows entirely from general theory.

Note also that the differential equation of Lemma 7.19 is a (non-homogeneous) cubic equation in the two functions $\wp$ and $\wp'$, which are Λ-periodic and thus well-defined on the quotient $\mathbb{C}/\Lambda$. We can therefore use it to obtain a map from $\mathbb{C}/\Lambda$ to an elliptic curve as follows.

Proposition 7.21 (Complex tori as elliptic curves)**.** *Consider the elliptic curve*

$$F = x_2^2 x_0 - c_3 x_1^3 - c_2 x_1^2 x_0 - c_1 x_1 x_0^2 - c_0 x_0^3$$

for the constants $c_3, c_2, c_1, c_0 \in \mathbb{C}$ of Lemma 7.19. There is a bijection

$$\Psi\colon \mathbb{C}/\Lambda \to V(F),\ z \mapsto (1 : \wp(z) : \wp'(z)).$$

Proof sketch. As $\wp$ and $\wp'$ are Λ-periodic and satisfy the differential equation of Lemma 7.19, it is clear that Ψ is well-defined as a map to $V(F)$. Strictly speaking, for $z = 0$ we have to note that $\wp$ and $\wp'$ have poles of order 2 and 3, respectively, so that the given expression for $\Psi(0)$ is of the form $(1 : \infty : \infty)$. But by Remark 7.12 we can write $\wp(z) = \frac{f(z)}{z^2}$ and $\wp'(z) = \frac{g(z)}{z^3}$ locally around the origin for some holomorphic functions f, g that do not vanish at 0, and so we have to interpret the expression for Ψ as

$$\Psi(0) = \lim_{z\to 0} (1 : \wp(z) : \wp'(z)) = \lim_{z\to 0} (z^3 : z f(z) : g(z)) = (0 : 0 : 1),$$

i. e. $\Psi(z)$ is well-defined at $z = 0$ as well.

Now let $(x_0 : x_1 : x_2) \in V(F)$ be a given point; we have to show that it has exactly one inverse image under Ψ. By what we have just said this is obvious for the "point at infinity" $(0 : 0 : 1)$, so let us assume that we are not at this point and hence pass to inhomogeneous coordinates where $x_0 = 1$. We thus have to show that there is exactly one (non-zero) $z \in \mathbb{C}/\Lambda$ with $\wp(z) = x_1$ and $\wp'(z) = x_2$.

Recall that $\wp$, and thus also $\wp - x_1$, has exactly one pole in $\mathbb{C}/\Lambda$, namely the origin, and that this pole is of order 2. Hence $\wp - x_1$ also has exactly two zeros (counted with multiplicities) in $\mathbb{C}/\Lambda$ by

Lemma 7.15, i. e. there are two $z \in \mathbb{C}/\Lambda$ with $\wp(z) = x_1$. For such a point z we then have by Lemma 7.19

$$\wp'(z)^2 = c_3\wp(z)^3 + c_2\wp(z)^2 + c_1\wp(z) + c_0 = c_3x_1^3 + c_2x_1^2 + c_1x_1 + c_0 = x_2^2$$

since $(1 : x_1 : x_2) \in V(F)$. So there are two possibilities:

- $\wp'(z) = 0$: Then $x_2 = 0$ as well, and z is a double zero (i. e. the only zero) of the function $\wp - x_1$. So there is exactly one $z \in \mathbb{C}/\Lambda$ with $\Psi(z) = (1 : \wp(z) : \wp'(z)) = (1 : x_1 : x_2)$.
- $\wp'(z) \neq 0$: Then z is only a simple zero of $\wp - x_1$. As $\wp$ is even and $\wp'$ odd by Remark 7.18, we see that $-z$ must be the other zero, and it satisfies $\wp'(-z) = -\wp'(z)$. Hence exactly one of the equations $\wp'(z) = x_2$ and $\wp'(-z) = x_2$ holds, and the corresponding point is the unique inverse image of $(1 : x_1 : x_2)$ under Ψ.

Altogether we conclude that Ψ is bijective, as we have claimed. □

Remark 7.22. In fact, the map Ψ of Proposition 7.21 is not just a bijection: Both $\mathbb{C}/\Lambda$ and $V(F)$ are 1-dimensional complex manifolds in a natural way, and Ψ is even an isomorphism between these two manifolds.

Remark 7.23 (Group structures on elliptic curves)**.** With Proposition 7.21 we are again in a similar situation as in Proposition 7.4: We have a bijection between a group $\mathbb{C}/\Lambda$ and a variety $V(F)$, so that the map Ψ of the above proposition can be used to construct a group structure on $V(F)$. In fact, we will see in Exercise 7.25 that this group structure is precisely the same as that obtained by the map Φ of Proposition 7.4 using divisors. But the algebraic properties of this group structure is a lot more obvious in this new picture: For example, the points of order n are easily read off to be the n^2 points

$$\frac{1}{n}(i\omega_1 + j\omega_2) \quad \text{for } 0 \leq i, j < n.$$

Exercise 7.24. Let Λ be a lattice in $\mathbb{C}$, and let $P \neq Q$ be points in $\mathbb{C}/\Lambda$. Show that there is no meromorphic function on $\mathbb{C}/\Lambda$ with a simple zero at P, a simple pole at Q, and which is holomorphic with non-zero value at all other points.

Note that we can view this as an analytic analogue of Proposition 6.34 for elliptic curves.

Exercise 7.25. Let F be an elliptic curve corresponding to a torus $\mathbb{C}/\Lambda$ as in Proposition 7.21. Show that the group structure on $V(F)$ induced by $\mathrm{Pic}^0 F$ as in Proposition 7.4 (using $(0:0:1)$ as the base point) is the same as the one induced by the natural group structure of $\mathbb{C}/\Lambda$.

Exercise 7.26. Let $\Lambda \subset \mathbb{C}$ be a lattice. Given two points $z, w \in \mathbb{C}/\Lambda$, it is very easy to find a natural number n such that $n \cdot w = z$ (in the group structure of $\mathbb{C}/\Lambda$), in case such a number exists. Why is this no contradiction to the idea of the cryptographic application in Example 7.9?

12

8. The Riemann-Roch Theorem

In the previous two chapters we have introduced and studied rational and regular functions on projective curves. As our last goal in these notes we now want to address the question *how many* such functions there are on a given projective curve (which, as before, will always be assumed to be smooth over an algebraically closed field).

But before we can try to solve this problem we first have to figure out what the precise question should be, i. e. which functions we want to consider and what exactly we mean by "how many". Note that we know already by Corollary 6.29 that global regular functions on a projective curve are always constant, and thus not very interesting. On the other hand, to obtain arbitrary rational functions we can take any quotient of two homogeneous polynomials of the same degree, so that we clearly get an infinite-dimensional vector space of such functions. Hence the most interesting question is to study something between regular and rational functions: rational functions which are everywhere regular, except for some specific points at which we allow poles of a given maximal order (or require zeros of a certain order). We will see that such functions form finite-dimensional vector spaces, so that we can then ask for their dimensions.

The conditions of allowing poles or requiring zeros at specified points is described best using the language of divisors. This leads to the following spaces that we will consider in this chapter.

Construction 8.1 ($L(D)$ and $l(D)$)**.** Let D be a divisor on a projective curve F. We set

$$L(D) := \{\varphi \in K(F)^* : \operatorname{div}\varphi + D \geq 0\} \cup \{0\}.$$

If $D = \sum_{P \in F} a_P \cdot P$, i. e. a_P denotes the coefficient of P in D, the condition $\operatorname{div}\varphi + D \geq 0$ obviously means $\mu_P(\varphi) + a_P \geq 0$, i. e. $\mu_P(\varphi) \geq -a_P$, for all points $P \in F$. Hence, except for the zero function, $L(D)$ consists by construction of all rational functions $\varphi \in K(F)^*$ that are just regular at all points of F, except that

(a) φ may have a pole of order at most a_P at P for all P with $a_P > 0$, and

(b) φ must have a zero of order at least $-a_P$ at P for all P with $a_P < 0$.

Note that $L(D)$ is a vector space over K: For all $\lambda \in K^*$ and $\varphi, \psi \in L(D)$, i. e. such that $\mu_P(\varphi) \geq -a_P$ and $\mu_P(\psi) \geq -a_P$ for all $P \in F$, we have

$$\mu_P(\varphi + \psi) \geq -a_P \text{ by Corollary 6.12} \quad \text{and} \quad \mu_P(\lambda\varphi) = \mu_P(\varphi) \geq -a_P \text{ by Construction 6.6 (b)}$$

for all P, and thus $\varphi + \psi \in L(D)$ and $\lambda\varphi \in L(D)$. Hence we can define

$$l(D) := \dim L(D) \quad \in \mathbb{N} \cup \{\infty\}.$$

As motivated above, it is the goal of this chapter to compute these dimensions $l(D)$. Unfortunately, we will not be able to do this for all D, since in general $l(D)$ depends on the precise position of the points occurring in D in a complicated way. However, the Riemann-Roch Theorem in Corollary 8.17 will allow to compute $l(D)$ in many cases just from the degree of D, which is of course easy to read off. The formula will also involve the *genus* of the curve – a concept that we have already seen over $\mathbb{C}$ from a topological point of view in Remark 5.12. As a byproduct of our work, we will therefore also give an algebraic definition of the genus of a curve, which is then applicable to any algebraically closed ground field.

But let us start with a few simple examples in which $l(D)$ is easy to determine.

Example 8.2.

(a) For the divisor $D = 0$ the space $L(D) = L(0)$ is by definition just the set of all rational functions that are regular at every point of the curve. Hence by Corollary 6.29 we have $L(0) = K$, and thus $l(0) = 1$.

(b) For any divisor D with $\deg D < 0$ we have $L(D) = \{0\}$ and thus $l(D) = 0$: If there was a non-zero element $\varphi \in L(D)$ we would have $\operatorname{div}\varphi + D \geq 0$, and hence $\deg \operatorname{div}\varphi + \deg D \geq 0$ by taking degrees. But this is a contradiction to $\deg D < 0$ since $\deg \operatorname{div}\varphi = 0$ by Remark 6.27 (b).

Remark 8.3. Let D be a divisor on a projective curve F.

(a) If D' is another divisor on F with $D \leq D'$ then $L(D) \subset L(D')$ and hence $l(D) \leq l(D')$, since $\operatorname{div}\varphi + D \geq 0$ clearly implies $\operatorname{div}\varphi + D' \geq \operatorname{div}\varphi + D \geq 0$.

(b) If $D' \sim D$ is linearly equivalent, i. e. $D - D' = \operatorname{div}\psi$ for a rational function $\psi \in K(F)^*$, then $L(D) \to L(D')$, $\varphi \mapsto \psi\varphi$ is an isomorphism of vector spaces (with inverse $\varphi \mapsto \frac{\varphi}{\psi}$) since the condition $\operatorname{div}\varphi + D \geq 0$ is equivalent to $\operatorname{div}(\psi\varphi) + D' \geq 0$. Hence we have $l(D) = l(D')$ in this case.

In particular, the notion $l(\cdot)$ is also well-defined for elements of the Picard group $\operatorname{Pic} F$. In the following, we will also use it in this extended way.

Many of our strategies to compute the numbers $l(D)$ will be inductive, i. e. relate $l(D)$ to $l(D \pm P)$ for a point P on the curve. Of particular importance will therefore be the following result, which tells us that $l(D)$ changes at most by 1 when adding or subtracting a point from D.

Lemma 8.4. *Let D be a divisor on a projective curve F.*

(a) *For any point $P \in F$ we have $l(D+P) = l(D)$ or $l(D+P) = l(D)+1$.*

(b) *For any divisor $D' \geq D$ we have $l(D) \leq l(D') \leq l(D) + \deg(D' - D)$.*

Proof.

(a) As $D \leq D + P$ we have $L(D) \subset L(D+P)$, and hence $l(D) \leq l(D+P)$, by Remark 8.3 (a).

Now let a_P be the coefficient of P in D, so that $a_P + 1$ is the coefficient of P in $D + P$. Consider the linear map

$$\Phi\colon L(D+P) \to K,\ \varphi \mapsto (t^{a_P+1}\varphi)(P),$$

where t is a local coordinate around P as in Proposition 6.10. Note that this evaluation of $t^{a_P+1}\varphi$ at P is well-defined, since for $\varphi \in L(D+P)\backslash\{0\}$ we have

$$\mu_P(t^{a_P+1}\varphi) = \mu_P(\varphi) + a_P + 1 \geq 0 \qquad (*)$$

(where the last inequality follows from Construction 8.1), so that $t^{a_P+1}\varphi$ is regular at P by Proposition 6.10 (b).

The kernel of Φ consists exactly of the rational functions for which $t^{a_P+1}\varphi$ has a zero at P, i. e. for which we have strict inequality in $(*)$. As this is equivalent to $\mu_P(\varphi) + a_P \geq 0$ and thus to $\operatorname{div}\varphi + D \geq 0$, we conclude that $\ker\Phi = L(D)$. The homomorphism theorem thus yields

$$L(D+P)/L(D) \cong \operatorname{im}\Phi \quad \subset K,$$

which means that $l(D+P) = l(D)$ (in case $\operatorname{im}\Phi = \{0\}$) or $l(D+P) = l(D)+1$ (in case $\operatorname{im}\Phi = K$).

(b) This follows immediately from (a) by induction on $\deg(D' - D)$, since D' is obtained from D by adding $\deg(D' - D)$ points. □

Remark 8.5. The proof of Lemma 8.4 (a) also has a simple analytic interpretation in case of the ground field $K = \mathbb{C}$. As the multiplicity of a rational function $\varphi \in L(D+P)$ at P is at least $-a_P - 1$, its Laurent expansion as in Remark 7.13 (a) can be taken to start with the power t^{-a_P-1} of an (analytic) local coordinate t. Inside $L(D+P)$, the subspace $L(D)$ now consists of exactly those functions for which the t^{-a_P-1}-coefficient of this expansion vanishes. As this coefficient is one complex number, its vanishing imposes one condition on $L(D+P)$ – which can be trivially satisfied by all elements of $L(D+P)$ already (in which case $l(D) = l(D+P)$) or not (in which case $l(D) = l(D+P) - 1$).

Corollary 8.6. *For any divisor D with $\deg D \geq 0$ on a projective curve F we have $l(D) \leq \deg D + 1$. In particular, the number $l(D)$ is always finite.*

Proof. Let $n = \deg D + 1$, and choose a point $P \in F$. Then $\deg(D - nP) = \deg D - n = -1 < 0$, so that $l(D - nP) = 0$ by Example 8.2 (b). It follows by Lemma 8.4 (b) that

$$l(D) \leq l(D - nP) + \deg(nP) = 0 + n = \deg D + 1. \qquad \square$$

Example 8.7.

(a) Let D be a divisor with $\deg D \geq 0$ on a projective curve F of degree 1 or 2. We claim that then $l(D) = \deg D + 1$, i. e. that we have equality in Corollary 8.6. In particular, together with Example 8.2 (b) this finishes the computation of all $l(D)$ on curves of degree 1 or 2.

To prove this, recall that $\mathrm{Pic}^0 F = \{0\}$ by Example 6.33, and hence $\mathrm{Pic} F \cong \mathbb{Z}$ by Remark 6.32, with an isomorphism given by the degree of divisors. If we pick any two distinct points $P, Q \in F$ this means first of all that $D \sim nP$ with $n := \deg D$. Moreover, as $P \sim Q$ there is a rational function $\varphi \in K(F)^*$ with $\operatorname{div} \varphi = Q - P$. We then have $\operatorname{div} \varphi^k = kQ - kP$ and hence $\varphi^k \in L(kP) \backslash L((k-1)P)$ for all $k \in \mathbb{N}_{>0}$, so that the inclusions

$$K \overset{8.2\,(a)}{=} L(0) \subset L(P) \subset L(2P) \subset \cdots \subset L(nP)$$

of Remark 8.3 (a) are all strict. Taking dimensions, we conclude that $l(nP) \geq n + 1$, hence in fact $l(nP) = n + 1$ by Corollary 8.6, and thus $l(D) = \deg D + 1$ by Remark 8.3 (b).

(b) Let P be a point on a projective curve F of degree at least 3. We will show that then $l(P) = 1$, i. e. that in this case we have a strict inequality in Corollary 8.6.

Consider any non-zero element $\varphi \in L(P)$. By definition, this rational function may then have a pole of order 1 at P but must be regular at all other points of F, so that $\operatorname{div} \varphi = Q - P$ for some point Q by Remark 6.27 (b). But by Proposition 6.34 this is impossible unless $Q = P$, which means that φ is a constant. Conversely, the constant functions are clearly contained in $L(P)$, and thus we see that $L(P) = K$, i. e. that $l(P) = 1$.

(c) Now consider the divisor $P - Q$ for two distinct points P and Q on a projective curve of degree at least 3. By (b) and Remark 8.3 (a) we have $L(P - Q) \subset L(P) = K$, so the elements of $L(P - Q)$ must be constant functions. But a constant does not have a zero at Q unless it is 0. Hence we see that $L(P - Q) = \{0\}$, and thus $l(P - Q) = 0$.

Exercise 8.8. Let F be a projective curve of degree d; without loss of generality we may assume that $F \neq z$. As usual, we will denote the vector space of homogeneous polynomials in x, y, z of degree n by $K[x, y, z]_n$.

For all $n \geq d$, show for the divisor $D := n \operatorname{div} z$:

(a) There is an exact sequence

$$0 \longrightarrow K[x, y, z]_{n-d} \xrightarrow{\cdot F} K[x, y, z]_n \xrightarrow{:z^n} L(D) \longrightarrow 0.$$

(b) $l(D) = \deg D + 1 - \binom{d-1}{2}$.

Remark 8.9 ($l(D)$ does not only depend on $\deg D$)**.** Note that on a projective curve F of degree at least 3 we have by Examples 8.2 (a) and 8.7 (c)

$$l(0) = 1 \quad \text{and} \quad l(P - Q) = 0$$

for any two distinct points $P, Q \in F$. In particular, as both divisors 0 and $P - Q$ have degree 0 we see that in general *the value $l(D)$ does not depend on the degree $\deg D$ alone*, but also on the exact positions of the points in D.

However, complementing the upper bound for $l(D)$ of Corollary 8.6 we can now also give a lower bound that depends only on $\deg D$. In fact, the difference between these two bounds turns out to be exactly the genus of the curve that we have already seen over $\mathbb{C}$ in Remark 5.12. We will use this observation as the definition of the genus in the algebraic setting.

Proposition and Definition 8.10. *Let F be a projective curve of degree d.*

(a) ***(Riemann's Theorem)*** *There is a unique smallest integer g, depending only on F, such that*

$$l(D) \geq \deg D + 1 - g \qquad (*)$$

for any divisor D. We call g the ***(algebraic) genus*** *of F.*

(b) ***(Algebraic degree-genus formula)*** *The algebraic genus of F is given by $g = \binom{d-1}{2}$.*

In particular, for $K = \mathbb{C}$ it coincides with the topological genus of Remark 5.12 and Proposition 5.16.

Proof. If we set $g = \binom{d-1}{2}$, Exercise 8.8 shows that there are divisors on F for which $(*)$ holds with equality. Hence, to prove both parts of the proposition, it suffices to prove that the inequality $(*)$ is true for every divisor D on F. To show this, note first:

(1) If $(*)$ holds for any divisor D, it also holds for any linearly equivalent divisor $D' \sim D$, since by Remarks 6.27 (b) and 8.3 (b) both sides of the inequality do not change when passing from D to D'.

(2) If $(*)$ holds for any divisor D, it also holds for any divisor $D' \leq D$: From $l(D) \geq \deg D + 1 - g$ it follows that

$$l(D') \overset{8.4}{\geq} l(D) - \deg(D - D') \geq \deg D + 1 - g - \deg(D - D') = \deg D' + 1 - g.$$

Now let D be any divisor on F, which we can write as $D = P_1 + \cdots + P_n - E$ for some points $P_1, \ldots, P_n \in F$ and an effective divisor E. As the points $P_1, \ldots, P_n$ are allowed to appear in E we may assume in this representation that $n \geq d$. For every $i = 1, \ldots, n$ choose a line l_i through P_i (which is not equal to F). Then the divisor

$$D' := D + \operatorname{div} \frac{z^n}{l_1 \cdot \cdots \cdot l_n}$$

is linearly equivalent to D, and satisfies

$$D' = P_1 + \cdots + P_n - E + n \operatorname{div} z - \sum_{i=1}^{n} \operatorname{div} l_i \leq P_1 + \cdots + P_n - E + n \operatorname{div} z - P_1 - \cdots - P_n \leq n \operatorname{div} z$$

since $\operatorname{div} l_i \geq P_i$ for all i. But now $(*)$ holds for $\operatorname{div} z^n$ by Exercise 8.8, hence also for D' by (2), and thus for D by (1). □

Summarizing, we now know by Corollary 8.6 and Proposition 8.10 (a) that

$$\deg D + 1 - g \leq l(D) \leq \deg D + 1$$

for every divisor D with $\deg D \geq 0$ on a projective curve F of genus g. We have also seen in Remark 8.9 already that we cannot expect an exact formula for $l(D)$ in terms of $\deg D$ alone. Nevertheless, one can make the above inequalities into an equality: It turns out that for every divisor D the difference between $l(D)$ and $\deg D + 1 - g$ can be identified as $l(D')$ for another divisor D' that is easily computable from D. To show this, we need the following special divisor on F.

Definition 8.11 (Canonical divisor)**.** Let F be a projective curve of degree d. For any line l (not equal to F) we call

$$K_F := (d-3) \operatorname{div} l \quad \in \operatorname{Pic} F$$

the **canonical divisor (class)** of F. (Note that for the element of $\operatorname{Pic} F$ it does not matter which line we take: For any other line l' we have $\operatorname{div} l \sim \operatorname{div} l'$ as $\operatorname{div} \frac{l}{l'} \in \operatorname{Prin} F$.)

Remark 8.12 (Canonical divisors are canonical)**.** It is hard to deny that our definition of the canonical divisor K_F of a projective curve F looks very artificial: It is not clear why divisor classes of lines and the choice of factor $d-3$ should lead to an object that plays a special role for F.

In fact, the usual definition of canonical divisors of curves in the literature is entirely different and much more natural (i. e. "canonical"): One can introduce *differential forms* on F in a way similar to the complex analytic setting in Remark 7.16, i. e. formal expressions of the form $\alpha = f\,dg$ for

rational functions f and g that satisfy the usual rules of differentiation. They are natural objects on F that do not require any choices to define them, and in the same way as for rational functions one can associate multiplicities $\mu_P(\alpha)$ to a differential form α at a point $P \in F$. Combining these multiplicities for all points $P \in F$ one obtains a divisor $\operatorname{div}\alpha \in \operatorname{Div} F$, again in the same way as for rational functions.

It turns out that the divisors of any two differential forms are linearly equivalent, so that we obtain a well-defined and natural element K_F of $\operatorname{Pic} F$ as the divisor class of any differential form. This is the usual definition of the canonical divisor class K_F. It is then a computation to show that in the case of a projective plane curve this canonical divisor is equal to the one of Definition 8.11. We just took this formula as a definition of K_F in order to avoid a detailed discussion of differential forms.

13

Lemma 8.13 (Degree of the canonical divisor)**.** *For any projective curve F of genus g we have* $\deg K_F = 2g-2$.

Proof. By Remark 6.27 (a) we have for a curve F of degree d

$$\deg K_F = (d-3)\deg \operatorname{div} l = (d-3)\,d = 2\binom{d-1}{2} - 2,$$

so the result follows from the degree-genus formula $g = \binom{d-1}{2}$ of Proposition 8.10 (b). □

The key property of the canonical divisor that will allow us to make Riemann's Theorem of Proposition 8.10 into an equality is the following.

Lemma 8.14. *For any point P on a projective curve F we have $l(K_F + P) = l(K_F)$.*

Proof. If $d := \deg F \leq 2$ then $g = 0$ by Proposition 8.10 (b), and hence $\deg K_F = -2$ by Lemma 8.13. So in this case the degrees of both K_F and $K_F + P$ are negative, which means by Example 8.2 (b) that $l(K_F + P) = l(K_F) = 0$. We can therefore assume from now on that $d \geq 3$.

Choose any line l through P that is not the tangent $T_P F$. The divisor $\operatorname{div} l - P$ is then effective and does not contain P. Moreover, in this proof we will use this line l in Definition 8.11 to regard K_F as a divisor (and not just a divisor class). It then clearly suffices to prove that $L(K_F + P) = L(K_F)$. By Remark 8.3 (a) the inclusion "$\supset$" is automatic, so we will show "$\subset$".

To do this, let $\varphi = \frac{f}{g}$ be a non-zero element of $L(K_F + P)$, so that $\operatorname{div}\varphi + K_F + P \geq 0$. By Definition 8.11 this can be rewritten as

$$\operatorname{div}(f l^{d-2}) \geq \operatorname{div} g + \operatorname{div} l - P \geq \operatorname{div} g.$$

Max Noether's Theorem as in Proposition 6.28 then implies that there is a homogeneous polynomial h of degree $d-2$ with

$$\operatorname{div} h = \operatorname{div}(f l^{d-2}) - \operatorname{div} g \geq \operatorname{div} l - P. \qquad (*)$$

For all $Q \neq P$ this means that $\mu_Q(F,h) \geq \mu_Q(F,l)$, and hence $\langle F,h\rangle \subset \langle F,l\rangle$ in $\mathscr{O}_{\mathbb{P}^2,Q}$ by Proposition 2.26. But then also $\langle l,h\rangle \subset \langle F,l\rangle$ in $\mathscr{O}_{\mathbb{P}^2,Q}$, which in turn yields $\mu_Q(l,h) \geq \mu_Q(F,l)$. Taking the sum of these numbers for all $Q \neq P$ we get

$$\sum_{Q\neq P} \mu_Q(l,h) \geq \sum_{Q\neq P} \mu_Q(F,l) = \deg(\operatorname{div} l - P) = d-1.$$

But as h has degree $d-2$, Bézout's Theorem as in Corollary 4.6 implies that h must contain l as a factor. Hence we have $\operatorname{div} h \geq \operatorname{div} l$, and so by $(*)$

$$\operatorname{div}(f l^{d-2}) - \operatorname{div} g \geq \operatorname{div} l,$$

which means that $\operatorname{div}\varphi + K_F \geq 0$, and thus $\varphi \in L(K_F)$. □

Remark 8.15. Over the complex numbers, Lemma 8.14 is just a simple consequence of the Residue Theorem: In Remark 7.16 we have already seen that the sum of the residues of a differential form α on a projective curve F is 0. In particular, it follows that α cannot have exactly one non-zero residue, and thus that it is impossible for α to have exactly one pole at a point P which is in addition of order

1 (since by definition the residue would then be non-zero there). Applying this to the differential form $\varphi\alpha$ for any rational function φ this means in the language of divisors that

$$\operatorname{div}(\varphi\alpha)+P\geq 0 \quad \text{implies} \quad \operatorname{div}(\varphi\alpha)\geq 0,$$

i. e. that

$$\operatorname{div}\varphi+\operatorname{div}\alpha+P\geq 0 \quad \text{implies} \quad \operatorname{div}\varphi+\operatorname{div}\alpha\geq 0.$$

But by Construction 8.1 this is just the same as saying that $L(\operatorname{div}\alpha+P)=L(\operatorname{div}\alpha)$. Hence we have $l(\operatorname{div}\alpha+P)=l(\operatorname{div}\alpha)$ – which is exactly the statement of Lemma 8.14 since $\operatorname{div}\alpha=K_F$ by Remark 8.12.

Using Lemma 8.14 we can now finally add an additional "correction term" to the inequality in Riemann's Theorem of Proposition 8.10 to make it into an equality. Surprisingly, it turns out that it essentially suffices to prove that the inequality still holds after adding the correction term, with equality then following from this very easily.

Lemma 8.16. *Let F be a projective curve of genus g. Then*

$$l(D)-l(K_F-D)\geq \deg D+1-g$$

for all divisors D on F.

Proof. We will prove the statement by descending induction on $\deg D$. For the start of the induction, note that for all divisors with $\deg D>2g-2$ we have $\deg(K_F-D)<0$ by Lemma 8.13, hence $l(K_F-D)=0$ by Example 8.2 (b), and so the statement is just Riemann's Theorem of Proposition 8.10.

For the induction step assume that the statement holds for a divisor D; we will show that it holds for $D-P$ for any point $P\in F$. As we already know that

$$\begin{aligned} l(D-P)-l(K_F-D+P) &\geq l(D)-1-(l(K_F-D)+1) && \text{(Lemma 8.4)}\\ &\geq \deg D+1-g-2 && \text{(induction assumption)}\\ &= \deg(D-P)-g \end{aligned}$$

it suffices to prove that the first inequality in this computation is strict. So assume for a contradiction that it is not, i. e. that $l(D-P)=l(D)-1$ and $l(K_F-D+P)=l(K_F-D)+1$. By Remark 8.3 (a) this means that $L(D-P)\subsetneq L(D)$ and $L(K_F-D)\subsetneq L(K_F-D+P)$, i. e. that there are rational functions

$$\begin{aligned} &\varphi\in L(D)\backslash L(D-P), && \text{i. e. } \operatorname{div}\varphi+D\geq 0 \text{ with equality at } P,\\ \text{and}\quad &\psi\in L(K_F-D+P)\backslash L(K_F-D), && \text{i. e. } \operatorname{div}\psi+K_F-D+P\geq 0 \text{ with equality at } P, \end{aligned}$$

where "equality at P" means that the point P appears with coefficient 0 on the left hand side of the inequalities. But then multiplying these two functions we obtain

$$\operatorname{div}(\varphi\psi)+K_F+P\geq 0 \text{ with equality at } P, \quad \text{i. e. } \varphi\psi\in L(K_F+P)\backslash L(K_F)$$

in contradiction to Lemma 8.14. □

Corollary 8.17 (Riemann-Roch). *Let D be a divisor on a projective curve F of genus g. Then*

$$l(D)-l(K_F-D)=\deg D+1-g.$$

Proof. Applying Lemma 8.16 to the divisor K_F-D we obtain

$$l(K_F-D)-l(D)\geq \deg(K_F-D)+1-g \overset{8.13}{=} 2g-2-\deg D+1-g,$$

or in other words

$$l(D)-l(K_F-D)\leq \deg D+1-g.$$

Combining this with the statement of Lemma 8.16 for the divisor D yields immediately the desired equation. □

Remark 8.18.

(a) For the divisor $D = 0$ we have $l(0) = 1$ by Example 8.2 (a). We thus get from Corollary 8.17

$$1 - l(K_F) = \deg 0 + 1 - g,$$

and hence $g = l(K_F)$. Sometimes in the literature this is taken as the definition of the genus of a projective curve.

(b) If D is a divisor on a projective curve F of genus g with $\deg D > 2g - 2$, then the divisor $K_F - D$ has negative degree by Lemma 8.13, so that $l(K_F - D) = 0$ by Example 8.2 (b), and thus we get by the Riemann-Roch Theorem

$$l(D) = \deg D + 1 - g.$$

Hence in this case of a divisor of large enough degree we can actually compute the dimension $l(D)$ just in terms of the degree of D. In fact, most applications of the Riemann-Roch theorem will just use this weaker statement.

Note that for curves of genus 0 this statement just reproduces our result for projective curves of degree at most 2 from Example 8.7 (a).

Exercise 8.19 ($l(D)$ for elliptic curves)**.** Let D be a divisor on an elliptic curve F, and denote by $\oplus$ the group structure of Chapter 7. Show that $l(D)$ is given by the following rules:

(a) If $\deg D < 0$ then $l(D) = 0$.

(b) If $\deg D > 0$ then $l(D) = \deg D$.

(c) If $\deg D = 0$ we can write $D = P_1 + \cdots + P_n - Q_1 - \cdots - Q_n$ for some $n \in \mathbb{N}$ and $P_1, \ldots, P_n, Q_1, \ldots, Q_n \in F$, and we have

$$l(D) = \begin{cases} 1 & \text{if } P_1 \oplus \cdots \oplus P_n = Q_1 \oplus \cdots \oplus Q_n, \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 8.20. Let P be a point on a projective curve F. Prove that there is a rational function on F that has a pole (of any order) at P, and is regular at all other points of F.

14

References

[F] W. Fulton, *Algebraic Curves*, http://www.math.lsa.umich.edu/~wfulton/CurveBook.pdf

[G1] A. Gathmann, *Algebraische Strukturen*, class notes TU Kaiserslautern (2023), https://agag-gathmann.math.rptu.de/ags

[G2] A. Gathmann, *Grundlagen der Mathematik*, class notes RPTU Kaiserslautern (2025/26), https://agag-gathmann.math.rptu.de/gdm

[G3] A. Gathmann, *Einführung in die Algebra*, class notes TU Kaiserslautern (2010/11), https://agag-gathmann.math.rptu.de/algebra

[G4] A. Gathmann, *Einführung in die Funktionentheorie*, class notes TU Kaiserslautern (2021/22), https://agag-gathmann.math.rptu.de/futheo

[G5] A. Gathmann, *Einführung in die Topologie*, class notes TU Kaiserslautern (2023), https://agag-gathmann.math.rptu.de/topo

[G6] A. Gathmann, *Commutative Algebra*, class notes TU Kaiserslautern (2014), https://agag-gathmann.math.rptu.de/commalg

[Ki] F. Kirwan, *Complex Algebraic Curves*, Cambridge University Press (1995)

[Ku] E. Kunz, *Introduction to Plane Algebraic Curves*, Birkhäuser (2005)

[W] Wikipedia entry *Curve25519* (2023), https://en.wikipedia.org/wiki/Curve25519

Index