# Algebraic Geometry

Andreas Gathmann

Class Notes TU Kaiserslautern 2014

# Contents

# 0. Introduction

In this introductory chapter we will explain in a very rough sketch what algebraic geometry is about and what it can be used for. We will stress the many correlations with other fields of research, such as complex analysis, topology, differential geometry, singularity theory, computer algebra, commutative algebra, number theory, enumerative geometry, and even theoretical physics. The goal of this chapter is just motivational; you will not find definitions or proofs here (and maybe not even a mathematically precise statement). In the same way, the exercises in this chapter are not designed to be solved in a mathematically precise way. Rather, they are just given as some "food for thought" if you want to think a little further about the examples presented here.

To start from something that you probably know, we can say that algebraic geometry is the combination of *linear algebra* and *algebra*:

- In linear algebra (as e. g. in the "Foundations of Mathematics" class [G2]), we study systems of linear equations in several variables.
- In algebra (as e. g. in the "Introduction to Algebra" class [G3]), one of the main topics is the study of polynomial equations in one variable.

Algebraic geometry combines these two fields of mathematics by studying systems of polynomial equations in several variables.

Given such a system of polynomial equations, what sort of questions can we ask? Note that we cannot expect in general to write down explicitly all the solutions: we know from algebra that even a single complex polynomial equation of degree $d > 4$ in one variable can in general not be solved exactly [G3, Problem 0.2]. So we are more interested in statements about the geometric structure of the set of solutions. For example, in the case of a complex polynomial equation of degree $d$, even if we cannot compute the solutions we know that there are exactly $d$ of them (if we count them with the correct multiplicities). Let us now see what sort of "geometric structure" we can find in polynomial equations in several variables.

**Example 0.1.** Probably the easiest example that is covered neither in linear algebra nor in algebra is that of a single polynomial equation in two variables. Let us consider the example

$$C_n = \{(x_1,x_2) \in \mathbb{C}^2 : x_2^2 = (x_1-1)(x_1-2)\cdots(x_1-2n)\} \quad \subset \mathbb{C}^2,$$

where $n \in \mathbb{N}_{>0}$. Note that in this case it is actually possible to write down all the solutions, because the equation is (almost) solved for $x_2$ already: we can pick $x_1$ to be any complex number, and then get two values for $x_2$ — unless $x_1 \in \{1,\dots,2n\}$, in which case there is only one value for $x_2$ (namely 0).

So it seems that $C_n$ looks like two copies of the complex plane, with the two copies of each point $1,\dots,2n$ identified: the complex plane parametrizes the values for $x_1$, and the two copies of it correspond to the two possible values for $x_2$, i. e. the two roots of the number $(x_1-1)\cdots(x_1-2n)$.

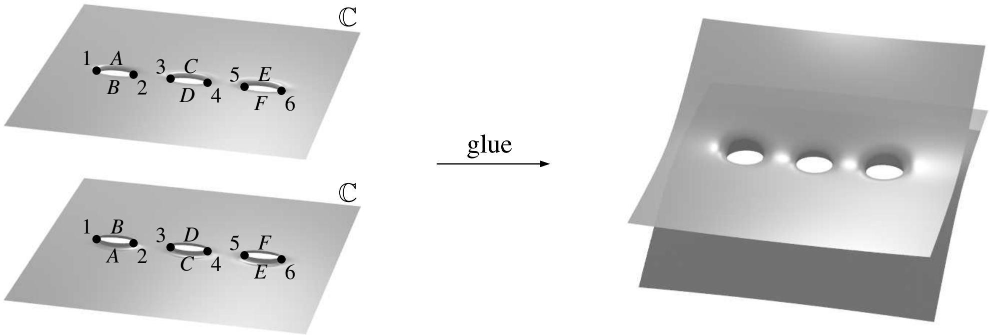
This is not the correct topological picture however, because a complex non-zero number does not have a distinguished first and second root that could correspond to the first and second copy of the complex plane. Rather, the two roots of a complex number get exchanged if you run around the origin once: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

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

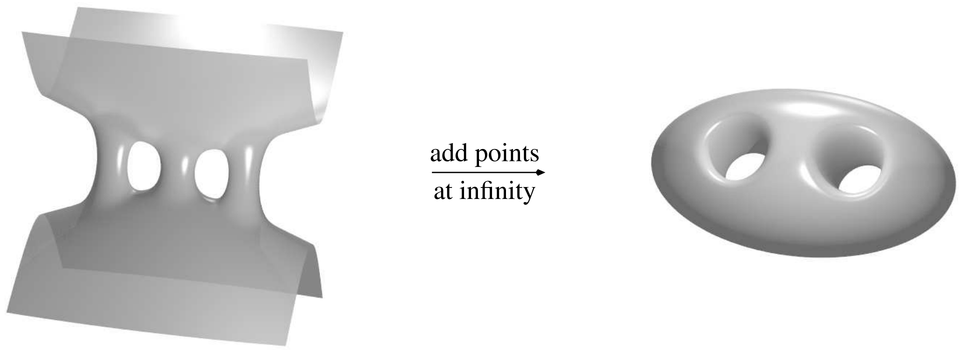
around the complex origin, the square root of this number would have to be defined by

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

which gives opposite values at $\varphi = 0$ and $\varphi = 2\pi$. In other words, if in $C_n$ we run around one of the points $1, \dots, 2n$ (i. e. around a point at which $x_2$ is the square root of 0), we go from one copy of the plane to the other. The way to draw this topologically is to cut the two planes along the real intervals $[1,2], \dots, [2n-1, 2n]$, and to glue the two planes along these lines as in this picture for $n = 3$ (lines marked with the same letter are to be identified):




To make the picture a little nicer, we can compactify our set by adding two points at infinity — one for each copy of the plane — in the same way as we can compactify the complex plane $\mathbb{C}$ by adding a point $\infty$. The precise construction of this compactification will be given in Example 5.6. If we do this here, we end up with a compact surface with $n-1$ handles:

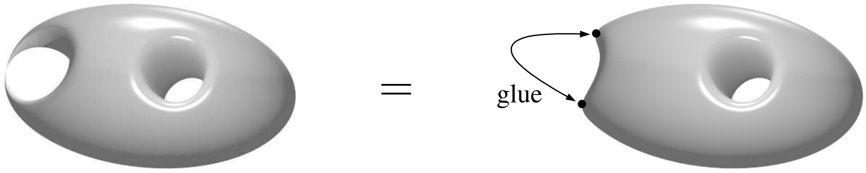



Such an object is called a surface *of genus* $n-1$; the example above shows a surface of genus 2.

**Example 0.2.** What happens in the previous Example 0.1 if we move the points $1, \dots, 2n$ for $x_1$ at which we have only one value for $x_2$, i. e. if we consider

$$C_n = \{(x_1, x_2) \in \mathbb{C}^2 : x_2^2 = f(x_1)\} \quad \subset \mathbb{C}^2$$

with $f$ some polynomial in $x_1$ of degree $2n$? Obviously, as long as the $2n$ roots of $f$ are still distinct, the topological picture above does not change. But if two of the roots approach each other and finally coincide, this has the effect of shrinking one of the tubes connecting the two planes until it reduces to a "singular point" (also called a *node*), as in the following picture on the left:




Obviously, we can view this as a surface with one handle less, where in addition we identify two of the points (as illustrated in the picture on the right). Note that we can still see the "handles" when we draw the surface like this, just that one of the handles results from the gluing of the two points.

Note that our examples so far were a little "cheated" because we said before that we want to figure out the geometric structure of equations that we cannot solve explicitly. In the examples above however, the polynomial equation was chosen so that we could solve it, and in fact we used this solution to construct the geometric picture. Let us now see what we can still do if we make the polynomial more complicated.

**Example 0.3.** Let $d \in \mathbb{N}_{>0}$, and consider

$$C_d = \{(x_1,x_2) \in \mathbb{C}^2 : f(x_1,x_2) = 0\} \quad \subset \mathbb{C}^2,$$

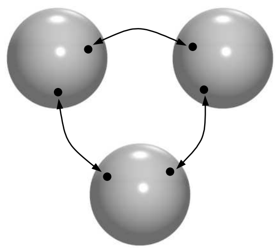
where $f$ is an arbitrary polynomial of degree $d$. This is an equation that we certainly cannot solve directly if $f$ is sufficiently general. Can we still deduce the geometric structure of $C_d$?

In fact, we can do this with the idea of Example 0.2. We saw there that the genus of the surface does not change if we perturb the polynomial equation, even if the surface acquires singular points (provided that we know how to compute the genus of such a singular surface). So why not deform the polynomial $f$ to something singular that is easier to analyze? Probably the easiest thing that comes into mind is to degenerate the polynomial $f$ of degree $d$ into a product of $d$ linear equations $l_1,\dots,l_d$: consider

$$C'_d = \{(x_1,x_2) \in \mathbb{C}^2 : l_1(x_1,x_2) \cdot \cdots \cdot l_d(x_1,x_2) = 0\} \quad \subset \mathbb{C}^2,$$

which should have the same "genus" as the original $C_d$.

It is easy to see what $C'_d$ looks like: of course it is just a union of $d$ complex planes. Any two of them intersect in a point, and we can certainly choose them so that no three of them intersect in a point. The picture below shows $C'_d$ for $d = 3$ (note that every complex plane is — after compactifying it with a point at infinity — just a sphere).



What is the genus of this surface? In the picture above it is obvious that we have one loop; so if $d = 3$ we get a surface of genus 1. In the general case we have $d$ spheres, and every two of them connect in a pair of points, so in total we have $\binom{d}{2}$ connections. But $d-1$ of them are needed to glue the $d$ spheres to a connected chain without loops; only the remaining ones then add a handle each. So the genus of $C'_d$ (and hence of $C_d$) is

$$\binom{d}{2} - (d-1) = \binom{d-1}{2}.$$

This is commonly called the *degree-genus formula* for complex plane curves. We will show it in Proposition 13.17.

**Remark 0.4** (Real vs. complex dimension)**.** One of the trivial but common sources for misunderstandings is whether we count dimensions over $\mathbb{C}$ or over $\mathbb{R}$. The examples considered above are *real surfaces* (the dimension over $\mathbb{R}$ is 2), but *complex curves* (the dimension over $\mathbb{C}$ is 1). We have used the word "surface" so far as this fitted best to the pictures that we have drawn. When looking at the theory however, it is usually more natural to call these objects (complex) curves. In what follows, we always mean the dimension over $\mathbb{C}$ unless stated otherwise.

**Exercise 0.5.** What do we get in Example 0.1 if we consider the equation

$$C'_n = \{(x_1,x_2) \in \mathbb{C}^2 : x_2^2 = (x_1-1)(x_1-2)\cdots(x_1-(2n-1))\} \subset \mathbb{C}^2$$

for $n \in \mathbb{N}_{>0}$ instead?

**Exercise 0.6.** In Example 0.3, we argued that a polynomial of degree $d$ in two complex variables gives rise to a surface of genus $\binom{d-1}{2}$. In Example 0.1 however, a polynomial of degree $2n$ gave us a surface of genus $n-1$. Can you see why these two results do not contradict each other?

**Remark 0.7.** Here is what we should learn from the examples considered so far:

- Algebraic geometry can make statements about the topological structure of objects defined by polynomial equations. It is therefore related to *topology* and *differential geometry* (where similar statements are deduced using analytic methods).
- The geometric objects considered in algebraic geometry need not be "smooth" (i. e. they need not be *manifolds*). Even if our primary interest is in smooth objects, degenerations to singular objects can greatly simplify a problem (as in Example 0.3). This is a main point that distinguishes algebraic geometry from other geometric theories (e. g. differential or symplectic geometry). Of course, this comes at a price: our theory must be strong enough to include such singular objects and make statements how things vary when we pass from smooth to singular objects. In this regard, algebraic geometry is related to *singularity theory* which studies precisely these questions.

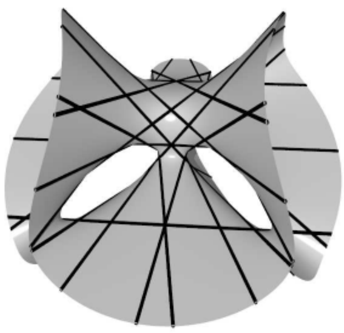
**Remark 0.8.** Maybe it looks a bit restrictive to allow only algebraic (polynomial) equations to describe our geometric objects. But in fact it is a deep theorem that for *compact* objects, we would not get anything different if we allowed *holomorphic* equations too. In this respect, algebraic geometry is very much related (and in certain cases identical) to *complex (analytic) geometry*. The easiest example of this correspondence is that a holomorphic map from the compactified complex plane $\mathbb{C} \cup \{\infty\}$ to itself must in fact be a rational map, i. e. a quotient of two polynomials.

**Example 0.9.** Let us now turn our attention to the next more complicated objects, namely complex surfaces in 3-dimensional space. We just want to give one example here: let $X$ be the *cubic surface*

$$X = \{(x_1,x_2,x_3) : 1 + x_1^3 + x_2^3 + x_3^3 - (1 + x_1 + x_2 + x_3)^3 = 0\} \quad \subset \mathbb{C}^3.$$

As $X$ has real dimension 4, it is impossible to draw pictures of it that reflect its topological properties correctly. Usually, we overcome this problem by just drawing the *real* part, i. e. we look for solutions of the equation over the real numbers. This then gives a real surface in $\mathbb{R}^3$ that we can draw. We should just be careful about which statements we can claim to see from this incomplete geometric picture.

The following picture shows the real part of the surface $X$.



In contrast to our previous examples, we have now used a *linear* projection to map the real 3-dimensional space onto the drawing plane (and not just a topologically correct picture).

We see that there are some lines contained in $X$. In fact, one can show that (after a suitable compactification) *every* smooth cubic surface has exactly 27 lines on it, see Chapter 11. This is another sort of question that one can ask about the solutions of polynomial equations, and that is not of topological nature: do they contain curves with special properties (in this case lines), and if so, how many? This branch of algebraic geometry is usually called *enumerative geometry*.

**Remark 0.10.** It is probably surprising that algebraic geometry, in particular enumerative geometry, is very much related to *theoretical physics*. In fact, many results in enumerative geometry have been found by physicists first.

Why are physicists interested e. g. in the number of lines on the cubic surface? We try to give a short answer to this (that is necessarily vague and incomplete): There is a branch of theoretical physics called *string theory* whose underlying idea is that the elementary particles (electrons, quarks, ...) might not be point-like, but rather 1-dimensional objects (the so-called strings), that are just so small that their 1-dimensional structure cannot be observed directly by any sort of physical measurement. When these particles move in time, they sweep out a surface in space-time. For some reason this surface has a natural complex structure coming from the underlying physical theory.

Now the same idea applies to space-time in general: string theorists believe that space-time is not 4-dimensional as we observe it, but rather has some extra dimensions that are again so small in size that we cannot observe them directly. (Think e. g. of a long tube with a very small diameter — of course this is a 2-dimensional object, but if you look at this tube from very far away you cannot see the small diameter any more, and the object looks like a 1-dimensional line.) These extra dimensions are parametrized by a space that sometimes has a complex structure too; it might for example be the complex cubic surface that we looked at above.

So in this case we are in fact looking at complex curves in a complex surface. A priori, these curves can sit in the surface in any way. But there are equations of motion that tell you how these curves will sit in the ambient space, just as in classical mechanics it follows from the equations of motion that a particle will move on a straight line if no forces apply to it. In our case, the equations of motion say that the curve must map *holomorphically* to the ambient space. As we said in Remark 0.8 above, this is equivalent to saying that we must have algebraic equations that describe the curve. So we are looking at exactly the same type of questions as we did in Example 0.9 above.

**Example 0.11.** Let us now have a brief look at curves in 3-dimensional space. Consider the example

$$C = \{(x_1,x_2,x_3) = (t^3,t^4,t^5) : t \in \mathbb{C}\} \quad \subset \mathbb{C}^3.$$

We have given this curve parametrically, but it is in fact easy to see that we can describe it equally well in terms of polynomial equations:

$$C = \{(x_1,x_2,x_3) : x_1^3 = x_2x_3,\ x_2^2 = x_1x_3,\ x_3^2 = x_1^2x_2\}.$$

What is striking here is that we have *three* equations, although we would expect that a 1-dimensional object in 3-dimensional space should be given by two equations. But in fact, if you leave out any of the above three equations, you are changing the set that it describes: if you leave out e. g. the last equation $x_3^2 = x_1^2x_2$, you would get the whole $x_3$-axis $\{(x_1,x_2,x_3) : x_1 = x_2 = 0\}$ as additional points that do satisfy the first two equations, but not the last one.

So we see another important difference to linear algebra: it is not clear whether a given object of codimension $d$ can be given by $d$ equations — in any case we have just seen that it is in general not possible to choose $d$ defining equations from a given set of such equations. Even worse, for a given set of equations it is in general a difficult task to figure out what dimension their solution has. There do exist algorithms to find this out for any given set of polynomials, but they are so complicated that you will in general want to use a computer program to do that for you. This is a simple example of an application of *computer algebra* to algebraic geometry.

**Exercise 0.12.** Show that if you replace the three equations defining the curve $C$ in Example 0.11 by

$$x_1^3 = x_2x_3,\ x_2^2 = x_1x_3,\ x_3^2 = x_1^2x_2 + \varepsilon$$

for a (small) non-zero number $\varepsilon \in \mathbb{C}$, the resulting set of solutions is in fact 0-dimensional, as you would expect from three equations in 3-dimensional space. So we see that very small changes in the equations can make a very big difference in the resulting solution set. Hence we usually cannot apply numerical methods to our problems: very small rounding errors can change the result completely.

**Remark 0.13.** Especially the previous Example 0.11 is already very algebraic in nature: the question that we asked there does not depend at all on the ground field being the complex numbers. In fact, this is a general philosophy: even if algebraic geometry describes geometric objects (when viewed over the complex numbers), most methods do not rely on this, and therefore should be established in purely algebraic terms. For example, the genus of a curve (that we introduced topologically in Example 0.1) can be defined in purely algebraic terms in such a way that all the statements from complex geometry (e. g. the degree-genus formula of Example 0.3) extend to this more general setting. Many geometric questions then reduce to pure *commutative algebra*, which is in some sense the foundation of algebraic geometry.

**Example 0.14.** The most famous application of algebraic geometry to ground fields other than the complex numbers is certainly Fermat's Last Theorem: this is just the statement that the algebraic curve *over the rational numbers*

$$C = \{(x_1,x_2) \in \mathbb{Q}^2 : x_1^n + x_2^n = 1\} \quad \subset \mathbb{Q}^2$$

contains only the trivial points where $x_1 = 0$ or $x_2 = 0$. Note that this is very different from the case of the ground field $\mathbb{C}$, where we have seen in Example 0.3 that $C$ is a curve of genus $\binom{n-1}{2}$. But a large part of the theory of algebraic geometry applies to the rational numbers (and related fields) as well, so if you look at the proof of Fermat's Theorem you will notice that it uses e. g. the concepts of algebraic curves and their genus at many places, although the corresponding point set $C$ contains only some trivial points. So, in some sense, we can view *(algebraic) number theory* as a part of algebraic geometry.

With this many relations to other fields of mathematics (and physics), it is obvious that we have to restrict our attention in this class to quite a small subset of the possible applications. Although we will develop the general theory of algebraic geometry, our focus will mainly be on geometric questions, neglecting number-theoretic aspects most of the time. So, for example, if we say "let $K$ be an algebraically closed field", feel free to read this as "let $K$ be the complex numbers" and think about geometry rather than algebra.

Every now and then we will quote results from or give applications to other fields of mathematics. This applies in particular to commutative algebra, which provides some of the basic foundations of algebraic geometry. So to fully understand algebraic geometry, you will need to get some background in commutative algebra as well, to the extent as covered e. g. in [AM] or [G5]. However, we will not assume this here — although this is probably not the standard approach it is perfectly possible to follow these notes without any knowledge of commutative algebra. To make this easier, all commutative algebra results that we will need will be stated clearly (and easy to understand), and you can learn the algebraic techniques to prove them afterwards. The only algebraic prerequisite needed for this class is some basic knowledge on groups, rings, fields, and vector spaces as e. g. taught in the "Algebraic Structures" and "Foundations of Mathematics" courses [G1, G2].

# 1. Affine Varieties

As explained in the introduction, the goal of algebraic geometry is to study solutions of polynomial equations in several variables over a fixed ground field. So let us now make the corresponding definitions.

**Convention 1.1.** Throughout these notes, $K$ will always denote a fixed base field (which we will require to be algebraically closed after our discussion of Hilbert's Nullstellensatz in Proposition 1.17). Rings are always assumed to be commutative with a multiplicative unit 1. 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

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

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).

**Definition 1.2** (Affine varieties)**.**

(a) We call

$$\mathbb{A}^n := \mathbb{A}^n_K := \{(c_1,\dots,c_n) : c_i\in K \text{ for } i=1,\dots,n\}$$

the **affine $n$-space** over $K$.

Note that $\mathbb{A}^n_K$ is just $K^n$ as a set. It is customary to use two different notations here since $K^n$ is also a $K$-vector space and a ring. We will usually use the notation $\mathbb{A}^n_K$ 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 polynomial

$$f=\sum_{i_1,\dots,i_n\in\mathbb{N}} a_{i_1,\dots,i_n}\, x_1^{i_1}\cdot\,\cdots\,\cdot x_n^{i_n} \quad \in K[x_1,\dots,x_n]$$

and a point $c=(c_1,\dots,c_n)\in\mathbb{A}^n$ we define the **value** of $f$ at $c$ to be

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

If there is no risk of confusion we will often denote a point in $\mathbb{A}^n$ by the same letter $x$ as we used for the formal variables, writing $f\in K[x_1,\dots,x_n]$ for the polynomial and $f(x)$ for its value at a point $x\in\mathbb{A}^n_K$.

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

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

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 $V(f_1,\dots,f_k)$.

**Remark 1.3.** Some authors refer to zero loci of polynomials in $\mathbb{A}^n$ as in Definition 1.2 (c) as *(affine) algebraic sets*, reserving the name "affine variety" for such zero loci that are in addition *irreducible* (a concept that we will introduce in Definition 2.6 (a)).

**Example 1.4.** Here are some simple examples of affine varieties:

(a) Affine $n$-space itself is an affine variety, since $\mathbb{A}^n=V(0)$. Similarly, the empty set $\emptyset=V(1)$ is an affine variety.

(b) Any single point in $\mathbb{A}^n$ is an affine variety: we have $(c_1,\dots,c_n)=V(x_1-c_1,\dots,x_n-c_n)$.

(c) Linear subspaces of $\mathbb{A}^n = K^n$ are affine varieties.

(d) If $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$ are affine varieties then so is the *product* $X \times Y \subset \mathbb{A}^n \times \mathbb{A}^m = \mathbb{A}^{n+m}$.

(e) All examples from the introduction in Chapter 0 are affine varieties: e. g. the curves of Examples 0.1 and 0.3, and the cubic surface of Example 0.9.

**Remark 1.5** (Affine varieties are zero loci of ideals)**.** Let $f$ and $g$ be polynomials that vanish on a certain subset $X \subset \mathbb{A}^n$. Then $f+g$ and $hf$ for any polynomial $h$ clearly vanish on $X$ as well. This means that the set $S \subset K[x_1,\dots,x_n]$ defining an affine variety $X = V(S)$ is certainly not unique: for any $f,g \in S$ and any polynomial $h$ we can add $f+g$ and $hf$ to $S$ without changing its zero locus. In other words, if

$$I = (S) = \{h_1 f_1 + \cdots + h_m f_m : m \in \mathbb{N},\, f_1,\dots,f_m \in S,\, h_1,\dots,h_m \in K[x_1,\dots,x_n]\}$$

is the ideal generated by $S$, then $V(I) = V(S)$. Hence any affine variety in $\mathbb{A}^n$ can be written as the zero locus of an ideal in $K[x_1,\dots,x_n]$.

**Example 1.6** (Affine varieties in $\mathbb{A}^1$)**.** Let $X$ be an affine variety in $\mathbb{A}^1$. By Remark 1.5 we can then write $X = V(I)$ for an ideal $I \trianglelefteq K[x]$. But $K[x]$ is a principal ideal domain [G1, Example 10.33 (a)]. Hence we have $I = (f)$ for some polynomial $f \in K[x]$, and thus $X = V(f)$.

As zero loci of non-zero polynomials in one variable are always finite, this means that any affine variety in $\mathbb{A}^1$ not equal to $\mathbb{A}^1$ itself must be a finite set. Conversely, any finite subset $\{a_1,\dots,a_n\} = V((x-a_1)\cdots(x-a_n))$ of $\mathbb{A}^1$ is an affine variety, and thus we conclude that the affine varieties in $\mathbb{A}^1$ are exactly the finite sets and $\mathbb{A}^1$ itself.

Unfortunately, for more than one variable we cannot use a similar argument to classify the affine varieties in $\mathbb{A}^n$ as the multivariate polynomial rings $K[x_1,\dots,x_n]$ are not principal ideal domains. However, we still have the following result that we will borrow from commutative algebra.

**Proposition 1.7** (**Hilbert's Basis Theorem** [G5, Proposition 7.13 and Remark 7.15])**.** *Every ideal in the polynomial ring $K[x_1,\dots,x_n]$ can be generated by finitely many elements.*

**Remark 1.8** (Affine varieties are zero loci of finitely many polynomials)**.** Let $X = V(S)$ be an affine variety. Then the ideal generated by $S$ can be written as $(S) = (f_1,\dots,f_m)$ for some $f_1,\dots,f_m \in S$ by Proposition 1.7, and hence $X = V(S) = V(f_1,\dots,f_m)$ by Remark 1.5. So every affine variety is the zero locus of finitely many polynomials.

**Exercise 1.9.** Prove that every affine variety $X \subset \mathbb{A}^n$ consisting of only finitely many points can be written as the zero locus of $n$ polynomials.

(Hint: interpolation.)

There is another reason why Remark 1.5 is important: it is in some sense the basis of algebraic geometry since it relates *geometric objects* (affine varieties) to *algebraic objects* (ideals). In fact, it will be the main goal of this first chapter to make this correspondence precise. We have already assigned affine varieties to ideals in Definition 1.2 (c) and Remark 1.5, so let us now introduce an operation that does the opposite job.

**Definition 1.10** (Ideal of a subset of $\mathbb{A}^n$)**.** Let $X \subset \mathbb{A}^n$ be any subset. Then

$$I(X) := \{f \in K[x_1,\dots,x_n] : f(x) = 0 \text{ for all } x \in X\}$$

is called the **ideal** of $X$ (note that this is indeed an ideal by Remark 1.5).

**Example 1.11** (Ideal of a point)**.** Let $a = (a_1,\dots,a_n) \in \mathbb{A}^n_K$ be a point. Then the ideal of the one-point set $\{a\}$ is $I(a) := I(\{a\}) = (x_1-a_1,\dots,x_n-a_n)$:

"$\subset$" If $f \in I(a)$ then $f(a) = 0$. This means that replacing each $x_i$ by $a_i$ in $f$ gives zero, i. e. that $f$ is zero modulo $(x_1-a_1,\dots,x_n-a_n)$. Hence $f \in (x_1-a_1,\dots,x_n-a_n)$.

"$\supset$" If $f \in (x_1-a_1,\dots,x_n-a_n)$ then $f = \sum_{i=1}^n (x_i-a_i) f_i$ for some $f_1,\dots,f_n \in K[x_1,\dots,x_n]$, and so certainly $f(a) = 0$, i. e. $f \in I(a)$.

We have now constructed operations

$$\begin{array}{rcl} \{\text{affine varieties in } \mathbb{A}^n\} & \longleftrightarrow & \{\text{ideals in } K[x_1,\dots,x_n]\} \\ X & \longmapsto & I(X) \\ V(J) & \longleftarrow & J \end{array}$$

and should check whether they actually give a *bijective* correspondence between ideals and affine varieties. The following lemma tells us the positive results in this direction. 01

**Lemma 1.12.** *Let $S$ and $S'$ be subsets of $K[x_1,\dots,x_n]$, and let $X$ and $X'$ be subsets of $\mathbb{A}^n$.*

(a) *If $X \subset X'$ then $I(X') \subset I(X)$.*
*If $S \subset S'$ then $V(S') \subset V(S)$.*
*We say that the operations $V(\,\cdot\,)$ and $I(\,\cdot\,)$ reverse inclusions.*

(b) *$X \subset V(I(X))$ and $S \subset I(V(S))$.*

(c) *If $X$ is an affine variety then $V(I(X)) = X$.*

*Proof.*

(a) Let $X \subset X'$. If $f \in I(X')$, i.e. $f(x) = 0$ for all $x \in X'$, then certainly also $f(x) = 0$ for all $x \in X$, and hence $f \in I(X)$. The second statement follows analogously.

(b) Let $x \in X$. Then $f(x) = 0$ for every $f \in I(X)$, and thus by definition we have $x \in V(I(X))$. Again, the second inclusion follows in the same way.

(c) By (b) it suffices to prove "$\subset$". As $X$ is an affine variety we can write $X = V(S)$ for some $S \subset K[x_1,\dots,x_n]$. Then $S \subset I(V(S))$ by (b), and thus $V(S) \supset V(I(V(S)))$ by (a). Replacing $V(S)$ by $X$ again now gives the required inclusion. □

By this lemma, the only thing left that would be needed for a bijective correspondence between ideals and affine varieties would be $I(V(J)) \subset J$ for any ideal $J$ (so that then $I(V(J)) = J$ by part (b)). Unfortunately, the following example shows that there are two reasons why this is not true in general.

**Example 1.13** (The inclusion $J \subset I(V(J))$ is strict in general)**.**

(a) Let $J \trianglelefteq \mathbb{C}[x]$ be a non-zero ideal. As $\mathbb{C}[x]$ is a principal ideal domain [G1, Example 10.33 (a)] and $\mathbb{C}$ is algebraically closed, we must have
$$J = \big( (x-a_1)^{k_1} \cdot \,\cdots\, \cdot (x-a_n)^{k_n} \big)$$
for some $n \in \mathbb{N}$, distinct $a_1,\dots,a_n \in \mathbb{C}$, and $k_1,\dots,k_n \in \mathbb{N}_{>0}$. Obviously, the zero locus of this ideal in $\mathbb{A}^1$ is $V(J) = \{a_1,\dots,a_n\}$. The polynomials vanishing on this set are precisely those that contain each factor $x - a_i$ for $i = 1,\dots,n$ at least once, i.e. we have
$$I(V(J)) = \big( (x-a_1) \cdot \,\cdots\, \cdot (x-a_n) \big).$$
If at least one of the numbers $k_1,\dots,k_n$ is greater than 1, this is a bigger ideal than $J$. In other words, *the zero locus of an ideal does not see powers of polynomials*: as a power $f^k$ of a polynomial $f$ has the same zero locus as $f$ itself, the information about this power is lost when applying the operation $I(V(\,\cdot\,))$.

(b) The situation is even worse for ground fields that are not algebraically closed: the ideal $J = (x^2+1) \trianglelefteq \mathbb{R}[x]$ has an empty zero locus in $\mathbb{A}^1$, and so we get $I(V(J)) = I(\emptyset) = \mathbb{R}[x]$. So in this case the complete information on the ideal $J$ is lost when applying the operation $I(V(\,\cdot\,))$.

To overcome the first of these problems, we just have to restrict our attention to ideals with the property that they contain a polynomial $f$ whenever they contain a power $f^k$ of it. The following definition accomplishes this.

**Definition 1.14** (Radicals and radical ideals)**.** Let $I$ be an ideal in a ring $R$.

(a) We call
$$\sqrt{I} := \{f \in R : f^n \in I \text{ for some } n \in \mathbb{N}\}$$
the **radical** of $I$.

(b) The ideal $I$ is said to be a **radical ideal** if $\sqrt{I} = I$.

**Remark 1.15.** Again let $I$ be an ideal in a ring $R$.

(a) The radical $\sqrt{I}$ of $I$ is always an ideal:
- We have $0 \in \sqrt{I}$, since $0 \in I$.
- If $f, g \in \sqrt{I}$, i. e. $f^n \in I$ and $g^m \in I$ for some $n, m \in \mathbb{N}$, then
$$(f+g)^{n+m} = \sum_{k=0}^{n+m} \binom{n+m}{k} f^k g^{n+m-k}$$
is again an element of $I$, since in each summand we must have that the power of $f$ is at least $n$ (in which case $f^k \in I$) or the power of $g$ is at least $m$ (in which case $g^{n+m-k} \in I$). Hence $f + g \in \sqrt{I}$.
- If $h \in R$ and $f \in \sqrt{I}$, i. e. $f^n \in I$ for some $n \in \mathbb{N}$, then $(hf)^n = h^n f^n \in I$, and hence $hf \in \sqrt{I}$.

Moreover, it is obvious that $I \subset \sqrt{I}$ (we can always take $n = 1$ in Definition 1.14 (a)). Hence $I$ is radical if and only if $\sqrt{I} \subset I$, i. e. if $f^n \in I$ for some $n \in \mathbb{N}$ implies $f \in I$.

(b) As expected from the terminology, the radical of $I$ is a radical ideal: if $f^n \in \sqrt{I}$ for some $f \in R$ and $n \in \mathbb{N}$ then $(f^n)^m = f^{nm} \in I$ for some $m \in \mathbb{N}$, and hence $f \in \sqrt{I}$.

(c) If $I$ is the ideal of an affine variety $X$ then $I$ is radical: if $f \in \sqrt{I}$ then $f^k$ vanishes on $X$, hence $f$ vanishes on $X$ and we also have $f \in I$.

**Example 1.16.** Continuing Example 1.13 (a), the radical of the ideal
$$J = \big((x-a_1)^{k_1} \cdot \cdots \cdot (x-a_n)^{k_n}\big) \quad \trianglelefteq \mathbb{C}[x]$$
consists of all polynomials $f \in \mathbb{C}[x]$ such that $(x-a_1)^{k_1} \cdot \cdots \cdot (x-a_n)^{k_n}$ divides $f^k$ for large enough $k$. This is obviously the set of all polynomials containing each factor $x - a_i$ for $i = 1, \dots, n$ at least once, i. e. we have
$$\sqrt{J} = \big((x-a_1) \cdot \cdots \cdot (x-a_n)\big).$$
One should note however that the explicit computation of radicals is in general hard and requires algorithms of computer algebra.

In our example at hand we therefore see that $I(V(J)) = \sqrt{J}$, resp. that $I(V(J)) = J$ if $J$ is radical. In fact, this holds in general for ideals in polynomial rings over algebraically closed fields. This statement is usually referred to as *Hilbert's Nullstellensatz* ("theorem of the zeroes"); it is another fact that we will quote here from commutative algebra.

**Proposition 1.17 (Hilbert's Nullstellensatz [G5, Corollary 10.14]).** *Let $K$ be an algebraically closed field. Then for every ideal $J \trianglelefteq K[x_1, \dots, x_n]$ we have $I(V(J)) = \sqrt{J}$. In particular, there is an inclusion-reversing one-to-one correspondence*
$$\begin{array}{rcl} \{\text{affine varieties in } \mathbb{A}^n\} & \longleftrightarrow & \{\text{radical ideals in } K[x_1, \dots, x_n]\} \\ X & \longmapsto & I(X) \\ V(J) & \longleftarrow & J. \end{array}$$

*Proof.* The main statement $I(V(J)) = \sqrt{J}$ is proven in [G5, Corollary 10.14]. The correspondence then follows from what we have already seen:
- $I(\,\cdot\,)$ maps affine varieties to radical ideals by Remark 1.15 (c);
- we have $V(I(X)) = X$ for any affine variety $X$ by Lemma 1.12 (c) and $I(V(J)) = J$ for any radical ideal $J$ by our main statement;

- the correspondence reverses inclusions by Lemma 1.12 (a). □

As we have already mentioned, this result is absolutely central for algebraic geometry since it allows us to translate geometric objects into algebraic ones. Note however that the introduction of radical ideals allowed us to solve Problem (a) in Example 1.13, but not Problem (b): for ground fields that are not algebraically closed the statement of Proposition 1.17 is clearly false since e. g. the ideal $J = (x^2+1) \trianglelefteq \mathbb{R}[x]$ is radical but has an empty zero locus, so that $I(V(J)) = \mathbb{R}[x] \neq (x^2+1) = \sqrt{J}$. Let us therefore agree:

$$\boxed{\text{From now on, our ground field } K \text{ will always be assumed to be algebraically closed.}}$$

**Remark 1.18.**

(a) Let $J \trianglelefteq K[x_1,\dots,x_n]$ be an ideal in the polynomial ring (over an algebraically closed field). If $J \neq K[x_1,\dots,x_n]$ then $J$ has a zero, i. e. $V(J)$ is non-empty: otherwise we would have $\sqrt{J} = I(V(J)) = I(\emptyset) = K[x_1,\dots,x_n]$ by Proposition 1.17, which means $1 \in \sqrt{J}$ and gives us the contradiction $1 \in J$. This statement can be thought of as a generalization of the algebraic closure property that a non-constant univariate polynomial has a zero. It is the origin of the name "Nullstellensatz" for Proposition 1.17.

(b) Another easy consequence of Proposition 1.17 is that polynomials and polynomial functions on $\mathbb{A}^n$ agree: if $f,g \in K[x_1,\dots,x_n]$ are two polynomials defining the same function on $\mathbb{A}^n$, i. e. such that $f(x) = g(x)$ for all $x \in \mathbb{A}^n$, then

$$f-g \in I(\mathbb{A}^n) = I(V(0)) = \sqrt{(0)} = (0)$$

and hence $f = g$ in $K[x_1,\dots,x_n]$. So $K[x_1,\dots,x_n]$ can be thought of as the ring of polynomial functions on $\mathbb{A}^n$. Note that this is false for general fields, since e. g. the polynomial $x^2+x \in \mathbb{Z}_2[x]$ defines the zero function on $\mathbb{A}^1_{\mathbb{Z}_2}$, although it is not the zero polynomial.

More generally, if $X$ is an affine variety then two polynomials $f,g \in K[x_1,\dots,x_n]$ define the same polynomial function on $X$, i. e. $f(x) = g(x)$ for all $x \in X$, if and only if $f-g \in I(X)$. So the quotient ring $K[x_1,\dots,x_n]/I(X)$ can be thought of as the ring of polynomial functions on $X$. Let us make this into a precise definition.

**Definition 1.19** (Polynomial functions and coordinate rings)**.** Let $X \subset \mathbb{A}^n$ be an affine variety. A **polynomial function** on $X$ is a map $X \to K$ that is of the form $x \mapsto f(x)$ for some $f \in K[x_1,\dots,x_n]$. By Remark 1.18 (b) the ring of all polynomial functions on $X$ is just the quotient ring

$$A(X) := K[x_1,\dots,x_n]/I(X).$$

It is usually called the **coordinate ring** of the affine variety $X$.

According to this definition, we can think of the elements of $A(X)$ in the following both as functions on $X$ and as elements of the quotient ring $K[x_1,\dots,x_n]/I(X)$. We can use this ring to define the concepts introduced so far in a relative setting, i. e. consider zero loci of ideals in $A(Y)$ and varieties contained in $Y$ for a fixed ambient affine variety $Y$ that is not necessarily $\mathbb{A}^n$:

**Construction 1.20** (Relative version of the correspondence between varieties and radical ideals)**.** Let $Y \subset \mathbb{A}^n$ be an affine variety. The following two constructions are then completely analogous to those in Definitions 1.2 (c) and 1.10:

(a) For a subset $S \subset A(Y)$ of polynomial functions on $Y$ we define its **zero locus** as

$$V(S) := V_Y(S) := \{x \in Y : f(x) = 0 \text{ for all } f \in S\} \quad \subset Y.$$

The subsets that are of this form are obviously precisely the affine varieties contained in $X$. They are called **affine subvarieties** of $Y$.

(b) For a subset $X \subset Y$ the **ideal** of $X$ in $Y$ is defined to be

$$I(X) := I_Y(X) := \{f \in A(Y) : f(x) = 0 \text{ for all } x \in X\} \quad \trianglelefteq A(Y).$$

With the same arguments as above, all results considered so far then hold in this relative setting as well. Let us summarize them here again:

**Proposition 1.21.** *Let $Y$ be an affine variety in $\mathbb{A}^n$.*

(a) *(**Hilbert's Basis Theorem**) Every ideal in $A(Y)$ can be generated by finitely many elements.*

(b) *(**Hilbert's Nullstellensatz**) For any ideal $J \trianglelefteq A(Y)$ we have $I_Y(V_Y(J)) = \sqrt{J}$. In particular, there is an inclusion-reversing one-to-one correspondence*

$$\begin{array}{rcl} \{\text{affine subvarieties of } Y\} & \longleftrightarrow & \{\text{radical ideals in } A(Y)\} \\ X & \longmapsto & I_Y(X) \\ V_Y(J) & \longleftarrow\!\shortmid & J. \end{array}$$

(c) *For a subvariety $X$ of $Y$ we have $A(X) \cong A(Y)/I_Y(X)$.*

*Proof.* As in our earlier version, the proof of (a) is covered by [G5, Proposition 7.13 and Remark 7.15], the proof of (b) by [G5, Corollary 10.14] and Proposition 1.17. The statement (c) follows in the same way as in Remark 1.18 (b). □

**Exercise 1.22.** Determine the radical of the ideal $(x_1^3 - x_2^6, x_1x_2 - x_2^3) \trianglelefteq \mathbb{C}[x_1,x_2]$.

**Exercise 1.23.** Let $X$ be an affine variety. Show that the coordinate ring $A(X)$ is a field if and only if $X$ is a single point.

In the rest of this chapter we want to study the basic properties of the operations $V(\cdot)$ and $I(\cdot)$.

**Lemma 1.24** (Properties of $V(\cdot)$)**.** *Let $X$ be an affine variety.*

(a) *If $J$ is any index set and $\{S_i : i \in J\}$ a family of subsets of $A(X)$ then $\bigcap_{i\in J} V(S_i) = V(\bigcup_{i\in J} S_i)$ in $X$.*

(b) *For $S_1, S_2 \subset A(X)$ we have $V(S_1) \cup V(S_2) = V(S_1S_2)$ in $X$, where as usual we set $S_1S_2 := \{fg : f \in S_1, g \in S_2\}$.*

*In particular, arbitrary intersections and finite unions of affine subvarieties of $X$ are again affine subvarieties of $X$.*

*Proof.*

(a) We have $x \in \bigcap_{i\in J} V(S_i)$ if and only if $f(x) = 0$ for all $f \in S_i$ for all $i \in J$, which is the case if and only if $x \in V(\bigcup_{i\in J} S_i)$.

(b) "$\subset$" If $x \in V(S_1) \cup V(S_2)$ then $f(x) = 0$ for all $f \in S_1$ or $g(x) = 0$ for all $g \in S_2$. In any case this means that $(fg)(x) = 0$ for all $f \in S_1$ and $g \in S_2$, i. e. that $x \in V(S_1S_2)$.

"$\supset$" If $x \notin V(S_1) \cup V(S_2)$, i. e. $x \notin V(S_1)$ and $x \notin V(S_2)$, then there are $f \in S_1$ and $g \in S_2$ with $f(x) \neq 0$ and $g(x) \neq 0$. Then $(fg)(x) \neq 0$, and hence $x \notin V(S_1S_2)$. □

**Remark 1.25** (Ideal-theoretic version of the properties of $V(\cdot)$)**.** If we want to consider zero loci of ideals rather than of general subsets of $A(X)$, then the properties of Lemma 1.24 take a slightly different form. To see this, let $J_1$ and $J_2$ be any ideals in $A(X)$.

(a) The ideal generated by $J_1 \cup J_2$ is just the *sum of ideals* $J_1 + J_2 = \{f + g : f \in J_1, g \in J_2\}$. So with Remark 1.5 the result of Lemma 1.24 (a) translates into

$$V(J_1) \cap V(J_2) = V(J_1 + J_2).$$

(b) In the same way as in (a), Lemma 1.24 (b) implies that $V(J_1) \cup V(J_2)$ is equal to the zero locus of the ideal generated by $J_1J_2$. Unfortunately, the usual convention is that for two *ideals* $J_1$ and $J_2$ (instead of arbitrary sets) the notation $J_1J_2$ denotes the ideal generated by all products $fg$ with $f \in J_1$ and $g \in J_2$, which is called the *product of the ideals* $J_1$ and $J_2$ — rather than the set of all such products $fg$ itself. So we get

$$V(J_1) \cup V(J_2) = V(J_1J_2)$$

with this modified definition of the product $J_1 J_2$.

(c) Another common operation on ideals is the *intersection* $J_1 \cap J_2$. In general, this ideal is different from the ones considered above, but we can show that there is always the relation
$$\sqrt{J_1 \cap J_2} = \sqrt{J_1 J_2}:$$
"$\subset$" If $f \in \sqrt{J_1 \cap J_2}$ then $f^n \in J_1 \cap J_2$ for some $n$. This means that $f^{2n} = f^n \cdot f^n \in J_1 J_2$, and hence that $f \in \sqrt{J_1 J_2}$.

"$\supset$" For $f \in \sqrt{J_1 J_2}$ we have $f^n \in J_1 J_2$ for some $n$. Then $f^n \in J_1 \cap J_2$, and thus $f \in \sqrt{J_1 \cap J_2}$.

By Proposition 1.21 (b) this means that $I(V(J_1 \cap J_2)) = I(V(J_1 J_2))$, and hence by applying $V(\cdot)$ that
$$V(J_1 \cap J_2) = V(J_1 J_2) = V(J_1) \cup V(J_2)$$
by (b).

Finally, for completeness let us also formulate the properties of Lemma 1.24 and Remark 1.25 in terms of the operation $I(\cdot)$ rather than $V(\cdot)$.

**Lemma 1.26** (Properties of $I(\cdot)$)**.** *Let $X$ be an affine variety, and let $Y_1$ and $Y_2$ be affine subvarieties of $X$. Then:*

(a) $I(Y_1 \cap Y_2) = \sqrt{I(Y_1) + I(Y_2)}$;

(b) $I(Y_1 \cup Y_2) = I(Y_1) \cap I(Y_2)$.

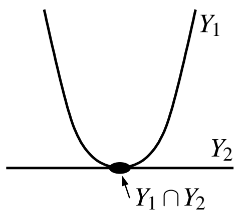
*Proof.*

(a) We have
$$\begin{aligned} I(Y_1 \cap Y_2) &= I(V(I(Y_1)) \cap V(I(Y_2))) && \text{(Proposition 1.21 (b))} \\ &= I(V(I(Y_1) + I(Y_2))) && \text{(Remark 1.25 (a))} \\ &= \sqrt{I(Y_1) + I(Y_2)}. && \text{(Proposition 1.21 (b))} \end{aligned}$$

(b) A polynomial function $f \in A(X)$ is contained in $I(Y_1 \cup Y_2)$ if and only if $f(x) = 0$ for all $x \in Y_1$ and all $x \in Y_2$, which is the case if and only if $f \in I(Y_1) \cap I(Y_2)$. □ 02

**Remark 1.27.** Recall from Remark 1.15 (c) that ideals of affine varieties are always radical. So in particular, Lemma 1.26 (b) shows that intersections of radical ideals in $A(X)$ are again radical — which could of course also be checked directly. In contrast, sums of radical ideals are in general not radical, and hence taking the radical in Lemma 1.26 (a) is really necessary.

In fact, there is also a geometric interpretation behind this fact. Consider for example the affine varieties $Y_1, Y_2 \subset \mathbb{A}^1_{\mathbb{C}}$ with ideals $I(Y_1) = (x_2 - x_1^2)$ and $I(Y_2) = (x_2)$ whose real points are shown in the picture on the right. Their intersection $Y_1 \cap Y_2$ is obviously the origin with ideal $I(Y_1 \cap Y_2) = (x_1, x_2)$. But
$$I(Y_1) + I(Y_2) = (x_2 - x_1^2, x_2) = (x_1^2, x_2)$$
is not a radical ideal; only its radical is equal to $I(Y_1 \cap Y_2) = (x_1, x_2)$.




The geometric meaning of the non-radical ideal $I(Y_1) + I(Y_2) = (x_1^2, x_2)$ is that $Y_1$ and $Y_2$ are tangent at the intersection point: if we consider the function $x_2 - x_1^2$ defining $Y_1$ on the $x_1$-axis $Y_2$ (where it is equal to $-x_1^2$) we see that it vanishes to order 2 at the origin. This means that $Y_1$ and $Y_2$ share the $x_1$-axis as common tangent direction, so that the intersection $Y_1 \cap Y_2$ can be thought of as "extending to an infinitesimally small amount in the $x_1$-direction", and we can consider $Y_1$ and $Y_2$ as "intersecting with multiplicity 2" at the origin. We will see later in Definition 12.23 (b) how such intersection multiplicities can be defined rigorously.

# 2. The Zariski Topology

In this chapter we will define a *topology* on an affine variety $X$, i. e. a notion of open and closed subsets of $X$. We will see that many properties of $X$ can be expressed purely in terms of this topology, e. g. its dimension or the question whether it consists of several components. The advantage of this is that these concepts can then easily be reused later in Chapter 5 when we consider more general varieties — which are still topological spaces, but arise in a slightly different way.

Compared to e. g. the standard topology on subsets of real vector spaces, the properties of our topology on affine varieties will be very unusual. Consequently, most concepts and results covered in a standard introductory course on topology will be trivial or useless in our case, so that we will only need the very first definitions of general topology. Let us quickly review them here.

**Remark 2.1** (Topologies)**.** A *topology* on a set $X$ is given by declaring some subsets of $X$ to be *closed*, such that the following properties hold:

(a) the empty set $\emptyset$ and the whole space $X$ are closed;

(b) arbitrary intersections of closed sets are closed;

(c) finite unions of closed sets are closed.

Given such a topology on $X$, a subset $U$ of $X$ is then called *open* if its complement $X\backslash U$ is closed. The *closure* $\overline{A}$ of a subset $A \subset X$ is defined to be the smallest closed subset containing $A$, or more precisely the intersection of all closed subsets containing $A$ (which is closed again by (b)).

A topology on $X$ induces a *subspace topology* on any subset $A \subset X$ by declaring the subsets of $A$ to be closed that are of the form $A \cap Y$ for a closed subset $Y$ of $X$ (or equivalently the subsets of $A$ to be open that are of the form $A \cap U$ for an open subset $U$ of $X$). Subsets of topological spaces will always be equipped with this subspace topology unless stated otherwise. Note that if $A$ is closed itself then the closed subsets of $A$ in the subspace topology are exactly the closed subsets of $X$ contained in $A$; if $A$ is open then the open subsets of $A$ in the subspace topology are exactly the open subsets of $X$ contained in $A$.

A map $\varphi : X \to Y$ between topological spaces is called *continuous* if inverse images of closed subsets of $Y$ under $\varphi$ are closed in $X$, or equivalently if inverse images of open subsets are open.

Note that the standard definition of closed subsets in $\mathbb{R}^n$ (or more generally in metric spaces) that you know from real analysis satisfies the conditions (a), (b), and (c), and leads with the above definitions to the well-known notions of open subsets, closures, and continuous functions.

With these preparations we can now define the standard topology used in algebraic geometry.

**Definition 2.2** (Zariski topology)**.** Let $X$ be an affine variety. We define the **Zariski topology** on $X$ to be the topology whose closed sets are exactly the affine subvarieties of $X$, i. e. the subsets of the form $V(S)$ for some $S \subset A(X)$. Note that this in fact a topology by Example 1.4 (a) and Lemma 1.24.

Unless stated otherwise, topological notions for affine varieties (and their subsets, using the subspace topology of Remark 2.1) will always be understood with respect to this topology.

**Remark 2.3.** Let $X \subset \mathbb{A}^n$ be an affine variety. Then we have just defined two topologies on $X$:

(a) the Zariski topology on $X$, whose closed subsets are the affine subvarieties of $X$; and

(b) the subspace topology of $X$ in $\mathbb{A}^n$, whose closed subsets are the sets of the form $X \cap Y$, with $Y$ a variety in $\mathbb{A}^n$.

These two topologies agree, since the subvarieties of $X$ are precisely the affine varieties contained in $X$ and the intersection of two affine varieties is again an affine variety. Hence it will not lead to confusion if we consider both these topologies to be the standard topology on $X$.

**Exercise 2.4.** Let $X \subset \mathbb{A}^n$ be an arbitrary subset. Prove that $V(I(X)) = \overline{X}$.

**Example 2.5** (Topologies on complex varieties)**.** Compared to the classical metric topology in the case of the ground field $\mathbb{C}$, the Zariski topology is certainly unusual:

(a) The metric unit ball $A = \{x \in \mathbb{A}^1_{\mathbb{C}} : |x| \le 1\}$ in $\mathbb{A}^1_{\mathbb{C}}$ is clearly closed in the classical topology, but not in the Zariski topology. In fact, by Example 1.6 the Zariski-closed subsets of $\mathbb{A}^1$ are only the finite sets and $\mathbb{A}^1$ itself. In particular, the closure of $A$ in the Zariski topology is all of $\mathbb{A}^1$.

Intuitively, we can say that the closed subsets in the Zariski topology are very "small", and hence that the open subsets are very "big" (see also Remark 2.18). Any Zariski-closed subset is also closed in the classical topology (since it is given by equations among polynomial functions, which are continuous in the classical topology), but as the above example shows only "very few" closed subsets in the classical topology are also Zariski-closed.

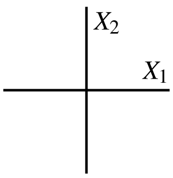
(b) Let $\varphi : \mathbb{A}^1 \to \mathbb{A}^1$ be *any* bijective map. Then $\varphi$ is continuous in the Zariski topology, since inverse images of finite subsets of $\mathbb{A}^1$ under $\varphi$ are finite.

This statement is essentially useless however, as we will not define morphisms of affine varieties as just being continuous maps with respect to the Zariski topology. In fact, this example gives us a strong hint that we should not do so.

(c) In general topology there is a notion of a *product topology*: if $X$ and $Y$ are topological spaces then $X \times Y$ has a natural structure of a topological space by saying that a subset of $X \times Y$ is open if and only if it is a union of products $U_i \times V_i$ for open subsets $U_i \subset X$ and $V_i \subset Y$ with $i$ in an arbitrary index set.

With this construction, note however that the Zariski topology of an affine product variety $X \times Y$ is not the product topology: e. g. the subset $V(x_1 - x_2) = \{(a,a) : a \in K\} \subset \mathbb{A}^2$ is closed in the Zariski topology, but not in the product topology of $\mathbb{A}^1 \times \mathbb{A}^1$. In fact, we will see in Proposition 4.10 that the Zariski topology is the "correct" one, whereas the product topology is useless in algebraic geometry.

But let us now start with the discussion of the topological concepts that are actually useful in the Zariski topology. The first ones concern *components* of an affine variety: the affine variety $X = V(x_1x_2) \subset \mathbb{A}^2$ as in the picture on the right can be written as the union of the two coordinate axes $X_1 = V(x_2)$ and $X_2 = V(x_1)$, which are themselves affine varieties. However, $X_1$ and $X_2$ cannot be decomposed further into finite unions of smaller affine varieties. The following notion generalizes this idea.



$X = X_1 \cup X_2$

**Definition 2.6** (Irreducible and connected spaces)**.** Let $X$ be a topological space.

(a) We say that $X$ is **reducible** if it can be written as $X = X_1 \cup X_2$ for closed subsets $X_1, X_2 \subsetneq X$. Otherwise $X$ is called **irreducible**.

(b) The space $X$ is called **disconnected** if it can be written as $X = X_1 \cup X_2$ for closed subsets $X_1, X_2 \subsetneq X$ with $X_1 \cap X_2 = \emptyset$. Otherwise $X$ is called **connected**.

**Remark 2.7.** Although we have given this definition for arbitrary topological spaces, you will usually want to apply the notion of irreducibility only in the Zariski topology. For example, in the classical topology, the complex plane $\mathbb{A}^1_{\mathbb{C}}$ is reducible because it can be written e. g. as the union of closed subsets as

$$\mathbb{A}^1_{\mathbb{C}} = \{z \in \mathbb{C} : |z| \le 1\} \cup \{z \in \mathbb{C} : |z| \ge 1\}.$$

In the Zariski topology however, $\mathbb{A}^1$ is irreducible by Example 1.6 (as it should be).

In contrast, the notion of connectedness can be used in the "usual" topology too and does mean there what you think it should mean.

In the Zariski topology, the algebraic characterization of the irreducibility and connectedness of affine varieties is the following.

**Proposition 2.8.** *Let $X$ be a disconnected affine variety, with $X = X_1 \cup X_2$ for two disjoint closed subsets $X_1, X_2 \subsetneq X$. Then $A(X) \cong A(X_1) \times A(X_2)$.*

*Proof.* By Proposition 1.21 (c) we have $A(X_1) \cong A(X)/I(X_1)$ and $A(X_2) \cong A(X)/I(X_2)$. Hence there is a ring homomorphism

$$\varphi : A(X) \to A(X_1) \times A(X_2), \;\; f \mapsto (\overline{f}, \overline{f}).$$

We have to show that it is bijective.

- $\varphi$ is injective: If $(\overline{f}, \overline{f}) = (\overline{0}, \overline{0})$ then $f \in I(X_1) \cap I(X_2) = I(X_1 \cup X_2) = I(X) = (0)$ by Lemma 1.26 (b).
- $\varphi$ is surjective: Let $(\overline{f_1}, \overline{f_2}) \in A(X_1) \times A(X_2)$. Note that
$$A(X) = I(\emptyset) = I(X_1 \cap X_2) = \sqrt{I(X_1) + I(X_2)}$$
by Lemma 1.26 (a). Thus $1 \in \sqrt{I(X_1) + I(X_2)}$, and hence $1 \in I(X_1) + I(X_2)$, which means $I(X_1) + I(X_2) = A(X)$. We can therefore find $g_1 \in I(X_1)$ and $g_2 \in I(X_2)$ with $f_1 - f_2 = g_1 - g_2$, so that $f_1 - g_1 = f_2 - g_2$. This latter element of $A(X)$ then maps to $(\overline{f_1}, \overline{f_2})$ under $\varphi$. □

**Proposition 2.9.** *An affine variety $X$ is irreducible if and only if $A(X)$ is an integral domain.*

*Proof.* "$\Rightarrow$": Assume that $A(X)$ is not an integral domain, i. e. there are non-zero $f_1, f_2 \in A(X)$ with $f_1 f_2 = 0$. Then $X_1 = V(f_1)$ and $X_2 = V(f_2)$ are closed, not equal to $X$ (since $f_1$ and $f_2$ are non-zero), and $X_1 \cup X_2 = V(f_1) \cup V(f_2) = V(f_1 f_2) = V(0) = X$. Hence $X$ is reducible.

"$\Leftarrow$": Assume that $X$ is reducible, with $X = X_1 \cup X_2$ for closed subsets $X_1, X_2 \subsetneq X$. By Proposition 1.21 (b) this means that $I(X_i) \neq (0)$ for $i = 1, 2$, and so we can choose non-zero $f_i \in I(X_i)$. Then $f_1 f_2$ vanishes on $X_1 \cup X_2 = X$. Hence $f_1 f_2 = 0 \in A(X)$, i. e. $A(X)$ is not an integral domain. □

**Remark 2.10.** If $X$ is an affine subvariety of an affine variety $Y$ we know by Proposition 1.21 (c) that $A(X) = A(Y)/I(X)$. So $A(X)$ is an integral domain, i. e. $X$ is irreducible, if and only if for all $f, g \in A(Y)$ the relation $fg \in I(X)$ implies $f \in I(X)$ or $g \in I(X)$. In commutative algebra, ideals with this property are called *prime ideals*. So in other words, in the one-to-one correspondence of Proposition 1.21 (b) between affine subvarieties of $Y$ and radical ideals in $A(Y)$ the irreducible subvarieties correspond exactly to prime ideals.

**Example 2.11.**

(a) A finite affine variety is irreducible if and only if it is connected: namely if and only if it contains at most one point.

(b) Any irreducible space is connected.

(c) The affine space $\mathbb{A}^n$ is irreducible (and thus connected) by Proposition 2.9 since its coordinate ring $A(\mathbb{A}^n) = K[x_1, \dots, x_n]$ is an integral domain. More generally, this holds for any affine variety given by linear equations, since again its coordinate ring is isomorphic to a polynomial ring.

(d) The union $X = V(x_1 x_2) \subset \mathbb{A}^2$ of the two coordinate axes $X_1 = V(x_2)$ and $X_2 = V(x_1)$ is not irreducible, since $X = X_1 \cup X_2$. But $X_1$ and $X_2$ themselves are irreducible by (c). Hence we have decomposed $X$ into a union of two irreducible spaces.

As already announced, we now want to see that such a decomposition into finitely many irreducible spaces is possible for any affine variety. In fact, this works for a much larger class of topological spaces, the so-called *Noetherian* spaces:

**Definition 2.12** (Noetherian topological spaces)**.** A topological space $X$ is called **Noetherian** if there is no infinite strictly decreasing chain

$$X_0 \supsetneq X_1 \supsetneq X_2 \supsetneq \cdots$$

of closed subsets of $X$.

**Lemma 2.13.** *Any affine variety is a Noetherian topological space.*

*Proof.* Let $X$ be an affine variety. Assume that there is an infinite chain $X_0 \supsetneq X_1 \supsetneq X_2 \supsetneq \cdots$ of subvarieties of $X$. By Proposition 1.21 (b) there is then a corresponding infinite chain

$$I(X_0) \subsetneq I(X_1) \subsetneq I(X_2) \subsetneq \cdots$$

of ideals in $A(X)$. It is checked immediately that the union $I := \bigcup_{n=0}^{\infty} I(X_i)$ is then an ideal as well [G1, Exercise 10.38 (a)]. By Proposition 1.21 (a) it is finitely generated, i. e. we have $I = (f_1, \dots, f_n)$ for some $f_1, \dots, f_n \in I$. All these polynomials have to lie in one of the ideals $I(X_m)$ — and in fact in the same one since these ideals form a chain. But then we have $I = (f_1, \dots, f_n) \subset I(X_m) \subsetneq I$, a contradiction. □

**Remark 2.14** (Subspaces of Noetherian spaces are Noetherian)**.** Let $A$ be a subset of a Noetherian topological space $X$. Then $A$ is also Noetherian: otherwise we would have an infinite strictly descending chain of closed subsets of $A$, which by definition of the subspace topology we can write as

$$A \cap Y_0 \supsetneq A \cap Y_1 \supsetneq A \cap Y_2 \supsetneq \cdots$$

for closed subsets $Y_0, Y_1, Y_2, \dots$ of $X$. Then

$$Y_0 \supset Y_0 \cap Y_1 \supset Y_0 \cap Y_1 \cap Y_2 \supset \cdots$$

is an infinite decreasing chain of closed subsets of $X$. In fact, in contradiction to our assumption it is also strictly decreasing, since $Y_0 \cap \cdots \cap Y_k = Y_0 \cap \cdots \cap Y_{k+1}$ for some $k \in \mathbb{N}$ would imply $A \cap Y_k = A \cap Y_{k+1}$ by intersecting with $A$.

Combining Lemma 2.13 with Remark 2.14 we therefore see that any subset of an affine variety is a Noetherian topological space. In fact, all topological spaces occurring in this class will be Noetherian, and thus we can safely restrict our attention to this class of spaces.

**Proposition 2.15** (Irreducible decomposition of Noetherian spaces)**.** *Every Noetherian topological space $X$ can be written as a finite union $X = X_1 \cup \cdots \cup X_r$ of irreducible closed subsets. If one assumes that $X_i \not\subset X_j$ for all $i \neq j$, then $X_1, \dots, X_r$ are unique (up to permutation). They are called the* ***irreducible components*** *of $X$.*

*Proof.* To prove existence, assume that there is a topological space $X$ for which the statement is false. In particular, $X$ is reducible, hence $X = X_1 \cup X_1'$ as in Definition 2.6 (a). Moreover, the statement of the proposition must be false for at least one of these two subsets, say $X_1$. Continuing this construction, one arrives at an infinite chain $X \supsetneq X_1 \supsetneq X_2 \supsetneq \cdots$ of closed subsets, which is a contradiction as $X$ is Noetherian.

To show uniqueness, assume that we have two decompositions

$$X = X_1 \cup \cdots \cup X_r = X_1' \cup \cdots \cup X_s'. \qquad (*)$$

Then for any fixed $i \in \{1, \dots, r\}$ we have $X_i \subset \bigcup_j X_j'$, so $X_i = \bigcup_j (X_i \cap X_j')$. But $X_i$ is irreducible, and so we must have $X_i = X_i \cap X_j'$, i. e. $X_i \subset X_j'$ for some $j$. In the same way we conclude that $X_j' \subset X_k$ for some $k$, so that $X_i \subset X_j' \subset X_k$. By assumption this is only possible for $i = k$, and consequently $X_i = X_j'$. Hence every set appearing on the left side of $(*)$ also appears on the right side (and vice versa), which means that the two decompositions agree. □

**Remark 2.16** (Computation of irreducible decompositions)**.** In general, the actual computation of the irreducible decomposition of an affine variety is quite difficult and requires advanced algorithmic methods of computer algebra. In fact, the corresponding question in commutative algebra is to find the isolated primes of a so-called *primary decomposition* of an ideal [G5, Chapter 8]. But in simple cases the irreducible decomposition might be easy to spot geometrically, as e. g. in Example 2.11 (d).

**Exercise 2.17.** Find the irreducible components of the affine variety $V(x_1 - x_2x_3, x_1x_3 - x_2^2) \subset \mathbb{A}^3_{\mathbb{C}}$. 03

**Remark 2.18** (Open subsets of irreducible spaces are dense)**.** We have already seen in Example 2.5 (a) that open subsets tend to be very "big" in the Zariski topology. Here are two precise statements along these lines. Let $X$ be an irreducible topological space, and let $U$ and $U'$ be non-empty open subsets of $X$. Then:

(a) The intersection $U \cap U'$ is never empty. In fact, by taking complements this is just equivalent to saying that the union of the two proper closed subsets $X \backslash U$ and $X \backslash U'$ is not equal to $X$, i. e. to the definition of irreducibility.

(b) The closure $\overline{U}$ of $U$ is all of $X$ — one says that $U$ is *dense* in $X$. This is easily seen: if $Y \subset X$ is any closed subset containing $U$ then $X = Y \cup (X \backslash U)$, and since $X$ is irreducible and $X \backslash U \neq X$ we must have $Y = X$.

**Exercise 2.19.** Let $A$ be a subset of a topological space $X$. Prove:

(a) If $Y \subset A$ is closed in the subspace topology of $A$ then $\overline{Y} \cap A = Y$.

(b) $A$ is irreducible if and only if $\overline{A}$ is irreducible.

**Exercise 2.20.** Let $\{U_i : i \in I\}$ be an open cover of a topological space $X$, and assume that $U_i \cap U_j \neq \emptyset$ for all $i, j \in I$. Show:

(a) If $U_i$ is connected for all $i \in I$ then $X$ is connected.

(b) If $U_i$ is irreducible for all $i \in I$ then $X$ is irreducible.

**Exercise 2.21.** Let $f : X \to Y$ be a continuous map of topological spaces. Prove:

(a) If $X$ is irreducible then so is $f(X)$.

(b) If $X$ is connected then so is $f(X)$.

**Exercise 2.22.** Let $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$ be irreducible affine varieties. Prove that the coordinate ring $A(X \times Y)$ of their product is an integral domain, and hence that $X \times Y$ is irreducible as well.

An important application of the notion of irreducibility is the definition of the dimension of an affine variety (or more generally of a topological space — but as with our other concepts above you will only want to apply it to the Zariski topology). Of course, at least in the case of complex varieties we have a geometric idea what the dimension of an affine variety should be: the number of coordinates that you need to describe $X$ locally around any point. Although there are algebraic definitions of dimension that mimic this intuitive one [G5, Proposition 11.31], the standard definition of dimension that we will give here uses only the language of topological spaces. Finally, all these definitions are of course equivalent and describe the intuitive notion of dimension, but it is actually quite hard to prove this rigorously.

The idea to construct the dimension in algebraic geometry using the Zariski topology is rather simple: if $X$ is an *irreducible* topological space, then any closed subset of $X$ not equal to $X$ should have smaller dimension. The resulting definition is the following.

**Definition 2.23** (Dimension and codimension)**.** Let $X$ be a non-empty topological space.

(a) The **dimension** $\dim X \in \mathbb{N} \cup \{\infty\}$ is the supremum over all $n \in \mathbb{N}$ such that there is a chain

$$\emptyset \neq Y_0 \subsetneq Y_1 \subsetneq \cdots \subsetneq Y_n \subset X$$

of length $n$ of irreducible closed subsets $Y_1, \dots, Y_n$ of $X$.

(b) If $Y \subset X$ is a non-empty irreducible closed subset of $X$ the **codimension** $\operatorname{codim}_X Y$ of $Y$ in $X$ is again the supremum over all $n$ such that there is a chain

$$Y \subset Y_0 \subsetneq Y_1 \subsetneq \cdots \subsetneq Y_n \subset X$$

of irreducible closed subsets $Y_1, \dots, Y_n$ of $X$ containing $Y$.

To avoid confusion, we will always denote the dimension of a $K$-vector space $V$ by $\dim_K V$, leaving the notation $\dim X$ (without an index) for the dimension of a topological space $X$ as above.

According to the above idea, one should imagine each $Y_i$ as having dimension $i$ in a maximal chain as in Definition 2.23 (a), so that finally $\dim X = n$. In the same way, each $Y_i$ in Definition 2.23 (b) should have dimension $i + \dim Y$ in a maximal chain, so that $n = \dim X - \dim Y$ can be thought of as the difference of the dimensions of $X$ and $Y$.

**Example 2.24.**

(a) If $X$ is a (non-empty) finite affine variety then $\dim X = 0$. In fact, since points are closed in $X$ all subsets of $X$ will be closed, and thus the only irreducible closed subsets of $X$ are single points. There are therefore only chains of length 0 of irreducible closed subsets of $X$.

(b) In contrast to (a), general finite topological spaces need not have dimension 0. For example, the two-pointed topological space $X = \{a,b\}$ whose closed subsets are exactly $\emptyset$, $\{a\}$, and $X$ has dimension 1 since $\{a\} \subsetneq X$ is a chain of length 1 of irreducible closed subsets of $X$ (and there are certainly no longer ones).

However, this will not be of further importance for us since all topological spaces occurring in this class will have the property that points are closed.

(c) By Example 1.6 the affine space $\mathbb{A}^1$ has dimension 1: maximal chains of irreducible closed subsets of $\mathbb{A}^1$ are $\{a\} \subsetneq \mathbb{A}^1$ for any $a \in \mathbb{A}^1$.

(d) It is easy to see that the affine space $\mathbb{A}^n$ for $n \in \mathbb{N}_{>0}$ has dimension *at least* $n$, since there is certainly a chain

$$V(x_1,\dots,x_n) \subsetneq V(x_2,\dots,x_n) \subsetneq \cdots \subsetneq V(x_n) \subsetneq V(0) = \mathbb{A}^n$$

of irreducible (linear) closed subsets of $\mathbb{A}^n$ of length $n$.

Of course, we would expect geometrically that the dimension of $\mathbb{A}^n$ is *equal* to $n$. Although this turns out to be true, the proof of this result is unfortunately rather difficult and technical. It is given in the "Commutative Algebra" class, where dimension is one of the major topics. In fact, our Definition 2.23 is easy to translate into commutative algebra: since irreducible closed subvarieties of an affine variety $X$ correspond exactly to prime ideals in $A(X)$ by Remark 2.10, the dimension of $X$ is the supremum over all $n$ such that there is a chain $I_0 \supsetneq I_1 \supsetneq \cdots \supsetneq I_n$ of prime ideals in $A(X)$ — and this can be studied algebraically.

Let us now quote the results on the dimension of affine varieties that we will use from commutative algebra. They are all very intuitive: besides the statement that $\dim \mathbb{A}^n = n$ they say that for irreducible affine varieties the codimension of $Y$ in $X$ is in fact the difference of the dimensions of $X$ and $Y$, and that cutting down an irreducible affine variety by one non-trivial equation reduces the dimension by exactly 1.

**Proposition 2.25** (Properties of dimension)**.** *Let $X$ and $Y$ be non-empty irreducible affine varieties.*

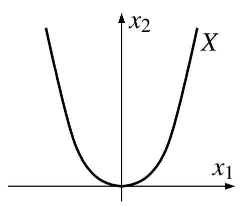
(a) *We have* $\dim(X \times Y) = \dim X + \dim Y$. *In particular,* $\dim \mathbb{A}^n = n$.

(b) *If $Y \subset X$ we have* $\dim X = \dim Y + \operatorname{codim}_X Y$. *In particular,* $\operatorname{codim}_X \{a\} = \dim X$ *for every point $a \in X$.*

(c) *If $f \in A(X)$ is non-zero every irreducible component of $V(f)$ has codimension* 1 *in $X$ (and hence dimension* $\dim X - 1$ *by (b)).*

*Proof.* Statement (a) is [G5, Proposition 11.9 (a) and Exercise 11.33 (a)], (b) is [G5, Example 11.13 (a)], and (c) is [G5, Corollary 11.19]. □

**Example 2.26.** Let $X = V(x_2 - x_1^2) \subset \mathbb{A}^2_{\mathbb{C}}$ be the affine variety whose real points are shown in the picture on the right. Then we have as expected:




(a) $X$ is irreducible by Proposition 2.9 since its coordinate ring $A(X) = \mathbb{C}[x_1,x_2]/(x_2 - x_1^2) \cong \mathbb{C}[x_1]$ is an integral domain.

(b) $X$ has dimension 1 by Proposition 2.25 (c), since it is the zero locus of one non-zero polynomial in the affine space $\mathbb{A}^2$, and $\dim \mathbb{A}^2 = 2$ by Proposition 2.25 (a).

**Remark 2.27** (Infinite-dimensional spaces)**.** One might be tempted to think that the "finiteness condition" of a Noetherian topological space $X$ ensures that $\dim X$ is always finite. This is not true however: if we equip the natural numbers $X = \mathbb{N}$ with the topology in which (except $\emptyset$ and $X$) exactly the subsets $Y_n := \{0,\dots,n\}$ for $n \in \mathbb{N}$ are closed, then $X$ is Noetherian, but has chains $Y_0 \subsetneq Y_1 \subsetneq \cdots \subsetneq Y_n$ of non-empty irreducible closed subsets of arbitrary length.

However, Proposition 2.25 (a) together with the following exercise shows that this cannot happen for arbitrary subsets of affine varieties. In fact, all topological spaces considered in this class will have finite dimension.

**Exercise 2.28.** Let $A$ be an arbitrary subset of a topological space $X$. Prove that $\dim A \le \dim X$.

**Remark 2.29.** Depending on where our chains of irreducible closed subvarieties end resp. start, we can break up the supremum in Definition 2.23 into irreducible components or local contributions:

(a) If $X = X_1 \cup \cdots \cup X_r$ is the irreducible decomposition of a Noetherian topological space as in Proposition 2.15, we have

$$\dim X = \max\{\dim X_1,\dots,\dim X_r\}:$$

"$\le$" Assume that $\dim X \ge n$, so that there is a chain $Y_0 \subsetneq \cdots \subsetneq Y_n$ of non-empty irreducible closed subvarieties of $X$. Then $Y_n = (Y_n \cap X_1) \cup \cdots \cup (Y_n \cap X_r)$ is a union of closed subsets. So as $Y_n$ is irreducible we must have $Y_n = Y_n \cap X_i$, and hence $Y_n \subset X_i$, for some $i$. But then $Y_0 \subsetneq \cdots \subsetneq Y_n$ is a chain of non-empty irreducible closed subsets in $X_i$, and hence $\dim X_i \ge n$.

"$\ge$" Let $\max\{\dim X_1,\dots,\dim X_r\} \ge n$. Then there is a chain of non-empty irreducible closed subsets $Y_0 \subsetneq \cdots \subsetneq Y_n$ in some $X_i$. This is also such a chain in $X$, and hence $\dim X \ge n$.

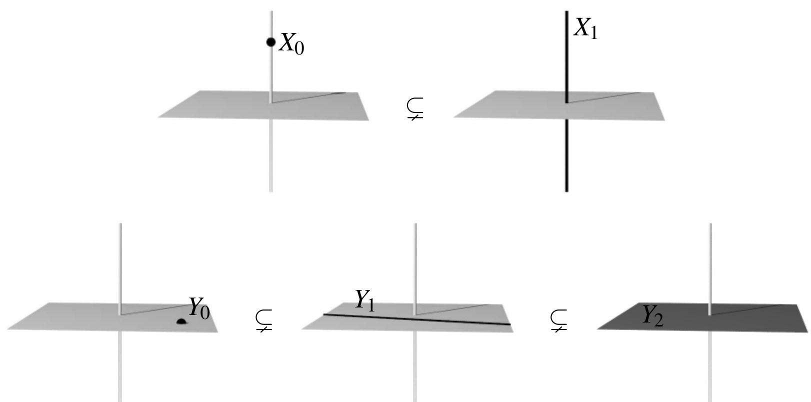
So for many purposes it suffices to consider the dimension of irreducible spaces.

(b) We always have $\dim X = \sup\{\operatorname{codim}_X\{a\} : a \in X\}$:

"$\le$" If $\dim X \ge n$ there is a chain $Y_0 \subsetneq \cdots \subsetneq Y_n$ of non-empty irreducible closed subsets of $X$. For any $a \in Y_0$ this chain then shows that $\operatorname{codim}_X\{a\} \ge n$.

"$\ge$" If $\operatorname{codim}_X\{a\} \ge n$ for some $a \in X$ there is a chain $\{a\} \subset Y_0 \subsetneq \cdots \subsetneq Y_n$ of non-empty irreducible closed subsets of $X$, which also shows that $\dim X \ge n$.

The picture below illustrates these two equations: the affine variety $X = V(x_1x_3, x_2x_3) \subset \mathbb{A}^3$ is a union of two irreducible components, a line $V(x_1,x_2)$ of dimension 1 and a plane $V(x_3)$ of dimension 2 (see Proposition 2.25 (a)). So by (a) we have $\dim X = 2$ (with a maximal chain of length 2 as in Definition 2.23 (a) given by $Y_0 \subsetneq Y_1 \subsetneq Y_2$).




As for (b), the codimension of the point $Y_0$ is 2, whereas the codimension of the point $X_0$ is 1, as illustrated by the chains in the picture. Note that this codimension of a point can be interpreted geometrically as the *local dimension* of $X$ at this point. Hence Proposition 2.25 (b) can also be interpreted as saying that the local dimension of an irreducible variety is the same at every point.

In practice, we will usually be concerned with affine varieties all of whose components have the same dimension. These spaces have special names that we want to introduce now. Note however that (as with the definition of a variety, see Remark 1.3) these terms are not used consistently throughout the literature — sometimes e. g. a curve is required to be irreducible, and sometimes it might be allowed to have additional components of dimension less than 1.

**Definition 2.30** (Pure-dimensional spaces)**.**

(a) A Noetherian topological space $X$ is said to be of **pure dimension** $n$ if every irreducible component of $X$ has dimension $n$.

(b) An affine variety is called ...

- an **affine curve** if it is of pure dimension 1;
- an **affine surface** if it is of pure dimension 2;
- an **affine hypersurface** of an irreducible affine variety $Y \supset X$ if it is of pure dimension $\dim Y - 1$.

**Exercise 2.31.** Let $X$ be the set of all $2 \times 3$ matrices over a field $K$ that have rank at most 1, considered as a subset of $\mathbb{A}^6 = \mathrm{Mat}(2 \times 3, K)$.

Show that $X$ is an affine variety. Is it irreducible? What is its dimension?

**Exercise 2.32.** Show that the ideal $I = (x_1x_2, x_1x_3, x_2x_3) \trianglelefteq \mathbb{C}[x_1, x_2, x_3]$ cannot be generated by fewer than three elements. What is the zero locus of $I$?

**Exercise 2.33.** Let $X$ be a topological space. Prove:

(a) If $\{U_i : i \in I\}$ is an open cover of $X$ then $\dim X = \sup\{\dim U_i : i \in I\}$.

(b) If $X$ is an irreducible affine variety and $U \subset X$ a non-empty open subset then $\dim X = \dim U$.

Does this statement hold more generally for any irreducible topological space?

**Exercise 2.34.** Prove the following (maybe at first surprising) statements:

(a) Every affine variety in the real affine space $\mathbb{A}^n_{\mathbb{R}}$ is the zero locus of one polynomial.

(b) Every Noetherian topological space is compact. In particular, every open subset of an affine variety is compact in the Zariski topology. (Recall that by definition a topological space $X$ is compact if every open cover of $X$ has a finite subcover.)

(c) The zero locus of a non-constant polynomial in $\mathbb{C}[x_1, x_2]$ is never compact in the classical topology of $\mathbb{A}^2_{\mathbb{C}} = \mathbb{C}^2$.

(For those of you who know commutative algebra: can you prove that an affine variety over $\mathbb{C}$ containing infinitely many points is *never* compact in the classical topology?)

# 3. The Sheaf of Regular Functions

After having defined affine varieties, our next goal must be to say what kind of maps between them we want to consider as morphisms, i. e. as "nice maps preserving the structure of the variety". In this chapter we will look at the easiest case of this: the so-called *regular functions*, i. e. maps to the ground field $K = \mathbb{A}^1$. They should be thought of as the analogue of continuous functions in topology, differentiable functions in real analysis, or holomorphic functions in complex analysis.

So what kind of nice "algebraic" functions should we consider on an affine variety $X$? First of all, as in the case of continuous or differentiable functions, we should not only aim for a definition of functions on all of $X$, but also on an arbitrary open subset $U$ of $X$. In contrast to the coordinate ring $A(X)$ of polynomial functions on the whole space $X$, this allows us to consider quotients $\frac{g}{f}$ of polynomial functions $f, g \in A(X)$ with $f \neq 0$ as well, since we can exclude the zero set $V(f)$ of the denominator from the domain of definition of the function.

But taking our functions to be quotients of polynomials turns out to be a little bit too restrictive. The problem with this definition would be that *it is not local*: recall that the condition on a function to be continuous or differentiable is local in the sense that it can be checked at every point, with the whole function then being continuous or differentiable if it has this property at every point. Being a quotient of polynomials however is not a condition of this type — we would have to find one *global* representation as a quotient of polynomials that is then valid at every point. Imposing such non-local conditions is usually not a good thing to do, since it would be hard in practice to find the required global representations of the functions.

The way out of this problem is to consider functions that are only *locally* quotients of polynomials, i. e. functions $\varphi : U \to K$ such that each point $a \in U$ has a neighborhood in $U$ in which $\varphi = \frac{g}{f}$ holds for two polynomials $f$ and $g$ (that may depend on $a$). In fact, we will see in Example 3.5 that passing from *global* to *local* quotients of polynomials really makes a difference. So let us now give the corresponding formal definition of regular functions.

**Definition 3.1** (Regular functions)**.** Let $X$ be an affine variety, and let $U$ be an open subset of $X$. A **regular function** on $U$ is a map $\varphi : U \to K$ with the following property: for every $a \in U$ there are polynomial functions $f, g \in A(X)$ with $f(x) \neq 0$ and

$$\varphi(x) = \frac{g(x)}{f(x)}$$

for all $x$ in an open subset $U_a$ with $a \in U_a \subset U$. The set of all such regular functions on $U$ will be denoted $\mathscr{O}_X(U)$.

**Notation 3.2.** We will usually write the condition "$\varphi(x) = \frac{g(x)}{f(x)}$ for all $x \in U_a$" of Definition 3.1 simply as "$\varphi = \frac{g}{f}$ on $U_a$". This is certainly an intuitive notation that should not lead to any confusion. However, a word of warning in particular for those of you who know commutative algebra already: this also means that (unless stated otherwise) the fraction $\frac{g}{f}$ of two elements of $A(X)$ will always denote the pointwise quotient of the two corresponding polynomial functions — and not the algebraic concept of an element in a localized ring as introduced later in Construction 3.12.

**Remark 3.3** ($\mathscr{O}_X(U)$ as a ring and $K$-algebra)**.** It is obvious that the set $\mathscr{O}_X(U)$ of regular functions on an open subset $U$ of an affine variety $X$ is a ring with pointwise addition and multiplication. However, it has an additional structure: it is also a $K$-vector space since we can multiply a regular function pointwise with a fixed scalar in $K$. In algebraic terms, this means that $\mathscr{O}_X(U)$ is a $K$-algebra, which is defined as follows.

**Definition 3.4** ($K$-algebras [G5, Definition 1.23 and Remark 1.24])**.**

(a) A $K$**-algebra** is a ring $R$ that is at the same time a $K$-vector space such that the ring multiplication is $K$-bilinear.

(b) For two $K$-algebras $R$ and $R'$ a **morphism** (or $K$**-algebra homomorphism**) from $R$ to $R'$ is a map $f: R \to R'$ that is a ring homomorphism as well as a $K$-linear map.

**Example 3.5** (Local $\neq$ global quotients of polynomials)**.** Consider the 3-dimensional affine variety $X = V(x_1x_4 - x_2x_3) \subset \mathbb{A}^4$ and the open subset

$$U = X \backslash V(x_2,x_4) = \{(x_1,x_2,x_3,x_4) \in X : x_2 \neq 0 \text{ or } x_4 \neq 0\} \quad \subset X.$$

Then

$$\varphi: U \to K, \ (x_1,x_2,x_3,x_4) \mapsto \begin{cases} \frac{x_1}{x_2} & \text{if } x_2 \neq 0, \\ \frac{x_3}{x_4} & \text{if } x_4 \neq 0 \end{cases} \tag{$*$}$$

is a regular function on $U$: it is well-defined since the defining equation for $X$ implies $\frac{x_1}{x_2} = \frac{x_3}{x_4}$ whenever $x_2 \neq 0$ and $x_4 \neq 0$, and it is obviously locally a quotient of polynomials. But none of the two representations in $(*)$ as quotients of polynomials can be used on all of $U$, since the first one does not work e. g. at the point $(0,0,0,1) \in U$, whereas the second one does not work at $(0,1,0,0) \in U$. Algebraically, this is just the statement that $A(X)$ is not a unique factorization domain [G5, Definition 8.1] because of the relation $x_1x_4 = x_2x_3$.

In fact, one can show that there is also no other global representation of $\varphi$ as a quotient of two polynomials. We will not need this statement here, and so we do not prove it — we should just keep in mind that representations of regular functions as quotients of polynomials will in general not be valid on the complete domain of definition of the function.

04

As a first result, let us prove the expected statement that zero loci of regular functions are always closed in their domain of definition.

**Lemma 3.6** (Zero loci of regular functions are closed)**.** *Let $U$ be an open subset of an affine variety $X$, and let $\varphi \in \mathscr{O}_X(U)$ be a regular function on $U$. Then*

$$V(\varphi) := \{x \in U : \varphi(x) = 0\}$$

*is closed in $U$.*

*Proof.* By Definition 3.1 any point $a \in U$ has an open neighborhood $U_a$ in $U$ on which $\varphi = \frac{g_a}{f_a}$ for some $f_a, g_a \in A(X)$ (with $f_a$ nowhere zero on $U_a$). So the set

$$\{x \in U_a : \varphi(x) \neq 0\} = U_a \backslash V(g_a)$$

is open in $X$, and hence so is their union over all $a \in U$, which is just $U \backslash V(\varphi)$. This means that $V(\varphi)$ is closed in $U$. □

**Remark 3.7** (Identity Theorem for regular functions)**.** A simple but remarkable consequence of Lemma 3.6 is the following: let $U \subset V$ be non-empty open subsets of an irreducible affine variety $X$. If $\varphi_1, \varphi_2 \in \mathscr{O}_X(V)$ are two regular functions on $V$ that agree on $U$, then they must agree on all of $V$: the locus $V(\varphi_1 - \varphi_2)$ where the two functions agree contains $U$ and is closed in $V$ by Lemma 3.6, hence it also contains the closure $\overline{U}$ of $U$ in $V$. But $\overline{V} = X$ by Remark 2.18 (b), hence $V$ is irreducible by Exercise 2.19 (b), which again by Remark 2.18 (b) means that the closure of $U$ in $V$ is $V$. Consequently, we have $\varphi_1 = \varphi_2$ on $V$.

Note that this statement is not really surprising since the open subsets in the Zariski topology are so big: over the ground field $\mathbb{C}$, for example, it is also true in the classical topology that the closure of $U$ in $V$ is $V$, and hence the equation $\varphi_1 = \varphi_2$ on $V$ already follows from $\varphi_1|_U = \varphi_2|_U$ by the (classical) continuity of $\varphi_1$ and $\varphi_2$. The interesting fact here is that the very same statement holds in complex analysis for holomorphic functions as well (or more generally, in real analysis for analytic functions): two holomorphic functions on a (connected) open subset $U \subset \mathbb{C}^n$ must be the same if they agree on any smaller open subset $V \subset U$. This is called the *Identity Theorem* for holomorphic functions. In complex analysis this is a real theorem because the open subset $V$ can be very small, so the statement that the extension to $U$ is unique is a lot more surprising than it is here in algebraic geometry. Still this is an example of a theorem that is true in literally the same way in both algebraic and complex geometry, although these two theories are quite different a priori. We will see another case of this in Example 3.14.

Let us now go ahead and compute the $K$-algebras $\mathscr{O}_X(U)$ in some cases. A particularly important result in this direction can be obtained if $U$ is the complement of the zero locus of a single polynomial function $f \in A(X)$. In this case it turns out that (in contrast to Example 3.5) the regular functions on $U$ can always be written with a single representation as a fraction, whose denominator is in addition a power of $f$.

**Definition 3.8** (Distinguished open subsets)**.** For an affine variety $X$ and a polynomial function $f \in A(X)$ on $X$ we call

$$D(f) := X \backslash V(f) = \{x \in X : f(x) \neq 0\}$$

the **distinguished open subset** of $f$ in $X$.

**Remark 3.9.** The distinguished open subsets of an affine variety $X$ behave nicely with respect to intersections and unions:

(a) For any $f, g \in A(X)$ we have $D(f) \cap D(g) = D(fg)$, since $f(x) \neq 0$ and $g(x) \neq 0$ is equivalent to $(fg)(x) \neq 0$ for all $x \in X$. In particular, finite intersections of distinguished open subsets are again distinguished open subsets.

(b) Any open subset $U \subset X$ is a finite union of distinguished open subsets: by definition of the Zariski topology it is the complement of an affine variety, which in turn is the zero locus of finitely many polynomial functions $f_1, \dots, f_k \in A(X)$ by Proposition 1.21 (a). Hence we have

$$U = X \backslash V(f_1, \dots, f_k) = D(f_1) \cup \dots \cup D(f_k).$$

We can therefore think of the distinguished open subsets as the "smallest" open subsets of $X$ — in topology, the correct notion for this would be to say that they form a *basis* of the Zariski topology on $X$.

**Proposition 3.10** (Regular functions on distinguished open subsets)**.** *Let $X$ be an affine variety, and let $f \in A(X)$. Then*

$$\mathscr{O}_X(D(f)) = \left\{ \frac{g}{f^n} : g \in A(X), n \in \mathbb{N} \right\}.$$

*In particular, setting $f = 1$ we see that $\mathscr{O}_X(X) = A(X)$, i. e. the regular functions on all of $X$ are exactly the polynomial functions.*

*Proof.* The inclusion "$\supset$" is obvious: every function of the form $\frac{g}{f^n}$ for $g \in A(X)$ and $n \in \mathbb{N}$ is clearly regular on $D(f)$.

For the opposite inclusion "$\subset$", let $\varphi : D(f) \to K$ be a regular function. By Definition 3.1 we obtain for every $a \in D(f)$ a local representation $\varphi = \frac{g_a}{f_a}$ for some $f_a, g_a \in A(X)$ which is valid on an open neighborhood of $a$ in $D(f)$. After possibly shrinking these neighborhoods we may assume by Remark 3.9 (b) that they are distinguished open subsets $D(h_a)$ for some $h_a \in A(X)$. Moreover, we can change the representations of $\varphi$ by replacing $g_a$ and $f_a$ by $g_a h_a$ and $f_a h_a$ (which does not change their quotient on $D(h_a)$) to assume that both $g_a$ and $f_a$ vanish on the complement $V(h_a)$ of $D(h_a)$. Finally, this means that $f_a$ vanishes on $V(h_a)$ and does not vanish on $D(h_a)$ — so $h_a$ and $f_a$ have the same zero locus, and we can therefore assume that $h_a = f_a$.

As a consequence, note that

$$g_a f_b = g_b f_a \quad \text{for all } a, b \in D(f): \qquad (*)$$

these two functions agree on $D(f_a) \cap D(f_b)$ since $\varphi = \frac{g_a}{f_a} = \frac{g_b}{f_b}$ there, and they are both zero otherwise since by our construction we have $g_a(x) = f_a(x) = 0$ for all $x \in V(f_a)$ and $g_b(x) = f_b(x) = 0$ for all $x \in V(f_b)$.

Now all our open neighborhoods cover $D(f)$, i. e. we have $D(f) = \bigcup_{a \in D(f)} D(f_a)$. Passing to the complement we obtain

$$V(f) = \bigcap_{a \in D(f)} V(f_a) = V\big(\{f_a : a \in D(f)\}\big),$$

and thus by Proposition 1.21 (b)

$$f \in I(V(f)) = I\big(V\big(\{f_a : a \in D(f)\}\big)\big) = \sqrt{(f_a : a \in D(f))}.$$

This means that $f^n = \sum_a k_a f_a$ for some $n \in \mathbb{N}$ and $k_a \in A(X)$ for finitely many elements $a \in D(f)$. Setting $g := \sum_a k_a g_a$, we then claim that $\varphi = \frac{g}{f^n}$ on all of $D(f)$: for all $b \in D(f)$ we have $\varphi = \frac{g_b}{f_b}$ and

$$g f_b = \sum_a k_a g_a f_b \overset{(*)}{=} \sum_a k_a g_b f_a = g_b f^n$$

on $D(f_b)$, and these open subsets cover $D(f)$. □

**Remark 3.11.** In the proof of Proposition 3.10 we had to use Hilbert's Nullstellensatz again. In fact, the statement is false if the ground field is not algebraically closed, as you can see from the example of the function $\frac{1}{x^2+1}$ that is regular on all of $\mathbb{A}^1_{\mathbb{R}}$, but not a polynomial function.

Proposition 3.10 is deeply linked to commutative algebra. Although we considered the quotients $\frac{g}{f^n}$ in this statement to be fractions of polynomial functions, there is also a purely algebraic construction of fractions in a (polynomial) ring — in the same way as we could regard the elements of $A(X)$ either geometrically as functions on $X$ or algebraically as elements in the quotient ring $K[x_1, \dots, x_n]/I(X)$. In these notes we will mainly use the geometric interpretation as functions, but it is still instructive to see the corresponding algebraic construction. For reasons that will become apparent in Lemma 3.21 it is called *localization*; it is also one of the central topics in the "Commutative Algebra" class.

**Construction 3.12** (Localizations [G5, Chapter 6]**).** Let $R$ be a ring. A **multiplicatively closed subset** of $R$ is a subset $S \subset R$ with $1 \in S$ and $fg \in S$ for all $f, g \in S$.

For such a multiplicatively closed subset $S$ we then consider pairs $(g, f)$ with $g \in R$ and $f \in S$, and call two such pairs $(g, f)$ and $(g', f')$ equivalent if there is an element $h \in S$ with $h(gf' - g'f) = 0$. The equivalence class of a pair $(g, f)$ will formally be written as a fraction $\frac{g}{f}$, the set of all such equivalence classes is denoted $S^{-1}R$. Together with the usual rules for the addition and multiplication of fractions, $S^{-1}R$ is again a ring. It is called the **localization** of $R$ at $S$.

By construction we can think of the elements of $S^{-1}R$ as formal fractions, with the numerators in $R$ and the denominators in $S$. In Proposition 3.10 our set of denominators is $S = \{f^n : n \in \mathbb{N}\}$ for a fixed element $f \in R$; in this case the localization $S^{-1}R$ is usually written as $R_f$. We will meet other sets of denominators later in Lemma 3.21 and Exercise 9.8 (a).

So let us now prove rigorously that the $K$-algebra $\mathscr{O}_X(D(f))$ of Proposition 3.10 can also be interpreted algebraically as a localization.

**Lemma 3.13** (Regular functions as localizations)**.** *Let $X$ be an affine variety, and let $f \in A(X)$. Then $\mathscr{O}_X(D(f))$ is isomorphic (as a $K$-algebra) to the localized ring $A(X)_f$.*

*Proof.* There is an obvious $K$-algebra homomorphism

$$A(X)_f \to \mathscr{O}_X(D(f)), \ \frac{g}{f^n} \mapsto \frac{g}{f^n}$$

that interprets a formal fraction in the localization $A(X)_f$ as an actual quotient of polynomial functions on $D(f)$. It is in fact well-defined: if $\frac{g}{f^n} = \frac{g'}{f^m}$ as formal fractions in the localization $A(X)_f$ then $f^k(gf^m - g'f^n) = 0$ in $A(X)$ for some $k \in \mathbb{N}$, which means that $gf^m = g'f^n$ and thus $\frac{g}{f^n} = \frac{g'}{f^m}$ as functions on $D(f)$.

The homomorphism is surjective by Proposition 3.10. It is also injective: if $\frac{g}{f^n} = 0$ as a function on $D(f)$ then $g = 0$ on $D(f)$ and hence $fg = 0$ on all of $X$, which means $f(g \cdot 1 - 0 \cdot f^n) = 0$ in $A(X)$ and thus $\frac{g}{f^n} = \frac{0}{1}$ as formal fractions in the localization $A(X)_f$. □

**Example 3.14** (Regular functions on $\mathbb{A}^2\backslash\{0\}$)**.** Probably the easiest case of an open subset of an affine variety that is not a distinguished open subset is the complement $U = \mathbb{A}^2\backslash\{0\}$ of the origin in the affine plane $X = \mathbb{A}^2$. We are going to see that

$$\mathscr{O}_{\mathbb{A}^2}(\mathbb{A}^2\backslash\{0\}) = K[x_1,x_2]$$

and thus that $\mathscr{O}_X(U) = \mathscr{O}_X(X)$, i. e. every regular function on $U$ can be extended to $X$. Note that this is another result that is true in the same way in complex analysis: there is a *Removable Singularity Theorem* that implies that every holomorphic function on $\mathbb{C}^2\backslash\{0\}$ can be extended holomorphically to $\mathbb{C}^2$.

To prove our claim let $\varphi \in \mathscr{O}_X(U)$. Then $\varphi$ is regular on the distinguished open subsets $D(x_1) = (\mathbb{A}^1\backslash\{0\}) \times \mathbb{A}^1$ and $D(x_2) = \mathbb{A}^1 \times (\mathbb{A}^1\backslash\{0\})$, and so by Proposition 3.10 we can write $\varphi = \frac{f}{x_1^m}$ on $D(x_1)$ and $\varphi = \frac{g}{x_2^n}$ on $D(x_2)$ for some $f,g \in K[x_1,x_2]$ and $m,n \in \mathbb{N}$. Of course we can do this so that $x_1 \nmid f$ and $x_2 \nmid g$.

On the intersection $D(x_1) \cap D(x_2)$ both representations of $\varphi$ are valid, and so we have $f x_2^n = g x_1^m$ on $D(x_1) \cap D(x_2)$. But the locus $V(f x_2^n - g x_1^m)$ where this equation holds is closed, and hence we see that $f x_2^n = g x_1^m$ also on $\overline{D(x_1) \cap D(x_2)} = \mathbb{A}^2$. In other words, we have $f x_2^n = g x_1^m$ in the polynomial ring $A(\mathbb{A}^2) = K[x_1,x_2]$.

Now if we had $m > 0$ then $x_1$ must divide $f x_2^n$, which is clearly only possible if $x_1 \,|\, f$. This is a contradiction, and so it follows that $m = 0$. But then $\varphi = f$ is a polynomial, as we have claimed.

**Exercise 3.15.** For those of you who know some commutative algebra already: generalize the proof of Example 3.14 to show that $\mathscr{O}_X(U) = \mathscr{O}_X(X) = A(X)$ for any open subset $U$ of an affine variety $X$ such that $A(X)$ is a unique factorization domain [G5, Definition 8.1] and $U$ is the complement of an irreducible subvariety of codimension at least 2 in $X$.

Recall that we have defined regular functions on an open subset $U$ of an affine variety as set-theoretic functions from $U$ to the ground field $K$ that satisfy some local property. Local constructions of function-like objects occur in many places in algebraic geometry (and also in other "topological" fields of mathematics), and so we will spend the rest of this chapter to formalize the idea of such objects. This will have the advantage that it gives us an "automatic" definition of morphisms between affine varieties in Chapter 4, and in fact also between more general varieties in Chapter 5.

**Definition 3.16** (Sheaves)**.** A **presheaf** $\mathscr{F}$ (of rings) on a topological space $X$ consists of the data:

- for every open set $U \subset X$ a ring $\mathscr{F}(U)$ (think of this as the ring of functions on $U$),
- for every inclusion $U \subset V$ of open sets in $X$ a ring homomorphism $\rho_{V,U} : \mathscr{F}(V) \to \mathscr{F}(U)$ called the *restriction map* (think of this as the usual restriction of functions to a subset),

such that

- $\mathscr{F}(\emptyset) = 0$,
- $\rho_{U,U}$ is the identity map on $\mathscr{F}(U)$ for all $U$,
- for any inclusion $U \subset V \subset W$ of open sets in $X$ we have $\rho_{V,U} \circ \rho_{W,V} = \rho_{W,U}$.

The elements of $\mathscr{F}(U)$ are usually called the **sections** of $\mathscr{F}$ over $U$, and the restriction maps $\rho_{V,U}$ are written as $\varphi \mapsto \varphi|_U$.

A presheaf $\mathscr{F}$ is called a **sheaf** of rings if it satisfies the following gluing property: if $U \subset X$ is an open set, $\{U_i : i \in I\}$ an arbitrary open cover of $U$ and $\varphi_i \in \mathscr{F}(U_i)$ sections for all $i$ such that $\varphi_i|_{U_i \cap U_j} = \varphi_j|_{U_i \cap U_j}$ for all $i,j \in I$, then there is a unique $\varphi \in \mathscr{F}(U)$ such that $\varphi|_{U_i} = \varphi_i$ for all $i$.

05

**Example 3.17.** Intuitively speaking, any "function-like" object forms a presheaf; it is a sheaf if the conditions imposed on the "functions" are local (i. e. if they can be checked on an open cover). The following examples illustrate this.

(a) Let $X$ be an affine variety. Then the rings $\mathscr{O}_X(U)$ of regular functions on open subsets $U \subset X$, together with the usual restriction maps of functions, form a sheaf $\mathscr{O}_X$ on $X$. In fact, the presheaf axioms are obvious, and the gluing property just means that a function $\varphi : U \to K$ is regular if it is regular on each element of an open cover of $U$ (which follows from the definition that $\varphi$ is regular if it is *locally* a quotient of polynomial functions). We call $\mathscr{O}_X$ the *sheaf of regular functions* on $X$.

(b) Similarly, on $X = \mathbb{R}^n$ the rings

$$\mathscr{F}(U) = \{\varphi : U \to \mathbb{R} \text{ continuous}\}$$

for open subsets $U \subset X$ form a sheaf $\mathscr{F}$ on $X$ with the usual restriction maps. In the same way we can consider on $X$ the sheaves of differentiable functions, analytic functions, arbitrary functions, and so on.

(c) On $X = \mathbb{R}^n$ let

$$\mathscr{F}(U) = \{\varphi : U \to \mathbb{R} \text{ constant function}\}$$

with the usual restriction maps. Then $\mathscr{F}$ is a presheaf, but not a sheaf, since being a constant function is not a local condition. More precisely, let $U_1$ and $U_2$ be any non-empty disjoint open subsets of $X$, and let $\varphi_1 \in \mathscr{F}(U_1)$ and $\varphi_2 \in \mathscr{F}(U_2)$ be constant functions with different values. Then $\varphi_1$ and $\varphi_2$ trivially agree on $U_1 \cap U_2 = \emptyset$, but there is still no *constant* function on $U = U_1 \cup U_2$ that restricts to both $\varphi_1$ on $U_1$ and $\varphi_2$ on $U_2$. Hence $\mathscr{F}$ does not satisfy the gluing property. Note however that we would obtain a sheaf if we considered *locally constant* functions instead of constant ones.

In order to get used to the language of sheaves let us now consider two common constructions with them.

**Definition 3.18** (Restrictions of (pre-)sheaves)**.** Let $\mathscr{F}$ be a presheaf on a topological space $X$, and let $U \subset X$ be an open subset. Then the **restriction** of $\mathscr{F}$ to $U$ is defined to be the presheaf $\mathscr{F}|_U$ on $U$ with

$$\mathscr{F}|_U(V) := \mathscr{F}(V)$$

for every open subset $V \subset U$, and with the restriction maps taken from $\mathscr{F}$. Note that if $\mathscr{F}$ is a sheaf then so is $\mathscr{F}|_U$.

**Construction 3.19** (Stalks of (pre-)sheaves)**.** Again let $\mathscr{F}$ be a presheaf on a topological space $X$. Fix a point $a \in X$ and consider pairs $(U, \varphi)$ where $U$ is an open neighborhood of $a$ and $\varphi \in \mathscr{F}(U)$. We call two such pairs $(U, \varphi)$ and $(U', \varphi')$ equivalent if there is an open subset $V$ with $a \in V \subset U \cap U'$ and $\varphi|_V = \varphi'|_V$ (it is easy to check that this is indeed an equivalence relation). The set of all such pairs modulo this equivalence relation is called the **stalk** $\mathscr{F}_a$ of $\mathscr{F}$ at $a$; it inherits a ring structure from the rings $\mathscr{F}(U)$. The elements of $\mathscr{F}_a$ are called **germs** of $\mathscr{F}$ at $a$.

**Remark 3.20.** The geometric interpretation of the germs of a sheaf is that they are functions (resp. sections of the sheaf) that are defined in an arbitrarily small neighborhood of the given point — we will also refer to these objects as *local functions* at this point. Hence e. g. on the real line the germ of a differentiable function at a point $a$ allows you to compute the derivative of this function at $a$, but none of the actual values of the function at any point except $a$. On the other hand, we have seen in Remark 3.7 that holomorphic functions on a (connected) open subset of $\mathbb{C}^n$ are already determined by their values on any smaller open set. So in this sense germs of holomorphic functions carry "much more information" than germs of differentiable functions.

In algebraic geometry, the situation is similar: let $\varphi_1$ and $\varphi_2$ be two regular functions on an open subset $U$ of an irreducible affine variety $X$. If there is a point of $U$ at which the germs of $\varphi_1$ and $\varphi_2$ are the same then $\varphi_1$ and $\varphi_2$ have to agree on a non-empty open subset, which means by Remark 3.7 that $\varphi_1 = \varphi_2$ on $U$. In other words, the germ of a regular function determines the function uniquely already. Note that the corresponding statement is clearly false for differentiable functions as we have seen above.

In fact, germs of regular functions on an affine variety $X$ can also be described algebraically in terms of localizations as introduced in Construction 3.12 — which is the reason why this algebraic concept is called "localization". As one might expect, such a germ at a point $a \in X$, i. e. a regular function in a small neighborhood of $a$, is given by an element in the localization of $A(X)$ for which we allow as denominators all polynomials that do not vanish at $a$.

**Lemma 3.21** (Germs of regular functions as localizations)**.** *Let $a$ be a point on an affine variety $X$, and let $S = \{f \in A(X) : f(a) \neq 0\}$. Then the stalk $\mathscr{O}_{X,a}$ of $\mathscr{O}_X$ at $a$ is isomorphic (as a $K$-algebra) to the localized ring*

$$S^{-1}A(X) = \left\{ \frac{g}{f} : f, g \in A(X), f(a) \neq 0 \right\}.$$

*It is called the **local ring** of $X$ at $a$.*

*Proof.* Note that $S$ is clearly multiplicatively closed, so that the localization $S^{-1}A(X)$ exists. Consider the $K$-algebra homomorphism

$$S^{-1}A(X) \to \mathscr{O}_{X,a}, \quad \frac{g}{f} \mapsto \overline{\left(D(f), \frac{g}{f}\right)}$$

that maps a formal fraction $\frac{g}{f}$ to the corresponding quotient of polynomial functions on the open neighborhood of $a$ where the denominator does not vanish. It is well-defined: if $\frac{g}{f} = \frac{g'}{f'}$ in the localization then $h(gf' - g'f) = 0$ for some $h \in S$. Hence the functions $\frac{g}{f}$ and $\frac{g'}{f'}$ agree on the open neighborhood $D(h)$ of $a$, and thus they determine the same element in the stalk $\mathscr{O}_{X,a}$.

The $K$-algebra homomorphism is surjective since by definition any regular function in a sufficiently small neighborhood of $a$ must be representable by a fraction $\frac{g}{f}$ with $g \in A(X)$ and $f \in S$. It is also injective: assume that a function $\frac{g}{f}$ represents the zero element in the stalk $\mathscr{O}_{X,a}$, i. e. that it is zero in an open neighborhood of $a$. By possibly shrinking this neighborhood we may assume by Remark 3.9 (b) that it is a distinguished open subset $D(h)$ containing $a$, i. e. with $h \in S$. But then $h(g \cdot 1 - 0 \cdot f)$ is zero on all of $X$, hence zero in $A(X)$, and thus $\frac{g}{f} = \frac{0}{1}$ in the localization $S^{-1}A(X)$. □

Local rings will become important later on when we construct tangent spaces (see Lemma 10.5) and vanishing multiplicities (see Definition 12.23). We will then mostly use their algebraic description of Lemma 3.21 and write the elements of $\mathscr{O}_{X,a}$ as quotients $\frac{g}{f}$ with $f, g \in A(X)$ such that $f(a) \neq 0$.

Algebraically, the most important property of the local ring $\mathscr{O}_{X,a}$ is that it has only one *maximal ideal* in the sense of the following lemma. In fact, in commutative algebra a local ring is *defined* to be a ring with only one maximal ideal.

**Lemma and Definition 3.22** (Maximal ideals)**.** *Let $a$ be a point on an affine variety $X$. Then every proper ideal of the local ring $\mathscr{O}_{X,a}$ is contained in the ideal*

$$I_a := I(a)\,\mathscr{O}_{X,a} := \left\{ \frac{g}{f} : f, g \in A(X), g(a) = 0, f(a) \neq 0 \right\}$$

*of all local functions vanishing at the point $a$. The ideal $I_a$ is therefore called the **maximal ideal** of $\mathscr{O}_{X,a}$.*

*Proof.* It is easily checked that $I_a$ is in fact an ideal. Now let $I \trianglelefteq \mathscr{O}_{X,a}$ be any ideal not contained in $I_a$. By definition, this means that there is an element $\frac{g}{f} \in I$ with $f(a) \neq 0$ and $g(a) \neq 0$. But then $\frac{f}{g}$ exists in $\mathscr{O}_{X,a}$ as well. Hence $1 = \frac{f}{g} \cdot \frac{g}{f} \in I$, and we conclude that $I = \mathscr{O}_{X,a}$. □

**Exercise 3.23.** Let $X \subset \mathbb{A}^n$ be an affine variety, and let $a \in X$. Show that $\mathscr{O}_{X,a} \cong \mathscr{O}_{\mathbb{A}^n,a}/I(X)\,\mathscr{O}_{\mathbb{A}^n,a}$, where $I(X)\,\mathscr{O}_{\mathbb{A}^n,a}$ denotes the ideal in $\mathscr{O}_{\mathbb{A}^n,a}$ generated by all quotients $\frac{f}{1}$ for $f \in I(X)$.

**Exercise 3.24.** Let $\mathscr{F}$ be a sheaf on a topological space $X$, and let $a \in X$. Show that the stalk $\mathscr{F}_a$ is a local object in the following sense: if $U \subset X$ is an open neighborhood of $a$ then $\mathscr{F}_a$ is isomorphic to the stalk of $\mathscr{F}|_U$ at $a$ on the topological space $U$.

**Remark 3.25** (Sheaves for other categories)**.** In Definition 3.16 we have constructed (pre-)sheaves of rings. In the same way one can define (pre-)sheaves of $K$-algebras, Abelian groups, or other suitable categories, by requiring that all $\mathscr{F}(U)$ are objects and all restriction maps are morphisms in the corresponding category. Note that the stalks of such a (pre-)sheaf then inherit this structure. For example, all our (pre-)sheaves considered so far have also been (pre-)sheaves of $K$-algebras for some field $K$, and thus their stalks are all $K$-algebras. In fact, starting in the next chapter we will restrict ourselves to this situation.

# 4. Morphisms

So far we have defined and studied regular functions on an affine variety $X$. They can be thought of as the morphisms (i. e. the "nice" maps) from open subsets of $X$ to the ground field $K = \mathbb{A}^1$. We now want to extend this notion of morphisms to maps to other affine varieties than just $\mathbb{A}^1$ (and in fact also to maps between more general varieties in Chapter 5). It turns out that there is a very natural way to define these morphisms once you know what the regular functions are on the source and target variety. So let us start by attaching the data of the regular functions to the structure of an affine variety, or rather more generally of a topological space.

**Definition 4.1** (Ringed spaces)**.**

(a) A **ringed space** is a topological space $X$ together with a sheaf of rings on $X$. In this situation the given sheaf will always be denoted $\mathscr{O}_X$ and called the **structure sheaf** of the ringed space. Usually we will write this ringed space simply as $X$, with the structure sheaf $\mathscr{O}_X$ being understood.

(b) An affine variety will always be considered as a ringed space together with its sheaf of regular functions as the structure sheaf.

(c) An open subset $U$ of a ringed space $X$ (e. g. of an affine variety) will always be considered as a ringed space with the structure sheaf being the restriction $\mathscr{O}_X|_U$ as in Definition 3.18.

With this idea that the regular functions make up the structure of an affine variety the obvious idea to define a morphism $f : X \to Y$ between affine varieties (or more generally ringed spaces) is now that they should preserve this structure in the sense that for any regular function $\varphi : U \to K$ on an open subset $U$ of $Y$ the composition $\varphi \circ f : f^{-1}(U) \to K$ is again a regular function.

However, there is a slight technical problem with this approach. Whereas there is no doubt about what the composition $\varphi \circ f$ above should mean for a regular function $\varphi$ on an affine variety, this notion is a priori undefined for general ringed spaces: recall that in this case by Definition 3.16 the structure sheaves $\mathscr{O}_X$ and $\mathscr{O}_Y$ are given by the data of arbitrary rings $\mathscr{O}_X(U)$ and $\mathscr{O}_Y(V)$ for open subsets $U \subset X$ and $V \subset Y$. So although we usually think of the elements of $\mathscr{O}_X(U)$ and $\mathscr{O}_Y(V)$ as functions on $U$ resp. $V$ there is nothing in the definition that guarantees us such an interpretation, and consequently there is no well-defined notion of composing these sections of the structure sheaves with the map $f : X \to Y$. So in order to be able to proceed without too much technicalities let us assume from now on that all our sheaves are in fact sheaves of functions with some properties:

**Convention 4.2** (Sheaves $=$ sheaves of $K$-valued functions)**.** For every sheaf $\mathscr{F}$ on a topological space $X$ we will assume from now on that the rings $\mathscr{F}(U)$ for open subsets $U \subset X$ are subrings of the rings of all functions from $U$ to $K$ (with the usual pointwise addition and multiplication) containing all constant functions, and that the restriction maps are the ordinary restrictions of such functions. In particular, this makes every sheaf also into a sheaf of $K$-algebras, with the scalar multiplication by elements of $K$ again given by pointwise multiplication. So in short we can say:

$$\boxed{\text{From now on, every sheaf is assumed to be a sheaf of } K\text{-valued functions.}}$$

With this convention we can now go ahead and define morphisms between ringed spaces as motivated above.

**Definition 4.3** (Morphisms of ringed spaces)**.** Let $f : X \to Y$ be a map of ringed spaces.

(a) For any map $\varphi : U \to K$ from an open subset $U$ of $Y$ to $K$ we denote the composition $\varphi \circ f : f^{-1}(U) \to K$ (which is well-defined by Convention 4.2) by $f^*\varphi$. It is called the **pull-back** of $\varphi$ by $f$.

(b) The map $f$ is called a **morphism** (of ringed spaces) if it is continuous, and if for all open subsets $U \subset Y$ and $\varphi \in \mathscr{O}_Y(U)$ we have $f^*\varphi \in \mathscr{O}_X(f^{-1}(U))$. So in this case pulling back by $f$ yields $K$-algebra homomorphisms

$$f^* : \mathscr{O}_Y(U) \to \mathscr{O}_X(f^{-1}(U)), \;\; \varphi \mapsto f^*\varphi.$$

(c) We say that $f$ is an **isomorphism** (of ringed spaces) if it has a two-sided inverse, i. e. if it is bijective, and both $f : X \to Y$ and $f^{-1} : Y \to X$ are morphisms.

Morphisms and isomorphisms of (open subsets of) affine varieties are morphisms (resp. isomorphisms) as ringed spaces.

**Remark 4.4.**

(a) The requirement of $f$ being continuous is necessary in Definition 4.3 (b) to formulate the second condition: it ensures that $f^{-1}(U)$ is open in $X$ if $U$ is open in $Y$, i. e. that $\mathscr{O}_X(f^{-1}(U))$ is well-defined.

(b) Without our Convention 4.2, i. e. for ringed spaces without a natural notion of a pull-back of elements of $\mathscr{O}_Y(U)$, one would actually have to include suitable ring homomorphisms $\mathscr{O}_Y(U) \to \mathscr{O}_X(f^{-1}(U))$ in the data needed to specify a morphism. In other words, in this case a morphism is no longer just a set-theoretic map satisfying certain properties. Although this would be the “correct” notion of morphisms of arbitrary ringed spaces, we will not do this here as it would clearly make our discussion of morphisms more complicated than necessary for our purposes.

**Remark 4.5** (Properties of morphisms)**.** The following two properties of morphisms are obvious from the definition:

(a) Compositions of morphisms are morphisms: if $f : X \to Y$ and $g : Y \to Z$ are morphisms of ringed spaces then so is $g \circ f : X \to Z$.

(b) Restrictions of morphisms are morphisms: if $f : X \to Y$ is a morphism of ringed spaces and $U \subset X$ and $V \subset Y$ are open subsets such that $f(U) \subset V$ then the restricted map $f|_U : U \to V$ is again a morphism of ringed spaces.

Conversely, morphisms satisfy a “gluing property” similar to that of a sheaf in Definition 3.16:

**Lemma 4.6** (Gluing property for morphisms)**.** *Let $f : X \to Y$ be a map of ringed spaces. Assume that there is an open cover $\{U_i : i \in I\}$ of $X$ such that all restrictions $f|_{U_i} : U_i \to Y$ are morphisms. Then $f$ is a morphism.*

*Proof.* By Definition 4.3 (b) we have to check two things:

(a) The map $f$ is continuous: Let $V \subset Y$ be an open subset. Then

$$f^{-1}(V) = \bigcup_{i \in I} (U_i \cap f^{-1}(V)) = \bigcup_{i \in I} (f|_{U_i})^{-1}(V).$$

But as all restrictions $f|_{U_i}$ are continuous the sets $(f|_{U_i})^{-1}(V)$ are open in $U_i$, and hence open in $X$. So $f^{-1}(V)$ is open in $X$, which means that $f$ is continuous.

Of course, this is just the well-known topological statement that continuity is a local property.

(b) The map $f$ pulls back sections of $\mathscr{O}_Y$ to sections of $\mathscr{O}_X$: Let $V \subset Y$ be an open subset and $\varphi \in \mathscr{O}_Y(V)$. Then $(f^*\varphi)|_{U_i \cap f^{-1}(V)} = (f|_{U_i \cap f^{-1}(V)})^*\varphi \in \mathscr{O}_X(U_i \cap f^{-1}(V))$ since $f|_{U_i}$ (and thus also $f|_{U_i \cap f^{-1}(V)}$ by Remark 4.5 (b)) is a morphism. By the gluing property for sheaves in Definition 3.16 this means that $f^*\varphi \in \mathscr{O}_X(f^{-1}(V))$. □

Let us now apply our definition of morphisms to (open subsets of) affine varieties. The following proposition can be viewed as a confirmation that our constructions above were reasonable: as one would certainly expect, a morphism to an affine variety $Y \subset \mathbb{A}^n$ is simply given by an $n$-tuple of regular functions whose image lies in $Y$.

**Proposition 4.7** (Morphisms between affine varieties)**.** *Let $U$ be an open subset of an affine variety $X$, and let $Y \subset \mathbb{A}^n$ be another affine variety. Then the morphisms $f : U \to Y$ are exactly the maps of the form*

$$f = (\varphi_1,\dots,\varphi_n) : U \to Y, \; x \mapsto (\varphi_1(x),\dots,\varphi_n(x))$$

*with $\varphi_i \in \mathscr{O}_X(U)$ for all $i = 1,\dots,n$.*

*In particular, the morphisms from $U$ to $\mathbb{A}^1$ are exactly the regular functions in $\mathscr{O}_X(U)$.*

*Proof.* First assume that $f : U \to Y$ is a morphism. For $i = 1,\dots,n$ the $i$-th coordinate function $y_i$ on $Y \subset \mathbb{A}^n$ is clearly regular on $Y$, and so $\varphi_i := f^* y_i \in \mathscr{O}_X(f^{-1}(Y)) = \mathscr{O}_X(U)$ by Definition 4.3 (b). But this is just the $i$-th component function of $f$, and so we have $f = (\varphi_1,\dots,\varphi_n)$.

Conversely, let now $f = (\varphi_1,\dots,\varphi_n)$ with $\varphi_1,\dots,\varphi_n \in \mathscr{O}_X(U)$ and $f(U) \subset Y$. First of all $f$ is continuous: let $Z$ be any closed subset of $Y$. Then $Z$ is of the form $V(g_1,\dots,g_m)$ for some $g_1,\dots,g_m \in A(Y)$, and

$$f^{-1}(Z) = \{x \in U : g_i(\varphi_1(x),\dots,\varphi_n(x)) = 0 \text{ for all } i = 1,\dots,m\}.$$

But the functions $x \mapsto g_i(\varphi_1(x),\dots,\varphi_n(x))$ are regular on $U$ since plugging in quotients of polynomial functions for the variables of a polynomial gives again a quotient of polynomial functions. Hence $f^{-1}(Z)$ is closed in $U$ by Lemma 3.6, and thus $f$ is continuous. Similarly, if $\varphi \in \mathscr{O}_Y(W)$ is a regular function on some open subset $W \subset Y$ then

$$f^*\varphi = \varphi \circ f : f^{-1}(W) \to K, \; x \mapsto \varphi(\varphi_1(x),\dots,\varphi_n(x))$$

is regular again, since if we replace the variables in a quotient of polynomial functions by other quotients of polynomial functions we obtain again a quotient of polynomial functions. Hence $f$ is a morphism. $\square$

For affine varieties themselves (rather than open subsets) we obtain as a consequence the following useful corollary that translates our geometric notion of morphisms entirely into the language of commutative algebra.

**Corollary 4.8.** *For any two affine varieties $X$ and $Y$ there is a one-to-one correspondence*

$$\begin{array}{ccc} \{\text{morphisms } X \to Y\} & \longleftrightarrow & \{K\text{-algebra homomorphisms } A(Y) \to A(X)\} \\ f & \longmapsto & f^*. \end{array}$$

*In particular, isomorphisms of affine varieties correspond exactly to $K$-algebra isomorphisms in this way.*

*Proof.* By Definition 4.3 any morphism $f : X \to Y$ determines a $K$-algebra homomorphism $f^* : \mathscr{O}_Y(Y) \to \mathscr{O}_X(X)$, i. e. $f^* : A(Y) \to A(X)$ by Proposition 3.10.

Conversely, let $g : A(Y) \to A(X)$ be a $K$-algebra homomorphism. Assume that $Y \subset \mathbb{A}^n$ and denote by $y_1,\dots,y_n$ the coordinate functions of $\mathbb{A}^n$. Then $\varphi_i := g(y_i) \in A(X) = \mathscr{O}_X(X)$ for all $i = 1,\dots,n$. If we set $f = (\varphi_1,\dots,\varphi_n) : X \to \mathbb{A}^n$ then $f(X) \subset Y = V(I(Y))$ since all polynomials $h \in I(Y)$ represent the zero element in $A(Y)$, and hence $h \circ f = g(h) = 0$. Hence $f : X \to Y$ is a morphism by Proposition 4.7. It has been constructed so that $f^* = g$, and so we get the one-to-one correspondence as stated in the corollary.

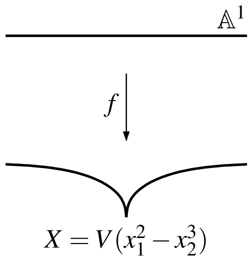
The additional statement about isomorphisms now follows immediately since $(f \circ g)^* = g^* \circ f^*$ and $(g \circ f)^* = f^* \circ g^*$ for all $f : X \to Y$ and $g : Y \to X$. $\square$

06

**Example 4.9** (Isomorphisms $\neq$ bijective morphisms)**.** Let $X = V(x_1^2 - x_2^3) \subset \mathbb{A}^2$ be the curve as in the picture below on the right. It has a "singular point" at the origin (a notion that we will introduce in Definition 10.7 (a)) where it does not look like the graph of a differentiable function.

Now consider the map

$$f : \mathbb{A}^1 \to X,\ t \mapsto (t^3, t^2)$$




which is a morphism by Proposition 4.7. Its corresponding $K$-algebra homomorphism $f^* : A(X) \to A(\mathbb{A}^1)$ as in Corollary 4.8 is given by

$$\begin{aligned} K[x_1,x_2]/(x_1^2 - x_2^3) &\to K[t] \\ \overline{x_1} &\mapsto t^3 \\ \overline{x_2} &\mapsto t^2 \end{aligned}$$

which can be seen by composing $f$ with the two coordinate functions of $\mathbb{A}^2$.

Note that $f$ is bijective with inverse map

$$f^{-1} : X \to \mathbb{A}^1,\ (x_1,x_2) \mapsto \begin{cases} \frac{x_1}{x_2} & \text{if } x_2 \neq 0 \\ 0 & \text{if } x_2 = 0. \end{cases}$$

But $f$ is not an isomorphism (i. e. $f^{-1}$ is not a morphism), since otherwise by Corollary 4.8 the map $f^*$ above would have to be an isomorphism as well — which is false since the linear polynomial $t$ is clearly not in its image. So we have to be careful not to confuse isomorphisms with bijective morphisms.

Another consequence of Proposition 4.7 concerns our definition of the product $X \times Y$ of two affine varieties $X$ and $Y$ in Example 1.4 (d). Recall from Example 2.5 (c) that $X \times Y$ does not carry the product topology — which might seem strange at first. The following proposition however justifies this choice, since it shows that our definition of the product satisfies the so-called *universal property* that giving a morphism to $X \times Y$ is the same as giving a morphism each to $X$ and $Y$. In fact, when we introduce more general varieties in the next chapter we will *define* their products using this certainly desirable universal property.

**Proposition 4.10** (Universal property of products)**.** *Let $X$ and $Y$ be affine varieties, and let $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ be the projection morphisms from the product onto the two factors. Then for every affine variety $Z$ and two morphisms $f_X : Z \to X$ and $f_Y : Z \to Y$ there is a unique morphism $f : Z \to X \times Y$ such that $f_X = \pi_X \circ f$ and $f_Y = \pi_Y \circ f$.*

*In other words, giving a morphism to the product $X \times Y$ is the same as giving a morphism to each of the factors $X$ and $Y$.*

$$\begin{tikzcd} Z \arrow[dr, dashed, "f"] \arrow[drr, bend left, "f_X"] \arrow[ddr, bend right, "f_Y"'] & & \\ & X \times Y \arrow[r, "\pi_X"'] \arrow[d, "\pi_Y"] & X \\ & Y & \end{tikzcd}$$

*Proof.* Obviously, the only way to obtain the relations $f_X = \pi_X \circ f$ and $f_Y = \pi_Y \circ f$ is to take the map $f : Z \to X \times Y$, $z \mapsto (f_X(z), f_Y(z))$. But this is clearly a morphism by Proposition 4.7: as $f_X$ and $f_Y$ must be given by regular functions in each coordinate, the same is then true for $f$. $\square$

**Remark 4.11.** If you know commutative algebra you will have noticed that the universal property of the product in Proposition 4.10 corresponds exactly to the universal property of tensor products using the translation between morphisms of affine varieties and $K$-algebra homomorphisms of Corollary 4.8. Hence the the coordinate ring $A(X \times Y)$ of the product is just the tensor product $A(X) \otimes_K A(Y)$.

**Exercise 4.12.** An *affine conic* is the zero locus in $\mathbb{A}^2$ of a single irreducible polynomial in $K[x_1,x_2]$ of degree 2. Show that every affine conic over a field of characteristic not equal to 2 is isomorphic to exactly one of the varieties $X_1 = V(x_2 - x_1^2)$ and $X_2 = V(x_1x_2 - 1)$, with an isomorphism given by a linear coordinate transformation followed by a translation.

**Exercise 4.13.** Let $X \subset \mathbb{A}^2$ be the zero locus of a single polynomial $\sum_{i+j\le d} a_{i,j} x_1^i x_2^j$ of degree at most $d$. Show that:

(a) Any line in $\mathbb{A}^2$ (i.e. any zero locus of a single polynomial of degree 1) not contained in $X$ intersects $X$ in at most $d$ points.

(b) Any affine conic (as in Exercise 4.12 over a field with $\operatorname{char} K \neq 2$) not contained in $X$ intersects $X$ in at most $2d$ points.

(This is a (very) special case of Bézout's theorem that we will prove in Chapter 12.)

**Exercise 4.14.** Let $f : X \to Y$ be a morphism of affine varieties and $f^* : A(Y) \to A(X)$ the corresponding homomorphism of the coordinate rings. Are the following statements true or false?

(a) $f$ is surjective if and only if $f^*$ is injective.

(b) $f$ is injective if and only if $f^*$ is surjective.

(c) If $f : \mathbb{A}^1 \to \mathbb{A}^1$ is an isomorphism then $f$ is affine linear, i. e. of the form $f(x) = ax+b$ for some $a, b \in K$.

(d) If $f : \mathbb{A}^2 \to \mathbb{A}^2$ is an isomorphism then $f$ is affine linear, i. e. it is of the form $f(x) = Ax+b$ for some $A \in \operatorname{Mat}(2\times 2, K)$ and $b \in K^2$.

**Construction 4.15** (Affine varieties from finitely generated $K$-algebras)**.** Corollary 4.8 allows us to construct affine varieties up to isomorphisms from finitely generated $K$-algebras: if $R$ is such an algebra we can pick generators $a_1, \dots, a_n$ for $R$ and obtain a surjective $K$-algebra homomorphism

$$g : K[x_1, \dots, x_n] \to R, \;\; f \mapsto f(a_1, \dots, a_n).$$

By the homomorphism theorem we therefore see that $R \cong K[x_1, \dots, x_n]/I$, where $I$ is the kernel of $g$. If we assume that $I$ is radical (which is the same as saying that $R$ does not have any nilpotent elements except 0) then $X = V(I)$ is an affine variety in $\mathbb{A}^n$ with coordinate ring $A(X) \cong R$.

Note that this construction of $X$ from $R$ depends on the choice of generators of $R$, and so we can get different affine varieties that way. However, Corollary 4.8 implies that all these affine varieties will be isomorphic since they have isomorphic coordinate rings — they just differ in their embeddings in affine spaces.

This motivates us to make a (very minor) redefinition of the term "affine variety" to allow for objects that are isomorphic to an affine variety in the old sense, but that do not come with an intrinsic description as the zero locus of some polynomials in a fixed affine space.

**Definition 4.16** (Slight redefinition of affine varieties)**.** From now on, an **affine variety** will be a ringed space that is isomorphic to an affine variety in the old sense of Definition 1.2 (c).

Note that all our concepts and results immediately carry over to an affine variety $X$ in this new sense: for example, all topological concepts are defined as $X$ is still a topological space, regular functions are just sections of the structure sheaf $\mathscr{O}_X$, the coordinate ring $A(X)$ can be considered to be $\mathscr{O}_X(X)$ by Proposition 3.10, and products involving $X$ can be defined using any embedding of $X$ in affine space (yielding a product that is unique up to isomorphisms).

Probably the most important examples of affine varieties in this new sense that do not look like affine varieties a priori are our distinguished open subsets of Definition 3.8:

**Proposition 4.17** (Distinguished open subsets are affine varieties)**.** *Let $X$ be an affine variety, and let $f \in A(X)$. Then the distinguished open subset $D(f)$ is an affine variety; its coordinate ring $A(D(f))$ is the localization $A(X)_f$.*

*Proof.* Clearly,

$$Y := \{(x,t) \in X \times \mathbb{A}^1 : t\, f(x) = 1\} \quad \subset X \times \mathbb{A}^1$$

is an affine variety as it is the zero locus of the polynomial $tf(x)-1$ in the affine variety $X \times \mathbb{A}^1$. It is isomorphic to $D(f)$ by the map

$$f: Y \to D(f),\ (x,t) \mapsto x \qquad \text{with inverse} \qquad f^{-1}: D(f) \to Y,\ x \mapsto \left(x, \frac{1}{f(x)}\right).$$

So $D(f)$ is an affine variety, and by Proposition 3.10 and Lemma 3.13 we see that its coordinate ring is $A(D(f)) = \mathscr{O}_X(D(f)) = A(X)_f$. □

**Example 4.18** ($\mathbb{A}^2 \backslash \{0\}$ is not an affine variety)**.** As in Example 3.14 let $X = \mathbb{A}^2$ and consider the open subset $U = \mathbb{A}^2 \backslash \{0\}$ of $X$. Then even in the new sense of Definition 4.16 the ringed space $U$ is not an affine variety: otherwise its coordinate ring would be $\mathscr{O}_X(U)$ by Proposition 3.10, and thus just the polynomial ring $K[x,y]$ by Example 3.14. But this is the same as the coordinate ring of $X = \mathbb{A}^2$, and hence Corollary 4.8 would imply that $U$ and $X$ are isomorphic, with the isomorphism given by the identity map. This is obviously not true, and hence we conclude that $U$ is not an affine variety.

However, we can cover $U$ by the two (distinguished) open subsets

$$D(x_1) = \{(x_1,x_2) : x_1 \neq 0\} \quad \text{and} \quad D(x_2) = \{(x_1,x_2) : x_2 \neq 0\}$$

which are affine by Proposition 4.17. This leads us to the idea that we should also consider ringed spaces that can be patched together from affine varieties. We will do this in the next chapter.

**Exercise 4.19.** Which of the following ringed spaces are isomorphic over $\mathbb{C}$?

(a) $\mathbb{A}^1 \backslash \{1\}$
(b) $V(x_1^2 + x_2^2) \subset \mathbb{A}^2$
(c) $V(x_2 - x_1^2, x_3 - x_1^3) \backslash \{0\} \subset \mathbb{A}^3$
(d) $V(x_1 x_2) \subset \mathbb{A}^2$
(e) $V(x_2^2 - x_1^3 - x_1^2) \subset \mathbb{A}^2$
(f) $V(x_1^2 - x_2^2 - 1) \subset \mathbb{A}^2$

# 5. Varieties

In this chapter we will now finally introduce the main objects of study in this class, the so-called *varieties*. As already announced in Example 4.18 they will be spaces that are not necessarily affine varieties themselves, but that can be covered by affine varieties. This idea is completely analogous to the definition of manifolds: recall that to construct them one first considers open subsets of $\mathbb{R}^n$ that are supposed to form the patches of your space, and then defines a manifold to be a topological space that looks locally like these patches. In our algebraic case we can say that the affine varieties form the basic patches of the spaces that we want to consider, and that general varieties are then spaces that look locally like affine varieties.

One of the main motivations for this generalization is that in the classical topology affine varieties over $\mathbb{C}$ are never bounded, and hence never compact, unless they are a finite set (see Exercise 2.34 (c)). As compact spaces are usually better-behaved than non-compact ones, it is therefore desirable to have a method to compactify an affine variety by "adding some points at infinity". Technically, this can be achieved by gluing it to other affine varieties that contain the points at infinity. The complete space can then obviously be covered by affine varieties. We will see this explicitly in Examples 5.5 (a) and 5.6, and much more generally when we construct projective varieties in Chapters 6 and 7.

So let us start by defining spaces that can be covered by affine varieties. They are called *prevarieties* since we will want to impose another condition on them later in Definition 5.19, which will then make them into varieties.

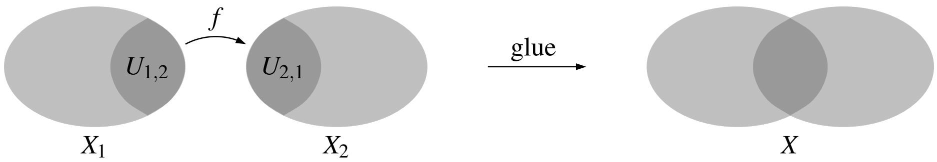
**Definition 5.1** (Prevarieties)**.** A **prevariety** is a ringed space $X$ that has a finite open cover by affine varieties. Morphisms of prevarieties are simply morphisms as ringed spaces. In accordance with Definition 3.1, the elements of $\mathscr{O}_X(U)$ for an open subset $U \subset X$ will be called **regular functions** on $U$.

**Remark 5.2.** Note that the open cover in Definition 5.1 is not part of the data needed to specify a prevariety — it is just required that such a cover exists. Any open subset of a prevariety that is an affine variety is called an *affine open set*.

**Example 5.3.** Of course, any affine variety is a prevariety. More generally, every open subset of an affine variety is a prevariety: it has a finite open cover by distinguished open subsets by Remark 3.9 (b), and these are affine open sets by Proposition 4.17.

The basic way to construct new prevarieties is to glue them together from previously known patches. For simplicity, let us start with the case when we only have two spaces to glue.

**Construction 5.4** (Gluing two prevarieties)**.** Let $X_1, X_2$ be two prevarieties (e. g. affine varieties), and let $U_{1,2} \subset X_1$ and $U_{2,1} \subset X_2$ be non-empty open subsets. Moreover, let $f : U_{1,2} \to U_{2,1}$ be an isomorphism. Then we can define a prevariety $X$ obtained by *gluing* $X_1$ and $X_2$ along $f$, as shown in the picture below:




- As a set, the space $X$ is just the disjoint union $X_1 \cup X_2$ modulo the equivalence relation given by $a \sim f(a)$ and $f(a) \sim a$ for all $a \in U_{1,2}$ (in addition to $a \sim a$ for all $a \in X_1 \cup X_2$). Note that there are then natural embeddings $i_1 : X_1 \to X$ and $i_2 : X_2 \to X$ that map a point to its equivalence class in $X_1 \cup X_2$.

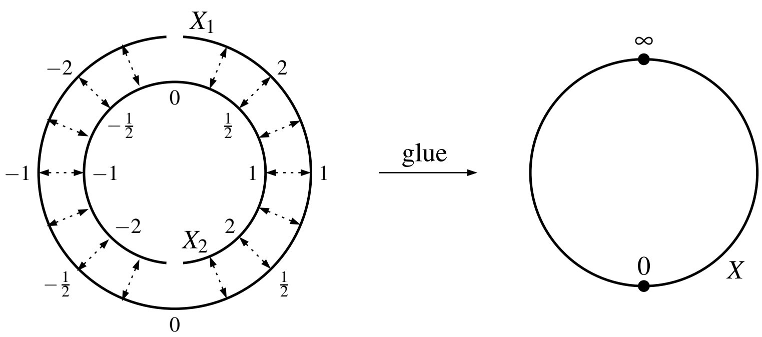
- As a topological space, we call a subset $U \subset X$ open if $i_1^{-1}(U) \subset X_1$ and $i_2^{-1}(U) \subset X_2$ are open. In topology, this is usually called the *quotient topology* of the two maps $i_1$ and $i_2$.
- As a ringed space, we define the structure sheaf $\mathscr{O}_X$ by
$$\mathscr{O}_X(U) = \{\varphi : U \to K : \; i_1^*\varphi \in \mathscr{O}_{X_1}(i_1^{-1}(U)) \text{ and } i_2^*\varphi \in \mathscr{O}_{X_2}(i_2^{-1}(U))\}$$
for all open subsets $U \subset X$, where $i_1^*$ and $i_2^*$ denote the pull-backs of Definition 4.3 (a). Intuitively, this means that a function on the glued space is regular if it is regular when restricted to both patches. It is obvious that this defines a sheaf on $X$.

With this construction it is checked immediately that the images of $i_1$ and $i_2$ are open subsets of $X$ that are isomorphic to $X_1$ and $X_2$. We will often drop the inclusion maps from the notation and say that $X_1$ and $X_2$ are open subsets of $X$. Since $X_1$ and $X_2$ can be covered by affine open subsets, the same is true for $X$, and thus $X$ is a prevariety.

**Example 5.5.** As the simplest example of the above gluing construction, let $X_1 = X_2 = \mathbb{A}^1$ and $U_{1,2} = U_{2,1} = \mathbb{A}^1 \backslash \{0\}$ in the notation of Construction 5.4. We consider two different choices of gluing isomorphism $f : U_{1,2} \to U_{2,1}$:

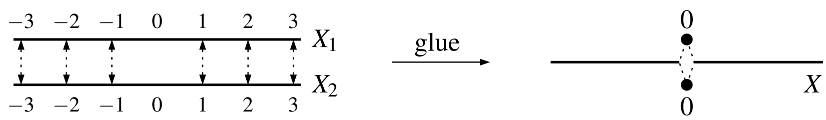
(a) Let $f : U_{1,2} \to U_{2,1}$, $x \mapsto \frac{1}{x}$. By construction, the affine line $X_1 = \mathbb{A}^1$ is an open subset of $X$. Its complement $X \backslash X_1 = X_2 \backslash U_{2,1}$ is a single point that corresponds to 0 in $X_2$ and therefore to "$\infty = \frac{1}{0}$" in the coordinate of $X_1$. Hence we can think of the glued space $X$ as $\mathbb{A}^1 \cup \{\infty\}$, and thus as a "compactification" of the affine line. We denote it by $\mathbb{P}^1$; it is a special case of the projective spaces that we will introduce in Chapter 6 (see Exercise 7.3 (a)).

In the case $K = \mathbb{C}$ the resulting space $X$ is just the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Its subset of real points is shown in the picture below (with the dotted arrows indicating the gluing isomorphism), it is homeomorphic to a circle.




As an example of gluing morphisms as in Lemma 4.6, the morphisms
$$X_1 \to X_2 \subset \mathbb{P}^1, \; x \mapsto x \quad \text{and} \quad X_2 \to X_1 \subset \mathbb{P}^1, \; x \mapsto x$$
(that correspond to a reflection across the horizontal axis in the picture above) glue together to a single morphism $\mathbb{P}^1 \to \mathbb{P}^1$ that can be thought of as $x \mapsto \frac{1}{x}$ in the interpretation of $\mathbb{P}^1$ as $\mathbb{A}^1 \cup \{\infty\}$.

(b) Let $f : U_{1,2} \to U_{2,1}$ be the identity map. Then the space $X$ obtained by gluing $X_1$ and $X_2$ along $f$ is shown in the picture below, it is "an affine line with two zero points".




Obviously this is a somewhat weird space. Speaking in analytic terms in the case $K = \mathbb{C}$, a sequence of points tending to zero would have two possible limits in $X$, namely the two zero points. Also, as in (a) the two morphisms
$$X_1 \to X_2 \subset X, \; x \mapsto x \quad \text{and} \quad X_2 \to X_1 \subset X, \; x \mapsto x$$

glue again to a morphism $g : X \to X$; this time it exchanges the two zero points and thus leads to the set $\{x \in X : g(x) = x\} = \mathbb{A}^1 \backslash \{0\}$ not being closed in $X$, although it is given by an equality of continuous maps.

Usually we want to exclude such spaces from the objects we consider. In Definition 5.19 we will therefore impose an additional condition on our prevarieties that ensures that the above phenomena do not occur (see e. g. Proposition 5.21 (b)).

**Example 5.6.** Consider again the complex affine curve

$$X = \{(x_1,x_2) \in \mathbb{A}^2_{\mathbb{C}} : x_2^2 = (x_1-1)(x_1-2)\cdots(x_1-2n)\}$$

of Example 0.1. We have already seen in the introduction that $X$ can (and should) be "compactified" by adding two points at infinity, corresponding to the limit $x_1 \to \infty$ and the two possible values for $x_2$. Let us now construct this compactified space rigorously as a prevariety.

To be able to add a limit point "$x_1 = \infty$" to our space, let us make a coordinate change $\tilde{x}_1 = \frac{1}{x_1}$ (where $x_1 \neq 0$), so that the equation of the curve becomes

$$x_2^2 \tilde{x}_1^{2n} = (1-\tilde{x}_1)(1-2\tilde{x}_1)\cdots(1-2n\tilde{x}_1).$$

If we make an additional coordinate change $\tilde{x}_2 = x_2 \tilde{x}_1^n$, this becomes

$$\tilde{x}_2^2 = (1-\tilde{x}_1)(1-2\tilde{x}_1)\cdots(1-2n\tilde{x}_1).$$

In these coordinates we can now add our two points at infinity, as they correspond to $\tilde{x}_1 = 0$ (and therefore $\tilde{x}_2 = \pm 1$).

Hence the "compactified curve" of Example 0.1 can be constructed as the prevariety obtained by gluing the two affine varieties

$$\begin{aligned} X_1 &= \{(x_1,x_2) \in \mathbb{A}^2_{\mathbb{C}} : x_2^2 = (x_1-1)(x_1-2)\cdots(x_1-2n)\} = X \\ \text{and} \quad X_2 &= \{(\tilde{x}_1,\tilde{x}_2) \in \mathbb{A}^2_{\mathbb{C}} : \tilde{x}_2^2 = (1-\tilde{x}_1)(1-2\tilde{x}_1)\cdots(1-2n\tilde{x}_1)\} \end{aligned}$$

along the isomorphism

$$f : U_{1,2} \to U_{2,1}, \ \ (x_1,x_2) \mapsto (\tilde{x}_1,\tilde{x}_2) = \left(\frac{1}{x_1}, \frac{x_2}{x_1^n}\right)$$

with inverse

$$f^{-1} : U_{2,1} \to U_{1,2}, \ \ (\tilde{x}_1,\tilde{x}_2) \mapsto (x_1,x_2) = \left(\frac{1}{\tilde{x}_1}, \frac{\tilde{x}_2}{\tilde{x}_1^n}\right),$$

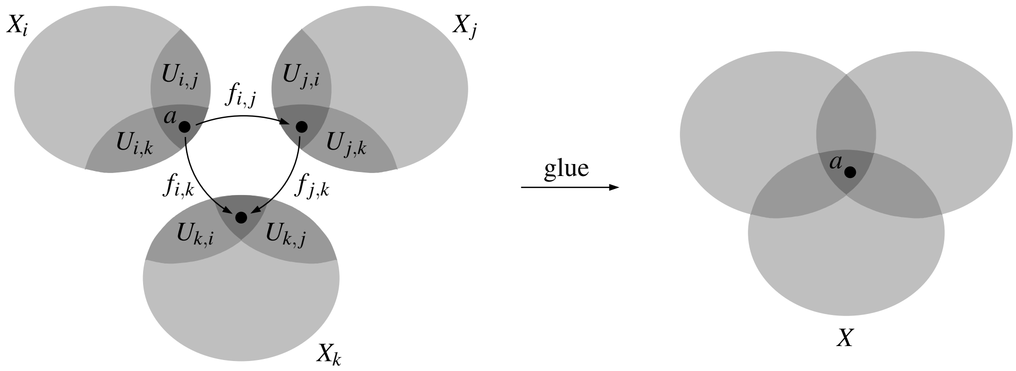
where $U_{1,2} = \{(x_1,x_2) : x_1 \neq 0\} \subset X_1$ and $U_{2,1} = \{(\tilde{x}_1,\tilde{x}_2) : \tilde{x}_1 \neq 0\} \subset X_2$.

Let us now turn to the general construction to glue more than two spaces together. In principle this works in the same way as in Construction 5.4; we just have an additional technical compatibility condition on the gluing isomorphisms.

**Construction 5.7** (General gluing construction)**.** For a finite index set $I$ let $X_i$ be a prevariety for all $i \in I$. Moreover, as in the picture below suppose that for all $i, j \in I$ with $i \neq j$ we are given open subsets $U_{i,j} \subset X_i$ and isomorphisms $f_{i,j} : U_{i,j} \to U_{j,i}$ such that for all distinct $i, j, k \in I$ we have

(a) $f_{j,i} = f_{i,j}^{-1}$;

(b) $U_{i,j} \cap f_{i,j}^{-1}(U_{j,k}) \subset U_{i,k}$, and $f_{j,k} \circ f_{i,j} = f_{i,k}$ on $U_{i,j} \cap f_{i,j}^{-1}(U_{j,k})$.




In analogy to Construction 5.4 we can then define a set $X$ by taking the disjoint union of all $X_i$ for $i \in I$, modulo the equivalence relation $a \sim f_{i,j}(a)$ for all $a \in U_{i,j} \subset X_i$ (in addition to $a \sim a$ for all $a$). In fact, the conditions (a) and (b) above ensure precisely that this relation is symmetric and transitive, respectively. It is obvious that we can now make $X$ into a prevariety by defining its topology and structure sheaf in the same way as in Construction 5.4. We call it the prevariety obtained by *gluing* the $X_i$ along the isomorphisms $f_{i,j}$.

**Exercise 5.8.** Show:

(a) Every morphism $f : \mathbb{A}^1 \backslash \{0\} \to \mathbb{P}^1$ can be extended to a morphism $\mathbb{A}^1 \to \mathbb{P}^1$.

(b) Not every morphism $f : \mathbb{A}^2 \backslash \{0\} \to \mathbb{P}^1$ can be extended to a morphism $\mathbb{A}^2 \to \mathbb{P}^1$.

(c) Every morphism $f : \mathbb{P}^1 \to \mathbb{A}^1$ is constant.

**Exercise 5.9.**

(a) Show that every isomorphism $f : \mathbb{P}^1 \to \mathbb{P}^1$ is of the form $f(x) = \frac{ax+b}{cx+d}$ for some $a,b,c,d \in K$, where $x$ is an affine coordinate on $\mathbb{A}^1 \subset \mathbb{P}^1$.

(b) Given three distinct points $a_1,a_2,a_3 \in \mathbb{P}^1$ and three distinct points $b_1,b_2,b_3 \in \mathbb{P}^1$, show that there is a unique isomorphism $f : \mathbb{P}^1 \to \mathbb{P}^1$ such that $f(a_i) = b_i$ for $i = 1,2,3$.

**Exercise 5.10.** If $X$ and $Y$ are affine varieties we have seen in Proposition 3.10 and Corollary 4.8 that there is a one-to-one correspondence

$$\begin{aligned}\{\text{morphisms } X \to Y\} &\longleftrightarrow \{K\text{-algebra homomorphisms } \mathscr{O}_Y(Y) \to \mathscr{O}_X(X)\} \\ f &\longmapsto f^*.\end{aligned}$$

Does this statement still hold

(a) if $X$ is an arbitrary prevariety (but $Y$ is still affine);

(b) if $Y$ is an arbitrary prevariety (but $X$ is still affine)?

We have just seen how we can construct prevarieties by gluing affine varieties. For the rest of the chapter let us now study some of their basic properties. Of course, all topological concepts (like connectedness, irreducibility, and dimension) carry over immediately to the case of prevarieties. The irreducible decomposition of Proposition 2.15 is also applicable since a prevariety is always Noetherian: 07

**Exercise 5.11.** Prove:

(a) Any prevariety is a Noetherian topological space.

(b) If $X = X_1 \cup \cdots \cup X_m$ is the irreducible decomposition of a prevariety $X$, then the local dimension $\operatorname{codim}_X \{a\}$ of $X$ at any point $a \in X$ is

$$\operatorname{codim}_X \{a\} = \max\{\dim X_i : a \in X_i\}.$$

(c) The statement of (a) would be false if we had defined a prevariety to be a ringed space that has an arbitrary (not necessarily finite) open cover by affine varieties.

As for properties of prevarieties involving the structure as ringed spaces, we should first of all figure out to what extent their subsets, images and inverse images under morphisms, and products are again prevarieties.

**Construction 5.12** (Open and closed subprevarieties)**.** Let $X$ be a prevariety.

(a) Let $U \subset X$ be an open subset. Then $U$ is again a prevariety (as usual with the structure sheaf $\mathscr{O}_U = \mathscr{O}_X|_U$ as in Definition 4.1 (c)): as $X$ can be covered by affine varieties, $U$ can be covered by open subsets of affine varieties, which themselves can be covered by affine varieties by Example 5.3.

We call $U$ (with this structure as a prevariety) an **open subprevariety** of $X$.

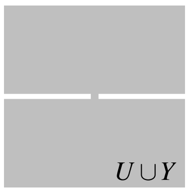
(b) The situation is more complicated for a closed subset $Y \subset X$: as an open subset $U$ of $Y$ is in general not open in $X$ we cannot define a structure sheaf on $Y$ by simply setting $\mathscr{O}_Y(U)$ to be $\mathscr{O}_X(U)$. Instead, we define $\mathscr{O}_Y(U)$ to be the $K$-algebra of functions $U \to K$ that are locally restrictions of functions on $X$, or formally

$$\mathscr{O}_Y(U) := \{\varphi : U \to K : \text{for all } a \in U \text{ there are an open neighborhood } V \text{ of } a \text{ in } X \text{ and } \psi \in \mathscr{O}_X(V) \text{ with } \varphi = \psi \text{ on } U \cap V\}.$$

By the local nature of this definition it is obvious that $\mathscr{O}_Y$ is a sheaf, thus making $Y$ into a ringed space. In fact, we will check in Exercise 5.13 that $Y$ is a prevariety in this way. We call it a **closed subprevariety** of $X$.

(c) If $U$ is an open and $Y$ a closed subset of $X$, then $U \cap Y$ is open in $Y$ and closed in $U$, and thus we can give it the structure of a prevariety by combining (a) with (b) — in fact, one can check that it does not matter whether we consider it to be an open subprevariety of $Y$ or a closed subprevariety of $U$. Intersections of open and closed subprevarieties (with this structure of a ringed space) are called **locally closed subprevarieties**. For example, $\{(x_1,x_2) \in \mathbb{A}^2 : x_1 = 0, x_2 \neq 0\}$ is a locally closed subprevariety of $\mathbb{A}^2$.

The reason why we consider all these seemingly special cases is that for a general subset of $X$ there is no way to make it into a prevariety in a natural way. Even worse, the notions of open and closed subprevarieties do not mix very well: whereas a finite union of open (resp. closed) subprevarieties is of course again an open (resp. closed) subprevariety, the same statement is not true if we try to combine open with closed spaces: in $X = \mathbb{A}^2$ the union of the open subprevariety $U = \mathbb{A}^1 \times (\mathbb{A}^1 \backslash \{0\})$ and the closed subprevariety $Y = \{0\}$ as in the picture on the right does not have a natural structure as a subprevariety of $\mathbb{A}^2$ (since it does not look like an affine variety in a neighborhood of the origin).




**Exercise 5.13.** Let $Y$ be a closed subset of a prevariety $X$, considered as a ringed space with the structure sheaf of Construction 5.12 (b). Prove for every affine open subset $U \subset X$ that the ringed space $U \cap Y$ (considered as an open subset of the ringed space $Y$ as in Definition 4.1 (c)) is isomorphic to the affine variety $U \cap Y$ (considered as an affine subvariety of the affine variety $U$).

In particular, this shows that Construction 5.12 (b) makes $Y$ into a prevariety, and that this prevariety is isomorphic to the affine variety $Y$ if $X$ is itself affine (and thus $Y$ an affine subvariety of $X$).

**Remark 5.14** (Properties of closed subprevarieties)**.** By Construction 5.12 (b), a regular function on (an open subset of) a closed subprevariety $Y$ of a prevariety $X$ is locally the restriction of a regular function on $X$. Hence:

(a) The inclusion map $Y \to X$ is a morphism (it is clearly continuous, and regular functions on $X$ are by construction still regular when restricted to $Y$).

(b) If $f : Z \to X$ is a morphism from a prevariety $Z$ such that $f(Z) \subset Y$ then we can also regard $f$ as a morphism from $Z$ to $Y$ (the pull-back of a regular function on $Y$ by $f$ is locally also a pull-back of a regular function on $X$, and hence regular as $f : Z \to X$ is a morphism).

**Remark 5.15** (Images and inverse images of subprevarieties)**.** Let $f : X \to Y$ be a morphism of prevarieties.

(a) The image of an open or closed subprevariety of $X$ under $f$ is not necessarily an open or closed subprevariety of $Y$. For example, for the affine variety $X = V(x_2x_3 - 1) \cup \{0\} \subset \mathbb{A}^3$ and the projection morphism $f : X \to \mathbb{A}^2$ onto the first two coordinates the image $f(X)$ is exactly the space $\mathbb{A}^1 \times (\mathbb{A}^1\backslash\{0\}) \cup \{0\}$ of Construction 5.12 which is neither an open nor a closed subprevariety of $\mathbb{A}^2$.

As a substitute, one can often consider the graph of $f$ instead of its image, see Proposition 5.21 (a).

(b) By continuity, the inverse image of an open (resp. closed) subprevariety of $Y$ under $f$ is clearly again an open (resp. closed) subprevariety of $X$.

As for the product $X \times Y$ of two prevarieties $X$ and $Y$, the natural idea to construct this space as a prevariety would be to choose finite affine open covers $\{U_i : i \in I\}$ and $\{V_j : j \in J\}$ of $X$ and $Y$, respectively, and then glue the affine product varieties $U_i \times V_j$ using Construction 5.7. If we did this directly however, we would still have to prove that the resulting space does not depend on the chosen affine covers. The best way out of this trouble is to define the product prevariety by a universal property analogous to Proposition 4.10. This will then ensure the uniqueness of the product, so that it suffices to prove its existence by gluing affine patches.

**Definition 5.16** (Products of prevarieties)**.** Let $X$ and $Y$ be prevarieties. A **product** of $X$ and $Y$ is a prevariety $P$ together with morphisms $\pi_X : P \to X$ and $\pi_Y : P \to Y$ satisfying the following *universal property*: for any two morphisms $f_X : Z \to X$ and $f_Y : Z \to Y$ from another prevariety $Z$ there is a unique morphism $f : Z \to P$ such that $\pi_X \circ f = f_X$ and $\pi_Y \circ f = f_Y$.

$$\begin{tikzcd}
Z \arrow[dr, dashed, "f"] \arrow[drr, bend left, "f_X"] \arrow[ddr, bend right, "f_Y"'] & & \\
& P \arrow[r, "\pi_X"'] \arrow[d, "\pi_Y"] & X \\
& Y &
\end{tikzcd}$$

As in the affine case in Proposition 4.10, this means that giving a morphism to the product $P$ is the same as giving a morphism to each of the factors $X$ and $Y$.

**Proposition 5.17** (Existence and uniqueness of products)**.** *Any two prevarieties $X$ and $Y$ have a product. Moreover, this product $P$ with morphisms $\pi_X : P \to X$ and $\pi_Y : P \to Y$ is unique up to unique isomorphism: if $P'$ with $\pi'_X : P' \to X$ and $\pi'_Y : P' \to Y$ is another product there is a unique isomorphism $f : P' \to P$ such that $\pi_X \circ f = \pi'_X$ and $\pi_Y \circ f = \pi'_Y$.*

*We will denote this product simply by $X \times Y$.*

*Proof.* To show existence we glue the affine products of Proposition 4.10 using Construction 5.7. More precisely, let $X$ and $Y$ be covered by affine varieties $U_i$ and $V_j$ for $i \in I$ and $j \in J$, respectively. Use Construction 5.7 to glue the affine products $U_i \times V_j$, where we glue any two such products $U_i \times V_j$ and $U_{i'} \times V_{j'}$ along the identity isomorphism of the common open subset $(U_i \cap U_{i'}) \times (V_j \cap V_{j'})$. Note that these isomorphisms obviously satisfy the conditions (a) and (b) of the construction, and that the resulting glued space $P$ is just the usual product $X \times Y$ as a set. Moreover, using Lemma 4.6 we can glue the affine projection morphisms $U_i \times V_j \to U_i \subset X$ and $U_i \times V_j \to V_j \subset Y$ to morphisms $\pi_X : P \to X$ and $\pi_Y : P \to Y$.

Let us now check the universal property of Definition 5.16 for our construction. If $f_X : Z \to X$ and $f_Y : Z \to Y$ are any two morphisms from a prevariety $Z$, the only way to achieve $\pi_X \circ f = f_X$ and $\pi_Y \circ f = f_Y$ is to define $f : Z \to P$ as $f(z) = (f_X(z), f_Y(z))$, where we identify $P$ set-theoretically with $X \times Y$. By Lemma 4.6, we can check that this is a morphism by restricting it to an affine open cover. If we first cover $Z$ by the open subsets $f_X^{-1}(U_i) \cap f_Y^{-1}(V_j)$ for all $i \in I$ and $j \in J$, and these subsets then by affine open subsets by Construction 5.12 (a), we may assume that every affine open subset in our open cover of $Z$ is mapped to a single (and hence affine) patch $U_i \times V_j$. But after this restriction to the affine case we know by Proposition 4.10 that $f$ is a morphism.

To show uniqueness, assume that $P'$ with $\pi'_X : P' \to X$ and $\pi'_Y : P' \to Y$ is another product. By the universal property of $P$ applied to the morphisms $\pi'_X : P' \to X$ and $\pi'_Y : P' \to Y$, we see that there is a unique morphism $f : P' \to P$ with $\pi_X \circ f = \pi'_X$ and $\pi_Y \circ f = \pi'_Y$. Reversing the roles of the two products, we get in the same way a unique morphism $g : P \to P'$ with $\pi'_X \circ g = \pi_X$ and $\pi'_Y \circ g = \pi_Y$. Finally, apply the universal property of $P$ to the two morphisms $\pi_X : P \to X$ and $\pi_Y : P \to Y$. Since both

$$\begin{aligned} \pi_X \circ (f \circ g) &= \pi'_X \circ g = \pi_X \\ \pi_Y \circ (f \circ g) &= \pi'_Y \circ g = \pi_Y \end{aligned} \quad \text{and} \quad \begin{aligned} \pi_X \circ \mathrm{id}_P &= \pi_X \\ \pi_Y \circ \mathrm{id}_P &= \pi_Y \end{aligned}$$

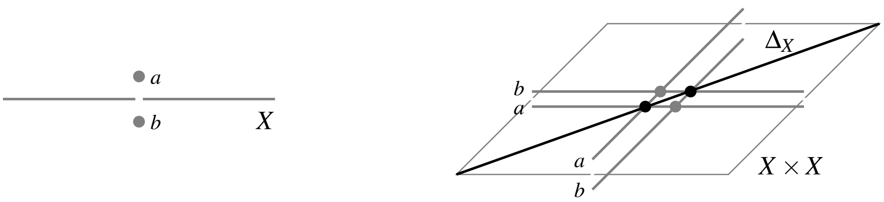
the uniqueness part of the universal property shows that $f \circ g = \mathrm{id}_P$. In the same way we see that $g \circ f = \mathrm{id}_{P'}$, so that $f$ is an isomorphism. □

**Remark 5.18.** Again, a check might be in order that our constructions were consistent: let $X$ and $Y$ be prevarieties, and let $X' \subset X$ and $Y' \subset Y$ be closed subprevarieties. Then we have defined two structures of prevarieties on the set-theoretic product $X' \times Y'$: the closed subprevariety structure of $X' \times Y'$ in $X \times Y$ as in Construction 5.12 (b), and the product prevariety structure of Definition 5.16. As expected, these two structures agree: in fact, by Definition 5.16 together with Remark 5.14 the set-theoretic identity map is a morphism between these two structures in both ways.

Finally, as already announced let us now impose a condition on prevarieties that excludes such unwanted spaces as the affine line with two zero points of Example 5.5 (b). In the theory of manifolds, this is usually done by requiring that the topological space satisfies the so-called *Hausdorff property*, i. e. that every two distinct points have disjoint open neighborhoods. This is obviously not satisfied in our case, since the two zero points do not have such disjoint open neighborhoods.

However, in the Zariski topology the Hausdorff property does not make too much sense, as non-empty open subsets of an irreducible space can never be disjoint by Remark 2.18 (a). So we need a suitable replacement for this condition that captures our geometric idea of the absence of doubled points also in the Zariski topology.

The solution to this problem is inspired by a proposition in general topology stating that the Hausdorff property of a topological space $X$ is equivalent to the condition that the so-called *diagonal* $\Delta_X = \{(x,x) : x \in X\}$ is closed in $X \times X$ (with the product topology). The picture below illustrates this in the case when $X$ is the affine line with two zero points $a$ and $b$: the product $X \times X$ then contains four zero points $(a,a)$, $(a,b)$, $(b,a)$, and $(b,b)$, but by definition only two of them, namely $(a,a)$ and $(b,b)$, are in $\Delta_X$. Hence the diagonal is not closed: the other two zero points are also contained in its closure.




Of course, this equivalence does not really help us directly in algebraic geometry since we do not use the product topology on $X \times X$. But the geometric idea to detect doubled points shown in the picture above on the right is still valid in the Zariski topology — and so we will just use the diagonal condition to *define* the property that we want:

**Definition 5.19** (Varieties)**.**

(a) A prevariety $X$ is called a **variety** (or **separated**) if the **diagonal**

$$\Delta_X := \{(x,x) : x \in X\}$$

is closed in $X \times X$.

(b) Analogously to Definition 2.30 (b), a variety of pure dimension 1 or 2 is called a **curve** resp. **surface**. If $X$ is a pure-dimensional variety and $Y$ a pure-dimensional subvariety of $X$ with $\dim Y = \dim X - 1$ we say that $Y$ is a **hypersurface** in $X$.

So by the argument given above, the affine line with two zero points of Example 5.5 (b) is not a variety. In contrast, the following lemma shows that most prevarieties that we will meet are also varieties. From now on we will almost always assume that our spaces are separated, and thus talk about varieties instead of prevarieties.

**Lemma 5.20.**

(a) *Affine varieties are varieties.*

(b) *Open, closed, and locally closed subprevarieties of varieties are varieties. We will therefore simply call them* ***open***, ***closed***, *and* ***locally closed subvarieties***, *respectively.*

*Proof.*

(a) If $X \subset \mathbb{A}^n$ then $\Delta_X = V(x_1 - y_1, \dots, x_n - y_n) \subset X \times X$, where $x_1, \dots, x_n$ and $y_1, \dots, y_n$ are the coordinates on the two factors, respectively. Hence $\Delta_X$ is closed.

(b) If $Y \subset X$ is an open, closed, or locally closed subvariety, consider the inclusion morphism $i : Y \times Y \to X \times X$ (which exists by the universal property of Definition 5.16). As $\Delta_Y = i^{-1}(\Delta_X)$ and $\Delta_X$ is closed by assumption, $\Delta_Y$ is closed as well by the continuity of $i$. □

For varieties, we have the following additional desirable properties in addition to the ones for prevarieties:

**Proposition 5.21** (Properties of varieties)**.** *Let $f, g : X \to Y$ be morphisms of prevarieties, and assume that $Y$ is a variety.*

(a) *The **graph** $\Gamma_f := \{(x, f(x)) : x \in X\}$ is closed in $X \times Y$.*

(b) *The set $\{x \in X : f(x) = g(x)\}$ is closed in $X$.*

*Proof.*

(a) By the universal property of products of prevarieties as in Definition 5.16 there is a morphism $(f, \mathrm{id}) : X \times Y \to Y \times Y$, $(x, y) \mapsto (f(x), y)$. As $Y$ is a variety we know that $\Delta_Y$ is closed, and hence so is $\Gamma_f = (f, \mathrm{id})^{-1}(\Delta_Y)$ by continuity.

(b) Similarly to (a), the given set is the inverse image of the diagonal $\Delta_Y$ under the morphism $X \to Y \times Y$, $x \mapsto (f(x), g(x))$. Hence it is closed again since $\Delta_Y$ is closed. □

**Exercise 5.22.** Show that the space $\mathbb{P}^1$ of Example 5.5 (a) is a variety.

**Exercise 5.23.** Let $X$ and $Y$ be prevarieties. Show:

(a) If $X$ and $Y$ are varieties then so is $X \times Y$.

(b) If $X$ and $Y$ are irreducible then so is $X \times Y$.

**Exercise 5.24.** Use diagonals to prove the following statements:

(a) The intersection of any two affine open subsets of a variety is again an affine open subset.

(b) If $X, Y \subset \mathbb{A}^n$ are two pure-dimensional affine varieties then every irreducible component of $X \cap Y$ has dimension at least $\dim X + \dim Y - n$.

**Exercise 5.25.** In Exercise 2.33 (b) we have seen that the dimension of a dense open subset $U$ of a topological space $X$ need not be the same as that of $X$.

However, show now that $\dim U = \dim X$ holds in this situation if $X$ is a variety. 08

# 6. Projective Varieties I: Topology

In the last chapter we have studied (pre-)varieties, i. e. topological spaces that are locally isomorphic to affine varieties. In particular, the ability to glue affine varieties together allowed us to construct *compact* spaces (in the classical topology over the ground field $\mathbb{C}$) as e. g. $\mathbb{P}^1$, whereas affine varieties themselves are never compact unless they consist of only finitely many points (see Exercise 2.34 (c)). Unfortunately, the description of a variety in terms of its affine patches and gluing isomorphisms is quite inconvenient in practice, as we have seen already in the calculations in the last chapter. It would therefore be desirable to have a global description of these spaces that does not refer to gluing methods.

Projective varieties form a very large class of "compact" varieties that do admit such a global description. In fact, the class of projective varieties is so large that it is not easy to construct a variety that is *not* (an open subset of) a projective variety — in this class we will certainly not see one.

In this chapter we will construct projective varieties as topological spaces, leaving their structure as ringed spaces to Chapter 7. To do this we first of all need projective spaces, which can be thought of as compactifications of affine spaces. We have already seen $\mathbb{P}^1$ as $\mathbb{A}^1$ together with a "point at infinity" in Example 5.5 (a); other projective spaces are just generalizations of this construction to higher dimensions. As we aim for a global description of these spaces however, their definition looks quite different from the one in Example 5.5 (a) at first.

**Definition 6.1** (Projective spaces)**.** Let $n \in \mathbb{N}$. We define **projective $n$-space** over $K$, denoted $\mathbb{P}^n_K$ or simply $\mathbb{P}^n$, to be the set of all 1-dimensional linear subspaces of the vector space $K^{n+1}$.

**Notation 6.2** (Homogeneous coordinates)**.** Obviously, a 1-dimensional linear subspace of $K^{n+1}$ is uniquely determined by a non-zero vector in $K^{n+1}$, with two such vectors spanning 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

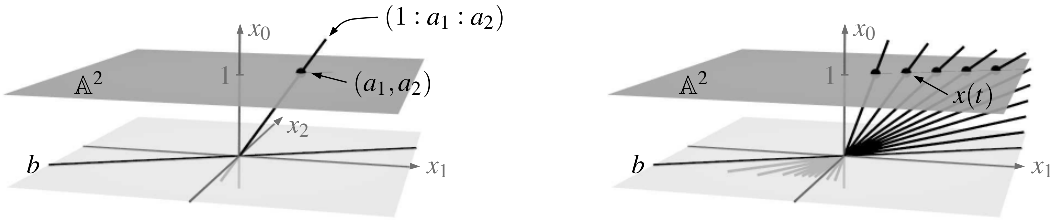
$$(x_0, \dots, x_n) \sim (y_0, \dots, y_n) \quad :\Leftrightarrow \quad x_i = \lambda y_i \text{ for some } \lambda \in K^* \text{ and all } i,$$

where $K^* = K \backslash \{0\}$ is the multiplicative group of units of $K$. This is usually written as $\mathbb{P}^n = (K^{n+1}\backslash\{0\})/K^*$, and the equivalence class of $(x_0, \dots, x_n)$ will be denoted by $(x_0 : \cdots : x_n) \in \mathbb{P}^n$ (the notations $[x_0 : \cdots : x_n]$ and $[x_0, \dots, x_n]$ are also common in the literature). So in the notation $(x_0 : \cdots : x_n)$ 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. They are called the **homogeneous coordinates** of the point (the reason for this name will become obvious in the course of this chapter). Note also that we will usually label the homogeneous coordinates of $\mathbb{P}^n$ by $x_0, \dots, x_n$ instead of by $x_1, \dots, x_{n+1}$. This choice is motivated by the following relation between $\mathbb{A}^n$ and $\mathbb{P}^n$.

**Remark 6.3** (Geometric interpretation of $\mathbb{P}^n$)**.** Consider the map

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

As in the picture below on the left we can embed the affine space $\mathbb{A}^n$ in $K^{n+1}$ at the height $x_0 = 1$, and then think of $f$ as mapping a point to the 1-dimensional linear subspace spanned by it.




The map $f$ is obviously injective, with image $U_0 := \{(x_0:\cdots:x_n) : x_0 \neq 0\}$. On this image the inverse of $f$ is given by

$$f^{-1} : U_0 \to \mathbb{A}^n, \ (x_0:\cdots:x_n) \mapsto \left(\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}\right),$$

it sends a line through the origin to its intersection point with $\mathbb{A}^n$ embedded in $K^{n+1}$. We can thus think of $\mathbb{A}^n$ as a subset $U_0$ of $\mathbb{P}^n$. The coordinates $\left(\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}\right)$ of a point $(x_0:\cdots:x_n) \in U_0 \subset \mathbb{P}^n$ are called its **affine coordinates**.

The remaining points of $\mathbb{P}^n$ are of the form $(0:x_1:\cdots:x_n)$; in the picture above they correspond to lines in the horizontal plane through the origin, such as e. g. $b$. By forgetting their first coordinate (which is zero anyway) they form a set that is naturally bijective to $\mathbb{P}^{n-1}$. Thinking of the ground field $\mathbb{C}$ we can regard them as *points at infinity*: consider e. g. in $\mathbb{A}^n_{\mathbb{C}}$ a parametrized line

$$x(t) = (a_1 + b_1 t, \dots, a_n + b_n t) \quad \text{for } t \in \mathbb{C}$$

for some starting point $(a_1, \dots, a_n) \in \mathbb{C}^n$ and direction vector $(b_1, \dots, b_n) \in \mathbb{C}^n \backslash \{0\}$. Of course, there is no limit point of $x(t)$ in $\mathbb{A}^n_{\mathbb{C}}$ as $t \to \infty$. But if we embed $\mathbb{A}^n_{\mathbb{C}}$ in $\mathbb{P}^n_{\mathbb{C}}$ as above we have

$$x(t) = (1:a_1 + b_1 t:\cdots:a_n + b_n t) = \left(\frac{1}{t} : \frac{a_1}{t} + b_1 : \cdots : \frac{a_n}{t} + b_n\right)$$

in homogeneous coordinates, and thus (in a suitable topology) we get a limit point $(0:b_1:\cdots:b_n) \in \mathbb{P}^n_{\mathbb{C}} \backslash \mathbb{A}^n_{\mathbb{C}}$ at infinity for $x(t)$ as $t \to \infty$. This limit point obviously remembers the *direction*, but not the *position* of the original line. Hence we can say that in $\mathbb{P}^n$ we have added a point at infinity to $\mathbb{A}^n$ in each direction. In other words, after extension to $\mathbb{P}^n$ two distinct lines in $\mathbb{A}^n$ will meet at infinity if and only if they are parallel, i. e. point in the same direction.

Usually, it is more helpful to think of the projective space $\mathbb{P}^n$ as the affine space $\mathbb{A}^n$ compactified by adding some points (parametrized by $\mathbb{P}^{n-1}$) at infinity, rather than as the set of 1-dimensional linear subspaces in $K^{n+1}$. In fact, after having given $\mathbb{P}^n$ the structure of a variety we will see in Proposition 7.2 and Exercise 7.3 (b) that with the above constructions $\mathbb{A}^n$ and $\mathbb{P}^{n-1}$ are open and closed subvarieties of $\mathbb{P}^n$, respectively.

**Remark 6.4** ($\mathbb{P}^n_{\mathbb{C}}$ is compact in the classical topology)**.** In the case $K = \mathbb{C}$ one can give $\mathbb{P}^n_{\mathbb{C}}$ a standard (quotient) topology by declaring a subset $U \subset \mathbb{P}^n$ to be open if its inverse image under the quotient map $\pi : \mathbb{C}^{n+1} \backslash \{0\} \to \mathbb{P}^n$ is open in the standard topology. Then $\mathbb{P}^n_{\mathbb{C}}$ is compact: let

$$S = \{(x_0, \dots, x_n) \in \mathbb{C}^{n+1} : |x_0|^2 + \cdots + |x_n|^2 = 1\}$$

be the unit sphere in $\mathbb{C}^{n+1}$. This is a compact space as it is closed and bounded. Moreover, as every point in $\mathbb{P}^n$ can be represented by a unit vector in $S$, the restricted map $\pi|_S : S \to \mathbb{P}^n$ is surjective. Hence $\mathbb{P}^n$ is compact as a continuous image of a compact set.

**Remark 6.5** (Homogeneous polynomials)**.** In complete analogy to affine varieties, we now want to define projective varieties to be subsets of $\mathbb{P}^n$ that can be given as the zero locus of some polynomials in the homogeneous coordinates. Note however that if $f \in K[x_0, \dots, x_n]$ is an arbitrary polynomial, it does not make sense to write down a definition like

$$V(f) = \{(x_0:\cdots:x_n) : f(x_0, \dots, x_n) = 0\} \quad \subset \mathbb{P}^n,$$

because the homogeneous coordinates are only defined up to a common scalar. For example, if $f = x_1^2 - x_0 \in K[x_0, x_1]$ then $f(1,1) = 0$ and $f(-1,-1) \neq 0$, although $(1:1) = (-1:-1)$ in $\mathbb{P}^1$. To get rid of this problem we have to require that $f$ is *homogeneous*, i. e. that all of its monomials have the same (total) degree $d$: in this case

$$f(\lambda x_0,\dots,\lambda x_n) = \lambda^d f(x_0,\dots,x_n) \text{ for all } \lambda \in K^*,$$

and so in particular we see that

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

so that the zero locus of $f$ is well-defined in $\mathbb{P}^n$. So before we can start with our discussion of projective varieties we have to set up some algebraic language to be able to talk about homogeneous elements in a ring (or $K$-algebra).

**Definition 6.6** (Graded rings and $K$-algebras)**.**

(a) A **graded ring** is a ring $R$ together with Abelian subgroups $R_d \subset R$ for all $d \in \mathbb{N}$, such that:

- Every element $f \in R$ has a unique decomposition $f = \sum_{d\in\mathbb{N}} f_d$ such that $f_d \in R_d$ for all $d \in \mathbb{N}$ and only finitely many $f_d$ are non-zero. In accordance with the direct sum notation in linear algebra, we usually write this condition as $R = \bigoplus_{d\in\mathbb{N}} R_d$.
- For all $d,e \in \mathbb{N}$ and $f \in R_d$, $g \in R_e$ we have $fg \in R_{d+e}$.

For $f \in R\backslash\{0\}$ the biggest number $d \in \mathbb{N}$ with $f_d \neq 0$ in the decomposition $f = \sum_{d\in\mathbb{N}} f_d$ as above is called the **degree** $\deg f$ of $f$. The elements of $R_d\backslash\{0\}$ are said to be **homogeneous** (of degree $d$). We call $f = \sum_{d\in\mathbb{N}} f_d$ and $R = \bigoplus_{d\in\mathbb{N}} R_d$ as above the **homogeneous decomposition** of $f$ and $R$, respectively.

(b) If $R$ is also a $K$-algebra in addition to (a), we say that it is a **graded $K$-algebra** if $\lambda f \in R_d$ for all $d \in \mathbb{N}$ and $f \in R_d$.

**Example 6.7.** The polynomial ring $R = K[x_0,\dots,x_n]$ is obviously a graded ring with

$$R_d = \left\{ \sum_{\substack{i_0,\dots,i_n\in\mathbb{N} \\ i_0+\cdots+i_n=d}} a_{i_0,\dots,i_n}\, x_0^{i_0}\cdot\;\cdots\;\cdot x_n^{i_n} : a_{i_0,\dots,i_n} \in K \text{ for all } i_0,\dots,i_n \right\}$$

for all $d \in \mathbb{N}$. In the following we will always consider it with this grading.

**Exercise 6.8.** Let $R \neq 0$ be a graded ring. Show that the multiplicative unit $1 \in R$ is homogeneous of degree 0.

Of course, we will also need ideals in graded rings. Naively, one might expect that we should consider ideals consisting of homogeneous elements in this case. However, as an ideal has to be closed under multiplication with *arbitrary* ring elements, it is virtually impossible that all of its elements are homogeneous. Instead, the correct notion of homogeneous ideal is the following.

**Definition 6.9** (Homogeneous ideals)**.** An ideal in a graded ring is called **homogeneous** if it can be generated by homogeneous elements.

**Lemma 6.10** (Properties of homogeneous ideals)**.** *Let $I$ and $J$ be ideals in a graded ring $R$.*

(a) *The ideal $I$ is homogeneous if and only if for all $f \in I$ with homogeneous decomposition $f = \sum_{d\in\mathbb{N}} f_d$ we also have $f_d \in I$ for all $d$.*

(b) *If $I$ and $J$ are homogeneous then so are $I+J$, $IJ$, $I\cap J$, and $\sqrt{I}$.*

(c) *If $I$ is homogeneous then the quotient $R/I$ is a graded ring with homogeneous decomposition $R/I = \bigoplus_{d\in\mathbb{N}} R_d/(R_d\cap I)$.*

*Proof.*

(a) "$\Rightarrow$": Let $I = (h_j : j \in J)$ for homogeneous elements $h_j \in R$ for all $j$, and let $f \in I$. Then $f = \sum_{j\in J} g_j h_j$ for some (not necessarily homogeneous) $g_j \in R$, of which only finitely many are non-zero. If we denote by $g_j = \sum_{e\in\mathbb{N}} g_{j,e}$ the homogeneous decompositions of these elements, the degree-$d$ part of $f$ for $d \in \mathbb{N}$ is

$$f_d = \sum_{\substack{j\in J, e\in\mathbb{N} \\ e+\deg h_j = d}} g_{j,e} h_j \in I.$$

"$\Leftarrow$": Now let $I = (h_j : j \in J)$ for arbitrary elements $h_j \in R$ for all $j$. If $h_j = \sum_{d\in\mathbb{N}} h_{j,d}$ is their homogeneous decomposition, we have $h_{j,d} \in I$ for all $j$ and $d$ by assumption, and thus $I = (h_{j,d} : j \in J, d \in \mathbb{N})$ can be generated by homogeneous elements.

(b) If $I$ and $J$ are generated by homogeneous elements, then clearly so are $I+J$ (which is generated by $I \cup J$) and $IJ$. Moreover, $I$ and $J$ then satisfy the equivalent condition of (a), and thus so does $I \cap J$.

It remains to be shown that $\sqrt{I}$ is homogeneous. We will check the condition of (a) for any $f \in \sqrt{I}$ by induction over the degree $d$ of $f$. Writing $f = f_0 + \cdots + f_d$ in its homogeneous decomposition, we get

$$f^n = (f_0 + \cdots + f_d)^n = f_d^n + (\text{terms of lower degree}) \quad \in I$$

for some $n \in \mathbb{N}$, hence $f_d^n \in I$ by (a), and thus $f_d \in \sqrt{I}$. But then $f - f_d = f_0 + \cdots + f_{d-1} \in \sqrt{I}$ as well, and so by the induction hypothesis we also see that $f_0, \dots, f_{d-1} \in \sqrt{I}$.

(c) It is clear that $R_d/(R_d \cap I) \to R/I$, $\overline{f} \mapsto \overline{f}$ is an injective group homomorphism, so that we can consider $R_d/(R_d \cap I)$ as a subgroup of $R/I$ for all $d$.

Now let $f \in R$ be arbitrary, with homogeneous decomposition $f = \sum_{d\in\mathbb{N}} f_d$. Then $\overline{f} = \sum_{d\in\mathbb{N}} \overline{f_d}$ with $\overline{f_d} \in R_d/(R_d \cap I)$, so $\overline{f}$ also has a homogeneous decomposition. Moreover, this decomposition is unique: if $\sum_{d\in\mathbb{N}} \overline{f_d} = \sum_{d\in\mathbb{N}} \overline{g_d}$ are two such decompositions of the same element in $R/I$ then $\sum_{d\in\mathbb{N}}(f_d - g_d)$ lies in $I$, hence by (a) is of the form $\sum_{d\in\mathbb{N}} h_d$ with all $h_d \in R_d \cap I$. But then

$$\sum_{d\in\mathbb{N}} (f_d - g_d - h_d) = 0,$$

which implies that $f_d - g_d - h_d = 0$, and thus $\overline{f_d} = \overline{g_d} \in R_d/(R_d \cap I)$ for all $d$. □

With this preparation we can now define projective varieties in the same way as affine ones. For simplicity, for a homogeneous polynomial $f \in K[x_0, \dots, x_n]$ and a point $x = (x_0 : \cdots : x_n) \in \mathbb{P}^n$ we will write the condition $f(x_0, \dots, x_n) = 0$ (which is well-defined by Remark 6.5) also as $f(x) = 0$.

**Definition 6.11** (Projective varieties and their ideals)**.** Let $n \in \mathbb{N}$.

(a) Let $S \subset K[x_0, \dots, x_n]$ be a set of homogeneous polynomials. Then the **(projective) zero locus** of $S$ is defined as

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

Subsets of $\mathbb{P}^n$ that are of this form are called **projective varieties**. For $S = (f_1, \dots, f_k)$ we will write $V(S)$ also as $V(f_1, \dots, f_k)$.

(b) For a homogeneous ideal $I \trianglelefteq K[x_0, \dots, x_n]$ we set

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

Obviously, if $I$ is the ideal generated by a set $S$ of homogeneous polynomials then $V(I) = V(S)$.

(c) If $X \subset \mathbb{P}^n$ is any subset we define its **ideal** to be

$$I(X) := \big(f \in K[x_0, \dots, x_n] \text{ homogeneous} : f(x) = 0 \text{ for all } x \in X\big) \quad \trianglelefteq K[x_0, \dots, x_n].$$

(Note that the homogeneous polynomials vanishing on $X$ do not form an ideal yet, so that we have to take the ideal generated by them.)

If we want to distinguish these projective constructions from the affine ones in Definitions 1.2 (c) and 1.10 we will denote them by $V_p(S)$ and $I_p(X)$, and the affine ones by $V_a(S)$ and $I_a(X)$, respectively.

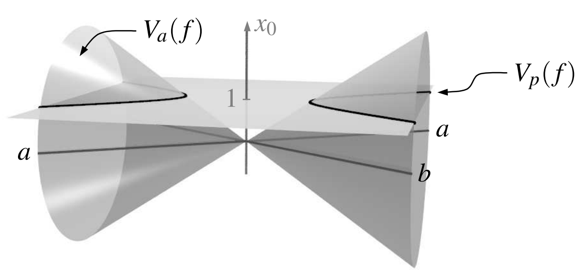
**Example 6.12.**

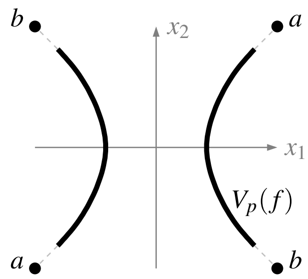
(a) As in the affine case, the empty set $\emptyset = V_p(1)$ and the whole space $\mathbb{P}^n = V_p(0)$ are projective varieties.

(b) If $f_1,\dots,f_r \in K[x_0,\dots,x_n]$ are homogeneous linear polynomials then $V_p(f_1,\dots,f_r) \subset \mathbb{P}^n$ is a projective variety. Projective varieties that are of this form are called **linear subspaces** of $\mathbb{P}^n$.

**Exercise 6.13.** Let $a \in \mathbb{P}^n$ be a point. Show that the one-point set $\{a\}$ is a projective variety, and compute explicit generators for the ideal $I_p(\{a\}) \trianglelefteq K[x_0,\dots,x_n]$.

**Example 6.14.** Let $f = x_1^2 - x_2^2 - x_0^2 \in \mathbb{C}[x_0,x_1,x_2]$. The real part of the affine zero locus $V_a(f) \subset \mathbb{A}^3$ of this homogeneous polynomial is the 2-dimensional cone shown in the picture below on the left. According to Definition 6.11, its projective zero locus $V_p(f) \subset \mathbb{P}^2$ is the set of all 1-dimensional linear subspaces contained in this cone — but we have seen in Remark 6.3 already that we should rather think of $\mathbb{P}^2$ as the affine plane $\mathbb{A}^2$ (embedded in $\mathbb{A}^3$ at $x_0 = 1$) together with some points at infinity. With this interpretation the real part of $V_p(f)$ consists of the hyperbola shown below on the right (whose equation $x_1^2 - x_2^2 - 1 = 0$ can be obtained by setting $x_0 = 1$ in $f$), together with two points $a$ and $b$ at infinity. In the 3-dimensional picture on the left, these two points correspond to the two 1-dimensional linear subspaces parallel to the plane at $x_0 = 1$, in the 2-dimensional picture of the affine part in $\mathbb{A}^2$ on the right they can be thought of as points at infinity in the corresponding directions. Note that, in the latter interpretation, "opposite" points at infinity are actually the same, since they correspond to the same 1-dimensional linear subspace in $\mathbb{C}^3$.




09

We see in this example that the affine and projective zero locus of $f$ carry essentially the same geometric information — the difference is just whether we consider the cone as a set of individual points, or as a union of 1-dimensional linear subspaces in $\mathbb{A}^3$. Let us now formalize and generalize this correspondence.

**Definition 6.15** (Cones)**.** Let $\pi : \mathbb{A}^{n+1}\backslash\{0\} \to \mathbb{P}^n$, $(x_0,\dots,x_n) \mapsto (x_0:\dots:x_n)$.

(a) An affine variety $X \subset \mathbb{A}^{n+1}$ is called a **cone** if $0 \in X$, and $\lambda x \in X$ for all $\lambda \in K$ and $x \in X$. In other words, it consists of the origin together with a union of lines through 0.

(b) For a cone $X \subset \mathbb{A}^{n+1}$ we call

$$\mathbb{P}(X) := \pi(X\backslash\{0\}) = \{(x_0:\dots:x_n) \in \mathbb{P}^n : (x_0,\dots,x_n) \in X\} \quad \subset \mathbb{P}^n$$

the **projectivization** of $X$.

(c) For a projective variety $X \subset \mathbb{P}^n$ we call

$$C(X) := \{0\} \cup \pi^{-1}(X) = \{0\} \cup \{(x_0,\dots,x_n) : (x_0:\dots:x_n) \in X\} \quad \subset \mathbb{A}^{n+1}$$

the **cone** over $X$ (we will see in Remark 6.17 that this is in fact a cone in the sense of (a)).

**Remark 6.16** (Cones and homogeneous ideals)**.** If $S \subset K[x_0,\dots,x_n]$ is a set of homogeneous polynomials with non-empty affine zero locus in $\mathbb{A}^{n+1}$ then $V_a(S)$ is a cone: clearly, we have $0 \in V_a(S)$ as every non-constant homogeneous polynomial vanishes at the origin. Moreover, let $\lambda \in K$ and $x \in V_a(S)$. Then $f(x) = 0$ for all $f \in S$, hence $f(\lambda x) = \lambda^{\deg f} f(x) = 0$, and so $\lambda x \in V_a(S)$ as well.

Conversely, the ideal $I(X)$ of a cone $X \subset \mathbb{A}^{n+1}$ is homogeneous: let $f \in I(X)$ with homogeneous decomposition $f = \sum_{d\in\mathbb{N}} f_d$. Then for all $x \in X$ we have $f(x) = 0$, and therefore also

$$0 = f(\lambda x) = \sum_{d\in\mathbb{N}} \lambda^d f_d(x)$$

for all $\lambda \in K$ since $X$ is a cone. This means that we have the zero polynomial in $\lambda$, i. e. that $f_d(x) = 0$ for all $d$, and thus $f_d \in I(X)$. Hence $I(X)$ is homogeneous by Lemma 6.10 (a).

**Remark 6.17** (Cones $\leftrightarrow$ projective varieties)**.** Let $S \subset K[x_0,\dots,x_n]$ be a set of homogeneous polynomials with non-empty affine zero locus in $\mathbb{A}^{n+1}$. Then $V_a(S)$ is a cone by Remark 6.16, and by construction we have

$$\mathbb{P}(V_a(S)) = V_p(S) \qquad \text{and} \qquad C(V_p(S)) = V_a(S).$$

In particular, the projectivization $\mathbb{P}(V_a(S))$ is a projective variety, and $C(V_p(S))$ is a cone. Moreover, as Remark 6.16 also shows that every cone is of the form $V_a(S)$ for a suitable set $S$ of homogeneous polynomials (namely generators for its homogeneous ideal), we obtain a one-to-one correspondence

$$\begin{array}{ccc} \{\text{cones in } \mathbb{A}^{n+1}\} & \longleftrightarrow & \{\text{projective varieties in } \mathbb{P}^n\} \\ X & \longmapsto & \mathbb{P}(X) \\ C(X) & \longleftarrow\!\shortmid & X. \end{array}$$

In other words, the correspondence works by passing from the affine to the projective zero locus (and vice versa) of the same set of homogeneous polynomials, as in Example 6.14. Note that in this way linear subspaces of $\mathbb{A}^{n+1}$ correspond exactly to linear subspaces of $\mathbb{P}^n$ in the sense of Example 6.12 (b).

Having defined projective varieties, we can now proceed with their study as in the affine case. First of all, we should associate a coordinate ring to a projective variety, and consider zero loci and ideals with respect to these coordinate rings.

**Construction 6.18** (Relative version of zero loci and ideals)**.** Let $Y \subset \mathbb{P}^n$ be a projective variety. In analogy to Definition 1.19 we call

$$S(Y) := K[x_0,\dots,x_n]/I(Y)$$

the **homogeneous coordinate ring** of $Y$. By Lemma 6.10 (c) it is a graded ring, so that it makes sense to talk about homogeneous elements of $S(Y)$. Moreover, the condition $f(x) = 0$ is still well-defined for a homogeneous element $f \in S(Y)$ and a point $x \in Y$, and thus we can define as in Definition 6.11

$$V(I) := \{x \in Y : f(x) = 0 \text{ for all homogeneous } f \in I\} \qquad \text{for a homogeneous ideal } I \trianglelefteq S(Y)$$

(and similarly for a set of homogeneous polynomials in $S(Y)$), and

$$I(X) := \big(f \in S(Y) \text{ homogeneous} : f(x) = 0 \text{ for all } x \in X\big) \qquad \text{for a subset } X \subset Y.$$

As before, in case of possible confusion we will decorate $V$ and $I$ with the subscript $Y$ and/or $p$ to denote the relative and projective situation, respectively. Subsets of $Y$ that are of the form $V_Y(I)$ for a homogeneous ideal $I \trianglelefteq S(Y)$ will be called **projective subvarieties** of $Y$; these are obviously exactly the projective varieties contained in $Y$.

**Remark 6.19.** Let $Y$ be a projective variety. The following results are completely analogous to the affine case:

(a) (*Hilbert's Basis Theorem*) Every homogeneous ideal in $S(Y)$ can be generated by finitely many elements. In fact, it is finitely generated by [G5, Proposition 7.13 and Remark 7.15], and hence also by homogeneous elements as we have seen in the proof of part "$\Leftarrow$" of Lemma 6.10 (a).

(b) The operations $V_Y(\cdot)$ and $I_Y(\cdot)$ reverse inclusions, we have $X = V_Y(I_Y(X))$ for every projective subvariety $X$ of $Y$, and $J \subset I_Y(V_Y(J))$ for any homogeneous ideal $J \trianglelefteq S(Y)$ — these statements follow literally in the same way as in Lemma 1.12.

(c) The ideal $I_Y(X)$ of a projective subvariety $X \subset Y$ is radical: by Lemma 6.10 (b) the radical $\sqrt{I_Y(X)}$ is homogeneous, so it suffices to prove that $f \in \sqrt{I_Y(X)}$ implies $f \in I_Y(X)$ for any homogeneous $f$. But this is obvious since $f^k = 0$ on $Y$ for some $k$ implies $f = 0$ on $Y$.

(d) By (c) the ideal $I_p(Y) \trianglelefteq K[x_0,\dots,x_n]$ is radical. Hence Proposition 1.17 implies that it is also the ideal of its affine zero locus $V_a(I_p(Y)) \subset \mathbb{A}^{n+1}$. But $V_p(I_p(Y)) = Y$ by (b), and so we see by Remark 6.17 that $V_a(I_p(Y)) = C(Y)$. Therefore we conclude that $I_p(Y) = I_a(C(Y))$, and thus that $S(Y) = A(C(Y))$. Hence every homogeneous coordinate ring of a projective variety can also be interpreted as a usual coordinate ring of an affine variety.

**Remark 6.20.** A remark that is sometimes useful is that every projective subvariety $X$ of a projective variety $Y \subset \mathbb{P}^n$ can be written as the zero locus of finitely many homogeneous polynomials in $S(Y)$ *of the same degree*. This follows easily from the fact that $V_p(f) = V_p(x_0^d f,\dots,x_n^d f)$ for all homogeneous $f \in S(Y)$ and every $d \in \mathbb{N}$. However, it is not true that every homogeneous ideal in $S(Y)$ can be generated by homogeneous elements of the same degree.

Of course, we would also expect a projective version of the Nullstellensatz as in Proposition 1.21 (b), i. e. that $I_Y(V_Y(J)) = \sqrt{J}$ for any homogeneous ideal $J$ in the homogeneous coordinate ring of a projective variety $Y$. This is *almost* true and can in fact be proved by reduction to the affine case — there is one exception however, since the origin in $\mathbb{A}^{n+1}$ does not correspond to a point in projective space $\mathbb{P}^n$:

**Example 6.21** (Irrelevant ideal)**.** Let $Y \subset \mathbb{P}^n$ be a non-empty projective variety, and let

$$I_0 := (\overline{x_0},\dots,\overline{x_n}) \trianglelefteq S(Y) = K[x_0,\dots,x_n]/I(Y).$$

Then $I_0$ is a homogeneous radical ideal, and its projective zero locus is empty since there is no point in $Y$ all of whose homogeneous coordinates are zero. Hence

$$I_Y(V_Y(I_0)) = I_Y(\emptyset) = S(Y),$$

which is not equal to $\sqrt{I_0} = I_0$. In fact, $I_0$ can never appear as the ideal of a projective variety, since $I_0 = I_Y(X)$ for some $X \subset Y$ would imply $X = V_Y(I_Y(X)) = V_Y(I_0) = \emptyset$ by Remark 6.19 (b), in contradiction to $I_Y(\emptyset) = S(Y)$.

We will see now however that this is the only counterexample to the projective version of the Nullstellensatz. The ideal $I_0$ above is therefore often called the **irrelevant ideal**. Note that by Proposition 1.21 (b) this is the unique radical ideal whose affine zero locus is $\{0\}$.

**Proposition 6.22 (Projective Nullstellensatz).** *Let $Y$ be a non-empty projective variety, and let $J \subset S(Y)$ be a homogeneous ideal such that $\sqrt{J}$ is not the irrelevant ideal. Then $I_p(V_p(J)) = \sqrt{J}$. In particular, we have an inclusion-reversing one-to-one correspondence*

$$\begin{array}{rcl} \{\text{projective subvarieties of } Y\} & \longleftrightarrow & \left\{\begin{array}{c}\text{homogeneous radical ideals in } S(Y) \\ \text{not equal to the irrelevant ideal}\end{array}\right\} \\ X & \longmapsto & I_p(X) \\ V_p(J) & \longleftarrow\!\shortmid & J. \end{array}$$

*Proof.* In this proof we will regard $J$ as an ideal in both $S(Y)$ and $A(C(Y))$ (see Remark 6.19 (d)), so that we can take both its projective and its affine zero locus. Note then that

$$\begin{aligned} I_p(V_p(J)) &= \big(f \in S(Y) \text{ homogeneous} : f(x) = 0 \text{ for all } x \in V_p(J)\big) \\ &= \big(f \in S(Y) \text{ homogeneous} : f(x) = 0 \text{ for all } x \in V_a(J)\backslash\{0\}\big). \end{aligned}$$

As the affine zero locus of polynomials is closed, we can rewrite this as

$$I_p(V_p(J)) = \big(f \in S(Y) \text{ homogeneous} : f(x) = 0 \text{ for all } x \in \overline{V_a(J)\backslash\{0\}}\big).$$

By Example 6.21 we have $V_a(J) = V_a(\sqrt{J}) \neq \{0\}$ since $\sqrt{J}$ is not the irrelevant ideal. But then $\overline{V_a(J)\backslash\{0\}} = V_a(J)$: if $V_a(J) = \emptyset$ this is trivial, and otherwise the cone $V_a(J)\backslash\{0\}$ contains at least one line without the origin, so that the origin lies in its closure. Hence we get

$$I_p(V_p(J)) = \big(f \in S(Y) \text{ homogeneous} : f(x) = 0 \text{ for all } x \in V_a(J)\big).$$

As the ideal of the cone $V_a(J)$ is homogeneous by Remark 6.16 this can be rewritten as $I_p(V_p(J)) = I_a(V_a(J))$, which is equal to $\sqrt{J}$ by the affine Nullstellensatz of Proposition 1.21 (b).

The one-to-one correspondence then follows together with the statement $V_p(I_p(X)) = X$ from Remark 6.19 (b) (note that $I_p(X)$ is always homogeneous by definition, radical by Remark 6.19 (c), and not equal to the irrelevant ideal by Example 6.21). □

**Remark 6.23** (Properties of $V_p(\cdot)$ and $I_p(\cdot)$)**.** The operations $V_p(\cdot)$ and $I_p(\cdot)$ satisfy the same properties as their affine counterparts in Lemma 1.24, Remark 1.25, and Lemma 1.26. More precisely, for any projective variety $X$ we have:

(a) For any family $\{S_i\}$ of subsets of $S(X)$ we have $\bigcap_i V_p(S_i) = V_p(\bigcup_i S_i)$; for any two subsets $S_1, S_2 \subset S(X)$ we have $V_p(S_1) \cup V_p(S_2) = V_p(S_1 S_2)$.

(b) If $J_1, J_2 \trianglelefteq S(X)$ are homogeneous ideals then

$$V_p(J_1) \cap V_p(J_2) = V_p(J_1 + J_2) \quad \text{and} \quad V_p(J_1) \cup V_p(J_2) = V_p(J_1 J_2) = V_p(J_1 \cap J_2).$$

(c) For subsets $Y_1, Y_2$ of a projective variety $X$ we have $I_p(Y_1 \cup Y_2) = I_p(Y_1) \cap I_p(Y_2)$. Moreover, $I_p(Y_1 \cap Y_2) = \sqrt{I_p(Y_1) + I_p(Y_2)}$ unless the latter is the irrelevant ideal (which is only possible if $Y_1$ and $Y_2$ are disjoint).

The proof of these statements is completely analogous to the affine case.

In particular, by (a) it follows that arbitrary intersections and finite unions of projective subvarieties of $X$ are again projective subvarieties, and hence we can define the Zariski topology on $X$ in the same way as in the affine case:

**Definition 6.24** (Zariski topology)**.** The **Zariski topology** on a projective variety $X$ is the topology whose closed sets are exactly the projective subvarieties of $X$, i. e. the subsets of the form $V_p(S)$ for some set $S \subset S(X)$ of homogeneous elements.

Of course, from now on we will always use this topology for projective varieties and their subsets. Note that, in the same way as in Remark 2.3, this is well-defined in the sense that the Zariski topology on a projective variety $X \subset \mathbb{P}^n$ agrees with the subspace topology of $X$ in $\mathbb{P}^n$. Moreover, since we want to consider $\mathbb{A}^n$ as a subset of $\mathbb{P}^n$ as in Remark 6.3 we should also check that the Zariski topology on $\mathbb{A}^n$ is the same as the subspace topology of $\mathbb{A}^n$ in $\mathbb{P}^n$. To do this, we need the following definition.

**Definition 6.25** (Homogenization)**.**

(a) Let

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

be a (non-zero) polynomial of degree $d$. We define its **homogenization** to be

$$\begin{aligned} f^h &:= x_0^d\, f\left(\frac{x_1}{x_0},\dots,\frac{x_n}{x_0}\right) \\ &= \sum_{i_1,\dots,i_n \in \mathbb{N}} a_{i_1,\dots,i_n}\, x_0^{d-i_1-\dots-i_n}\, x_1^{i_1} \cdot \dots \cdot x_n^{i_n} \quad \subset K[x_0,\dots,x_n]; \end{aligned}$$

obviously this is a homogeneous polynomial of degree $d$.

(b) The **homogenization** of an ideal $I \trianglelefteq K[x_1,\dots,x_n]$ is defined to be the ideal $I^h$ in $K[x_0,\dots,x_n]$ generated by all $f^h$ for $f \in I$.

**Example 6.26.** For $f = x_1^2 - x_2^2 - 1 \in K[x_1,x_2]$ we have $f^h = x_1^2 - x_2^2 - x_0^2 \in K[x_0,x_1,x_2]$.

**Remark 6.27.** If $f,g \in K[x_1,\dots,x_n]$ are polynomials of degree $d$ and $e$, respectively, then $fg$ has degree $d+e$, and so we get

$$(fg)^h = x_0^{d+e} f\left(\frac{x_1}{x_0},\dots,\frac{x_n}{x_0}\right) \cdot g\left(\frac{x_1}{x_0},\dots,\frac{x_n}{x_0}\right) = f^h \cdot g^h.$$

However, $(f+g)^h$ is clearly not equal to $f^h + g^h$ in general — in fact, $f^h + g^h$ is usually not even homogeneous. This is the reason why in Definition 6.25 (b) we have to take the ideal generated by all homogenizations of polynomials in $I$, instead of just all these homogenizations themselves.

**Remark 6.28** ($\mathbb{A}^n$ as an open subset of $\mathbb{P}^n$)**.** Recall from Remark 6.3 that we want to identify the subset $U_0 = \{(x_0 : \cdots : x_n) \in \mathbb{P}^n : x_0 \neq 0\}$ of $\mathbb{P}^n$ with $\mathbb{A}^n$ by the bijective map

$$F : \mathbb{A}^n \to U_0,\ (x_1,\dots,x_n) \mapsto (1 : x_1 : \cdots : x_n).$$

Obviously, $U_0$ is an open subset of $\mathbb{P}^n$. Moreover, with the above identification the subspace topology of $U_0 = \mathbb{A}^n \subset \mathbb{P}^n$ is the affine Zariski topology:

(a) If $X = V_p(S) \cap \mathbb{A}^n$ is closed in the subspace topology (for a subset $S \subset K[x_0,\dots,x_n]$ of homogeneous polynomials) then $X = V(f(1,\cdot) : f \in S)$ is also Zariski closed.

(b) If $X = V(S) \subset \mathbb{A}^n$ is Zariski closed (with $S \subset K[x_1,\dots,x_n]$) then $X = V_p(f^h : f \in S) \cap \mathbb{A}^n$ is closed in the subspace topology as well.

In other words we can say that the map $F : \mathbb{A}^n \to U_0$ above is a homeomorphism. In fact, after having given $\mathbb{P}^n$ the structure of a variety we will see in Proposition 7.2 that it is even an isomorphism of varieties.

Having defined the Zariski topology on projective varieties (or more generally on subsets of $\mathbb{P}^n$) we can now immediately apply all topological concepts of Chapter 2 to this new situation. In particular, the notions of connectedness, irreducibility, and dimension are well-defined for projective varieties (and have the same geometric interpretation as in the affine case). Let us study some examples using these concepts.

**Remark 6.29** ($\mathbb{P}^n$ is irreducible of dimension $n$)**.** Of course, by symmetry of the coordinates, it follows from Remark 6.28 that all subsets $U_i = \{(x_0 : \cdots : x_n) : x_i \neq 0\}$ of $\mathbb{P}^n$ for $i = 0,\dots,n$ are homeomorphic to $\mathbb{A}^n$ as well. As these subsets cover $\mathbb{P}^n$ and have non-empty intersections, we conclude by Exercise 2.20 (b) that $\mathbb{P}^n$ is irreducible, and by Exercise 2.33 (a) that $\dim \mathbb{P}^n = n$.

**Exercise 6.30.** Let $L_1, L_2 \subset \mathbb{P}^3$ be two disjoint lines (i. e. 1-dimensional linear subspaces in the sense of Example 6.12 (b)), and let $a \in \mathbb{P}^3 \backslash (L_1 \cup L_2)$. Show that there is a unique line $L \subset \mathbb{P}^3$ through $a$ that intersects both $L_1$ and $L_2$.

Is the corresponding statement for lines and points in $\mathbb{A}^3$ true as well?

**Exercise 6.31.**

(a) Prove that a graded ring $R$ is an integral domain if and only if for all *homogeneous* elements $f,g \in R$ with $fg = 0$ we have $f = 0$ or $g = 0$.

(b) Show that a projective variety $X$ is irreducible if and only if its homogeneous coordinate ring $S(X)$ is an integral domain.

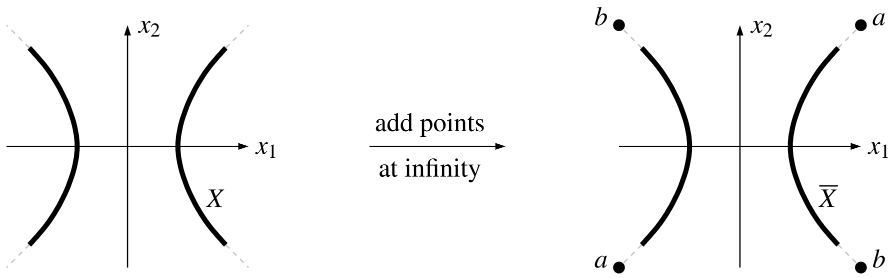
**Exercise 6.32.** In this exercise we want to show that an intersection of projective varieties is never empty unless one would expect it to be empty for dimensional reasons — so e. g. the phenomenon of parallel non-intersecting lines in the plane does not occur in projective space (which we have seen already in Remark 6.3).

So let $X,Y \subset \mathbb{P}^n$ be non-empty projective varieties. Show:

(a) The dimension of the cone $C(X) \subset \mathbb{A}^{n+1}$ is $\dim X + 1$.

(b) If $\dim X + \dim Y \geq n$ then $X \cap Y \neq \emptyset$.

We have just seen in Remark 6.28 (b) that we can use homogenizations of polynomials to describe an affine variety $X \subset \mathbb{A}^n$ in terms of their homogeneous coordinates on $\mathbb{A}^n \subset \mathbb{P}^n$. Let us now finish this chapter by showing that this construction can also be used to compute the closure of $X$ in $\mathbb{P}^n$. As this will be a "compact" space in the sense of Remarks 6.3 and 6.4 we can think of this closure $\overline{X}$ as being obtained by compactifying $X$ by some "points at infinity". For example, if we start with the affine hyperbola $X = V_a(x_1^2 - x_2^2 - 1) \subset \mathbb{A}^2$ in the picture below on the left, its closure $\overline{X} \subset \mathbb{P}^2$ adds the two points $a$ and $b$ at infinity as in Example 6.14.




We know already in this example that $\overline{X} = V_p(x_1^2 - x_2^2 - x_0^2)$, i. e. that the closure is the projective zero locus of the homogenization of the original polynomial $x_1^2 - x_2^2 - 1$. Let us now prove the corresponding general statement.

**Proposition 6.33** (Computation of the projective closure)**.** *Let $I \trianglelefteq K[x_1,\dots,x_n]$ be an ideal. Consider its affine zero locus $X = V_a(I) \subset \mathbb{A}^n$, and its closure $\overline{X}$ in $\mathbb{P}^n$.*

(a) *We have $\overline{X} = V_p(I^h)$.*

(b) *If $I = (f)$ is a principal ideal then $\overline{X} = V_p(f^h)$.*

*Proof.*

(a) Clearly, the set $V_p(I^h)$ is closed and contains $X$: if $x = (x_1,\dots,x_n) \in X$ then $f(x_1,\dots,x_n) = 0$ and thus $f^h(1,x_1,\dots,x_n) = 0$ for all $f \in I$, which implies that $(1:x_1:\cdots:x_n) \in V_p(I^h)$.

In order to show that $V_p(I^h)$ is the smallest closed set containing $X$ let $Y \supset X$ be any closed set; we have to prove that $Y \supset V_p(I^h)$. As $Y$ is closed we have $Y = V_p(J)$ for some homogeneous ideal $J$. Now any homogeneous element of $J$ can be written as $x_0^d f^h$ for some $d \in \mathbb{N}$ and $f \in K[x_1,\dots,x_n]$, and for this element we have

$$\begin{aligned}
& x_0^d f^h \text{ is zero on } X && (X \text{ is a subset of } Y)\\
\Rightarrow\ & f \text{ is zero on } X && (x_0 \neq 0 \text{ on } X \subset \mathbb{A}^n)\\
\Rightarrow\ & f \in I_a(X) = I_a(V_a(I)) = \sqrt{I} && (\text{Proposition 1.17})\\
\Rightarrow\ & f^m \in I \text{ for some } m \in \mathbb{N} &&\\
\Rightarrow\ & (f^h)^m = (f^m)^h \in I^h \text{ for some } m \in \mathbb{N} && (\text{Remark 6.27})\\
\Rightarrow\ & f^h \in \sqrt{I^h} &&\\
\Rightarrow\ & x_0^d f^h \in \sqrt{I^h}. &&
\end{aligned}$$

We therefore conclude that $J \subset \sqrt{I^h}$, and so $Y = V_p(J) \supset V_p(\sqrt{I^h}) = V_p(I^h)$ as desired.

(b) As $(f) = \{fg : g \in K[x_1,\dots,x_n]\}$, we have

$$\overline{X} = V((fg)^h : g \in K[x_1,\dots,x_n]) = V(f^h g^h : g \in K[x_1,\dots,x_n]) = V(f^h)$$

by (a) and Remark 6.27. □ 10

**Example 6.34.** In contrast to Proposition 6.33 (b), for general ideals it usually does not suffice to only homogenize a set of generators. As an example, consider the ideal $I = (x_1, x_2 - x_1^2) \trianglelefteq K[x_1,x_2]$ with affine zero locus $X = V_a(I) = \{0\} \subset \mathbb{A}^2$. This one-point set is also closed in $\mathbb{P}^2$, and thus

$\overline{X} = \{(1:0:0)\}$ is just the corresponding point in homogeneous coordinates. But if we homogenize the two given generators of $I$ we obtain the homogeneous ideal $(x_1, x_0x_2 - x_1^2)$ with projective zero locus $\{(1:0:0),(0:0:1)\} \supsetneq \overline{X}$.

For those of you who know some computer algebra: one can show however that it suffices to homogenize a *Gröbner basis* of $I$. This makes the problem of finding $\overline{X}$ computationally feasible since in contrast to Proposition 6.33 (a) we only have to homogenize finitely many polynomials.

**Exercise 6.35.** Sketch the set of real points of the complex affine curve $X = V(x_1^3 - x_1x_2^2 + 1) \subset \mathbb{A}^2_{\mathbb{C}}$ and compute the points at infinity of its projective closure $\overline{X} \subset \mathbb{P}^2_{\mathbb{C}}$.

# 7. Projective Varieties II: Ringed Spaces

After having defined projective varieties as topological spaces, we will now give them the structure of ringed spaces to make them into varieties in the sense of Chapter 5. In other words, we have to define a suitable notion of regular functions on (open subsets of) projective varieties.

Of course, as in the affine case in Definition 3.1 the general idea is that a regular function should be a $K$-valued function that is locally a quotient of two polynomials. However, note that in contrast to the affine situation the elements of the homogeneous coordinate ring $S(X)$ of a projective variety $X$ are not well-defined functions on $X$: even if $f \in S(X)$ is homogeneous of degree $d$ we only have $f(\lambda x) = \lambda^d f(x)$ for all $x \in X$ and $\lambda \in K$. So the only way to obtain well-defined functions is to consider quotients of homogeneous polynomials of the same degree, so that the factor $\lambda^d$ cancels out:

**Definition 7.1** (Regular functions on projective varieties)**.** Let $U$ be an open subset of a projective variety $X$. A **regular function** on $U$ is a map $\varphi : U \to K$ with the following property: for every $a \in U$ there are $d \in \mathbb{N}$ and $f, g \in S(X)_d$ with $f(x) \neq 0$ and

$$\varphi(x) = \frac{g(x)}{f(x)}$$

for all $x$ in an open subset $U_a$ with $a \in U_a \subset U$.

It is obvious that the sets $\mathscr{O}_X(U)$ of regular functions on $U$ are subrings of the $K$-algebras of all functions from $U$ to $K$, and — by the local nature of the definition — that they form a sheaf $\mathscr{O}_X$ on $X$.

With this definition, let us check first of all that the open subsets of a projective variety where one of the coordinates is non-zero are affine varieties, so that projective varieties are prevarieties in the sense of Definition 5.1.

**Proposition 7.2** (Projective varieties are prevarieties)**.** *Let $X \subset \mathbb{P}^n$ be a projective variety. Then*

$$U_i = \{(x_0 : \cdots : x_n) \in X : x_i \neq 0\} \quad \subset X$$

*is an affine variety for all $i = 0, \dots, n$. In particular, $X$ is a prevariety.*

*Proof.* By symmetry it suffices to prove the statement for $i = 0$. Let $X = V_p(h_1, \dots, h_r)$ for some homogeneous polynomials $h_1, \dots, h_r \in K[x_0, \dots, x_n]$, and set $g_j(x_1, \dots, x_n) = h_j(1, x_1, \dots, x_n)$ for all $j = 1, \dots, r$. If $Y = V_a(g_1, \dots, g_r)$ we claim that

$$F : Y \to U_0, \ (x_1, \dots, x_n) \mapsto (1 : x_1 : \cdots : x_n)$$

is an isomorphism with inverse

$$F^{-1} : U_0 \to Y, \ (x_0 : \cdots : x_n) \mapsto \left( \frac{x_1}{x_0}, \dots, \frac{x_n}{x_0} \right).$$

In fact, it is clear by construction that these two maps are well-defined and inverse to each other. Moreover, similarly to Remark 6.28 they are continuous: the inverse image of a closed set $V_p(f_1, \dots, f_s) \cap U_0$ under $F$ is the closed set $V_a(f_1(1, \cdot), \dots, f_s(1, \cdot))$, and the image of a closed set $V_a(f_1, \dots, f_s) \subset Y$ under $F$ is the closed set $V_p(f_1^h, \dots, f_s^h) \cap U_0$.

Finally, we have to check that $F$ and $F^{-1}$ pull back regular functions to regular functions: a regular function on (an open subset of) $U_0$ is by Definition 7.1 locally of the form $\frac{p(x_0, \dots, x_n)}{q(x_0, \dots, x_n)}$ (with nowhere vanishing denominator) for two homogeneous polynomials $p$ and $q$ of the same degree. Then

$$F^* \frac{p(x_0, \dots, x_n)}{q(x_0, \dots, x_n)} = \frac{p(1, x_1, \dots, x_n)}{q(1, x_1, \dots, x_n)}$$

is a quotient of polynomials and thus a regular function on $Y$. Conversely, $F^{-1}$ pulls back a quotient $\frac{p(x_1,\dots,x_n)}{q(x_1,\dots,x_n)}$ of two polynomials to

$$(F^{-1})^* \frac{p(x_1,\dots,x_n)}{q(x_1,\dots,x_n)} = \frac{p\left(\frac{x_1}{x_0},\dots,\frac{x_n}{x_0}\right)}{q\left(\frac{x_1}{x_0},\dots,\frac{x_n}{x_0}\right)},$$

which is a regular function on $U_0$ since it can be rewritten as a quotient of two homogeneous polynomials of the same degree (by multiplying both the numerator and the denominator by $x_0^m$ for $m = \max(\deg p, \deg q)$). Hence $F$ is an isomorphism by Definition 4.3 (b), and so $U_0$ is an affine open subset of $X$.

In particular, as the open subsets $U_i$ for $i = 0,\dots,n$ cover $X$ we conclude that $X$ is a prevariety. □

**Exercise 7.3.** Check that Definition 7.1 (together with Proposition 7.2) is compatible with our earlier constructions in the following cases:

(a) The prevariety $\mathbb{P}^1$ is the same as the one introduced in Example 5.5 (a).

(b) If $X \subset \mathbb{P}^n$ is a projective variety then its structure sheaf as defined above is the same as the closed subprevariety structure of $X$ in $\mathbb{P}^n$ as in Construction 5.12 (b).

**Exercise 7.4.** Let $m,n \in \mathbb{N}_{>0}$. Use Exercise 6.32 to prove that $\mathbb{P}^m \times \mathbb{P}^n$ is not isomorphic to $\mathbb{P}^{m+n}$.

We have already mentioned that the major advantage of (subprevarieties of) projective varieties is that they have a global description with homogeneous coordinates that does not refer to gluing techniques. In fact, the following proposition shows that many morphisms between projective varieties can also be constructed without gluing.

**Lemma 7.5** (Morphisms of projective varieties)**.** *Let $X \subset \mathbb{P}^n$ be a projective variety, and let $f_0,\dots,f_m \in S(X)$ be homogeneous elements of the same degree. Then on the open subset $U := X \backslash V(f_0,\dots,f_m)$ these elements define a morphism*

$$f : U \to \mathbb{P}^m, \;\; x \mapsto (f_0(x):\cdots:f_m(x)).$$

*Proof.* First of all note that $f$ is well-defined set-theoretically: by definition of $U$ the image point can never be $(0:\cdots:0)$; and if we rescale the homogeneous coordinates $x_0,\dots,x_n$ of $x \in U$ we get

$$\begin{aligned}(f_0(\lambda x_0:\cdots:\lambda x_n):\cdots:f_m(\lambda x_0:\cdots:\lambda x_n))\\ &= (\lambda^d f_0(x_0:\cdots:x_n):\cdots:\lambda^d f_m(x_0:\cdots:x_n))\\ &= (f_0(x_0:\cdots:x_n):\cdots:f_m(x_0:\cdots:x_n)),\end{aligned}$$

where $d$ is the common degree of the $f_0,\dots,f_m$. To check that $f$ is a morphism we want to use the gluing property of Lemma 4.6. So let $\{V_i : i = 0,\dots,m\}$ be the affine open cover of $\mathbb{P}^m$ with $V_i = \{(y_0:\cdots:y_m) : y_i \neq 0\}$ for all $i$. Then the open subsets $U_i := f^{-1}(V_i) = \{x \in X : f_i(x) \neq 0\}$ cover $U$, and in the affine coordinates on $V_i$ the map $f|_{U_i}$ is given by the quotients of polynomials $\frac{f_j}{f_i}$ for $j = 0,\dots,m$ with $j \neq i$, which are regular functions on $U_i$ by Definition 7.1. Hence $f|_{U_i}$ is a morphism by Proposition 4.7, and so $f$ is a morphism by Lemma 4.6. □

**Example 7.6.**

(a) Let $A \in \mathrm{GL}(n+1,K)$ be an invertible matrix. Then $f : \mathbb{P}^n \to \mathbb{P}^n$, $x \mapsto Ax$ is a morphism with inverse $f^{-1} : \mathbb{P}^n \to \mathbb{P}^n$, $x \mapsto A^{-1}x$, and hence an isomorphism. We will refer to these maps as **projective automorphisms** of $\mathbb{P}^n$. In fact, we will see in Proposition 13.4 that these are the only isomorphisms of $\mathbb{P}^n$.

(b) Let $a = (1:0:\cdots:0) \in \mathbb{P}^n$ and $L = V(x_0) \cong \mathbb{P}^{n-1}$. Then the map

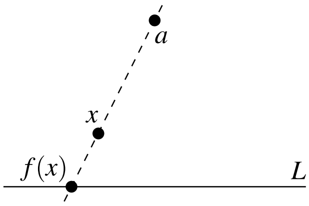
$$f : \mathbb{P}^n \backslash \{a\} \to \mathbb{P}^{n-1}, \;\; (x_0:\cdots:x_n) \mapsto (x_1:\cdots:x_n)$$

given by forgetting one of the homogeneous coordinates is a morphism by Lemma 7.5. It can be interpreted geometrically as in the picture below on the left: for $x = (x_0 : \cdots : x_n) \in \mathbb{P}^n \backslash \{a\}$ the unique line through $a$ and $x$ is clearly given parametrically by

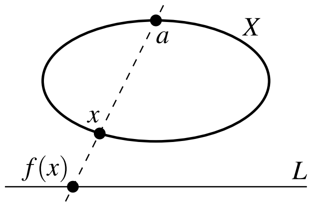
$$\{(s : t x_1 : \cdots : t x_n) : (s : t) \in \mathbb{P}^1\},$$

and its intersection point with $L$ is just $(0 : x_1 : \cdots : x_n)$, i. e. $f(x)$ with the identification $L \cong \mathbb{P}^{n-1}$. We call $f$ the **projection** from $a$ to the linear subspace $L$. Note however that the picture below is only schematic and does not show a standard affine open subset $U_i = \{(x_0 : \cdots : x_n) : x_i \neq 0\}$, since none of these subsets contains both $a$ and (a non-empty open subset of) $L$.

Of course, the same construction works for any point $a \in \mathbb{P}^n$ and any linear subspace $L$ of dimension $n-1$ not containing $a$ — the corresponding morphism then differs from the special one considered above by a projective automorphism as in (a).



(b)



(c)

(c) The projection morphism $f : \mathbb{P}^n \backslash \{a\} \to \mathbb{P}^{n-1}$ as in (b) cannot be extended to the point $a$. The intuitive reason for this is that the line through $a$ and $x$ (and thus also the point $f(x)$) does not have a well-defined limit as $x$ approaches $a$. This changes however if we restrict the projection to a suitable projective variety: for $X = V(x_0x_2 - x_1^2)$ as in the schematic picture above on the right consider the map

$$f : X \to \mathbb{P}^1, \ (x_0 : x_1 : x_2) \mapsto \begin{cases} (x_1 : x_2) & \text{if } (x_0 : x_1 : x_2) \neq (1:0:0), \\ (x_0 : x_1) & \text{if } (x_0 : x_1 : x_2) \neq (0:0:1). \end{cases}$$

It is clearly well-defined since the equation $x_0x_2 - x_1^2 = 0$ implies $(x_1 : x_2) = (x_0 : x_1)$ whenever both these points in $\mathbb{P}^1$ are defined. Moreover, it extends the projection as in (b) to all of $X$ (which includes the point $a$), and it is a morphism since it is patched together from two projections as above. Geometrically, the image $f(a)$ is the intersection of the tangent to $X$ at $a$ with the line $L$.

This geometric picture also tells us that $f$ is bijective: for every point $y \in L$ the restriction of the polynomial $x_0x_2 - x_1^2$ defining $X$ to the line through $a$ and $y$ has degree 2, and thus this line intersects $X$ in two points (counted with multiplicities), of which one is $a$. The other point is then the unique inverse image $f^{-1}(y)$. In fact, it is easy to check that $f$ is even an isomorphism since its inverse is

$$f^{-1} : \mathbb{P}^1 \to X, \ (y_0 : y_1) \mapsto (y_0^2 : y_0y_1 : y_1^2),$$

which is a morphism by Lemma 7.5.

Note that the example of the morphism $f$ above also shows that we cannot expect every morphism between projective varieties to have a global description by homogeneous polynomials as in Lemma 7.5.

(d) Now let $X \subset \mathbb{P}^2$ be any *projective conic*, i. e. the zero locus of a single irreducible homogeneous polynomial $f \in K[x_0, x_1, x_2]$ of degree 2. Assuming that $\operatorname{char} K \neq 2$, we know by Exercise 4.12 that the affine part $X \cap \mathbb{A}^2$ is isomorphic to $V_a(x_2 - x_1^2)$ or $V_a(x_1x_2 - 1)$ by a linear transformation followed by a translation. Extending this map to a projective automorphism of $\mathbb{P}^2$ as in (a), the projective conic $X$ thus becomes isomorphic to $V_p(x_0x_2 - x_1^2)$ or $V_p(x_1x_2 - x_0^2)$ by Proposition 6.33 (b). So by (c) we see that every projective conic is isomorphic to $\mathbb{P}^1$.

**Exercise 7.7.** Let us say that $n+2$ points in $\mathbb{P}^n$ are *in general position* if for any $n+1$ of them their representatives in $K^{n+1}$ are linearly independent.

Now let $a_1,\dots,a_{n+2}$ and $b_1,\dots,b_{n+2}$ be two sets of points in $\mathbb{P}^n$ in general position. Show that there is an isomorphism $f:\mathbb{P}^n\to\mathbb{P}^n$ with $f(a_i)=b_i$ for all $i=1,\dots,n+2$.

**Exercise 7.8.** Show by example that the homogeneous coordinate ring of a projective variety is *not* invariant under isomorphisms, i.e. that there are isomorphic projective varieties $X,Y$ such that the rings $S(X)$ and $S(Y)$ are not isomorphic.

**Exercise 7.9.** Let $f:\mathbb{P}^n\to\mathbb{P}^m$ be a morphism. Prove:

(a) If $X\subset\mathbb{P}^m$ is the zero locus of a single homogeneous polynomial in $K[x_0,\dots,x_m]$ then every irreducible component of $f^{-1}(X)$ has dimension at least $n-1$.

(b) If $n>m$ then $f$ must be constant.

Let us now verify that projective varieties are separated, i. e. that they are varieties and not just prevarieties. In other words, we have to check that the diagonal $\Delta_X$ of a projective variety $X$ is closed in the product $X\times X$. By Lemma 5.20 (b) it suffices to show this for $X=\mathbb{P}^n$.

For the proof of this statement it is useful to first find a good description of the product of projective spaces — note that by Exercise 7.4 such products are not just again projective spaces. Of course, we could just parametrize these products by two sets of homogeneous coordinates. It turns out however that we can also use a single set of homogeneous coordinates and thus embed products of projective spaces as a projective variety into a bigger projective space.

**Construction 7.10** (Segre embedding)**.** Consider $\mathbb{P}^m$ with homogeneous coordinates $x_0,\dots,x_m$ and $\mathbb{P}^n$ with homogeneous coordinates $y_0,\dots,y_n$. Set $N=(m+1)(n+1)-1$ and let $\mathbb{P}^N$ be the projective space with homogeneous coordinates labeled $z_{i,j}$ for $0\le i\le m$ and $0\le j\le n$. Then there is an obviously well-defined set-theoretic map

$$f:\mathbb{P}^m\times\mathbb{P}^n\to\mathbb{P}^N$$

11 given by $z_{i,j}=x_iy_j$ for all $i,j$. It satisfies the following properties:

**Proposition 7.11.** *Let $f:\mathbb{P}^m\times\mathbb{P}^n\to\mathbb{P}^N$ be the map of Construction 7.10. Then:*

(a) *The image $X=f(\mathbb{P}^m\times\mathbb{P}^n)$ is a projective variety given by*

$$X=V_p(z_{i,j}\,z_{k,l}-z_{i,l}\,z_{k,j}:0\le i,k\le m,0\le j,l\le n).$$

(b) *The map $f:\mathbb{P}^m\times\mathbb{P}^n\to X$ is an isomorphism.*

*In particular, $\mathbb{P}^m\times\mathbb{P}^n\cong X$ is a projective variety. The isomorphism $f:\mathbb{P}^m\times\mathbb{P}^n\to X\subset\mathbb{P}^N$ is called the **Segre embedding**; the coordinates $z_{0,0},\dots,z_{m,n}$ above will be referred to as **Segre coordinates** on $\mathbb{P}^m\times\mathbb{P}^n$.*

*Proof.*

(a) It is obvious that the points of $f(\mathbb{P}^m\times\mathbb{P}^n)$ satisfy the given equations. Conversely, consider a point $z\in\mathbb{P}^N$ with homogeneous coordinates $z_{0,0},\dots,z_{m,n}$ that satisfy the given equations. At least one of these coordinates must be non-zero; we can assume without loss of generality that it is $z_{0,0}$. Let us pass to affine coordinates by setting $z_{0,0}=1$. Then we have $z_{i,j}=z_{i,0}z_{0,j}$ for all $i=0,\dots,m$ and $j=0,\dots,n$. Hence by setting $x_i=z_{i,0}$ and $y_j=z_{0,j}$ (in particular $x_0=y_0=1$) we obtain a point of $\mathbb{P}^m\times\mathbb{P}^n$ that is mapped to $z$ by $f$.

(b) Continuing the above notation, let $z\in X$ be a point with $z_{0,0}=1$. If $f(x,y)=z$ for some $(x,y)\in\mathbb{P}^m\times\mathbb{P}^n$, it follows that $x_0\neq 0$ and $y_0\neq 0$, so we can pass to affine coordinates here as well and assume that $x_0=1$ and $y_0=1$. But then it follows that $x_i=z_{i,0}$ and $y_j=z_{0,j}$ for all $i$ and $j$, i. e. $f$ is injective and thus as a map onto its image also bijective.

The same calculation shows that $f$ and $f^{-1}$ are given (locally in affine coordinates) by polynomial maps. Hence $f$ is an isomorphism. □

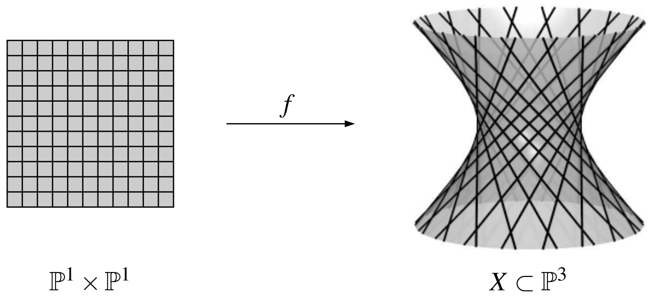
**Example 7.12.** According to Proposition 7.11, the product $\mathbb{P}^1 \times \mathbb{P}^1$ is (isomorphic to) the surface

$$X = \{(z_{0,0}:z_{0,1}:z_{1,0}:z_{1,1}) : z_{0,0}z_{1,1} = z_{1,0}z_{0,1}\} \quad \subset \mathbb{P}^3$$

by the isomorphism

$$f : \mathbb{P}^1 \times \mathbb{P}^1 \to X, \ \ ((x_0:x_1),(y_0:y_1)) \mapsto (x_0y_0:x_0y_1:x_1y_0:x_1y_1).$$

In particular, the "lines" $\{a\} \times \mathbb{P}^1$ and $\mathbb{P}^1 \times \{a\}$ in $\mathbb{P}^1 \times \mathbb{P}^1$ where the first or second factor is constant, respectively, are mapped to lines in $X \subset \mathbb{P}^3$. The following schematic picture shows these two families of lines on the surface $X$ (whose set of real points is a hyperboloid).




**Corollary 7.13.** *Every projective variety is a variety.*

*Proof.* We have already seen in proposition 7.2 that every projective variety is a prevariety. So by Lemma 5.20 (b) it only remains to be shown that $\mathbb{P}^n$ is separated, i. e. that the diagonal $\Delta_{\mathbb{P}^n}$ is closed in $\mathbb{P}^n \times \mathbb{P}^n$. We can describe this diagonal as

$$\Delta_{\mathbb{P}^n} = \{((x_0:\cdots:x_n),(y_0:\cdots:y_n)) : x_iy_j - x_jy_i = 0 \text{ for all } i,j\},$$

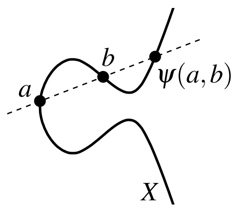
because these equations mean exactly that the matrix

$$\begin{pmatrix} x_0 & x_1 & \cdots & x_n \\ y_0 & y_1 & \cdots & y_n \end{pmatrix}$$

has rank (at most) 1, i. e. that $(x_0:\cdots:x_n) = (y_0:\cdots:y_n)$. In particular, it follows that $\Delta_{\mathbb{P}^n}$ is closed as the zero locus of the homogeneous linear polynomials $z_{i,j} - z_{j,i}$ in the Segre coordinates $z_{i,j} = x_iy_j$ of $\mathbb{P}^n \times \mathbb{P}^n$. □

**Remark 7.14.** If $X \subset \mathbb{P}^m$ and $Y \subset \mathbb{P}^n$ are projective varieties then $X \times Y$ is a closed subset of $\mathbb{P}^m \times \mathbb{P}^n$. As the latter is a projective variety by the Segre embedding we see that $X \times Y$ is a projective variety as well (namely a projective subvariety of $\mathbb{P}^m \times \mathbb{P}^n$).

**Exercise 7.15.** Let $X \subset \mathbb{P}^2$ be a curve given as the zero locus of a homogeneous polynomial of degree 3. Moreover, let $U \subset X \times X$ be the set of all $(a,b) \in X \times X$ such that $a \neq b$ and the unique line through $a$ and $b$ meets $X$ in exactly three distinct points. Of course, two of these points are then $a$ and $b$; we will denote the third one by $\psi(a,b) \in X$.




Show that $U \subset X \times X$ is open, and that $\psi : U \to X$ is a morphism.

**Exercise 7.16.**

(a) Prove that for every projective variety $Y \subset \mathbb{P}^n$ of pure dimension $n-1$ there is a homogeneous polynomial $f$ such that $I(Y) = (f)$. You may use the commutative algebra fact that every polynomial in $K[x_0,\dots,x_n]$ admits a unique decomposition into prime elements [G5, Remark 8.6].

(b) If $X$ is a projective variety of dimension $n$, show by example that in general not every projective variety $Y \subset X$ of dimension $n-1$ is of the form $V(f)$ for a homogeneous polynomial $f \in S(X)$. (One possibility is to consider the Segre embedding $X$ of $\mathbb{P}^1 \times \mathbb{P}^1$ in $\mathbb{P}^3$, and $Y = \mathbb{P}^1 \times \{0\} \subset \mathbb{P}^1 \times \mathbb{P}^1$.)

The most important property of projective varieties is that they are compact in the classical topology if the ground field is $\mathbb{C}$. We have seen this already for projective spaces in Remark 6.4, and it then follows for projective varieties as well since they are closed subsets of them. However, Exercises 2.34 (c) and 5.11 (a) show unfortunately that every prevariety is compact in the Zariski topology, and so in particular that compactness in the Zariski topology does not capture the same geometric idea as in the classical case. We therefore need an alternative description of the intuitive compactness property that works in our algebraic setting of the Zariski topology.

The key idea to achieve this is that compact sets should be mapped to compact sets again under continuous maps. In our language, this means that images of morphisms between projective varieties should be closed. This property (that we have already seen to be false for general varieties in Remark 5.15 (a)) is what we want to prove now. We start with a special case which contains all the hard work, and from which the general case will then follow easily.

**Definition 7.17** (Closed maps)**.** A map $f : X \to Y$ between topological spaces is called **closed** if $f(A) \subset Y$ is closed for every closed subset $A \subset X$.

**Proposition 7.18.** *The projection map $\pi : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^m$ is closed.*

*Proof.* Let $Z \subset \mathbb{P}^n \times \mathbb{P}^m$ be a closed set. By Remark 6.20 we can write $Z = V(f_1,\dots,f_r)$ for homogeneous polynomials $f_1,\dots,f_r$ of the same degree $d$ in the Segre coordinates of $\mathbb{P}^n \times \mathbb{P}^m$, i.e. for bihomogeneous polynomials of degree $d$ in both the coordinates $x_0,\dots,x_n$ of $\mathbb{P}^n$ and $y_0,\dots,y_m$ of $\mathbb{P}^m$. Now consider a fixed point $a \in \mathbb{P}^m$; we will determine if it is contained in the image $\pi(Z)$. To do this, let $g_i = f_i(\,\cdot\,,a) \in K[x_0,\dots,x_n]$ for $i = 1,\dots,r$. Then

$$\begin{aligned} a \notin \pi(Z) &\Leftrightarrow \text{there is no } x \in \mathbb{P}^n \text{ with } (x,a) \in Z \\ &\Leftrightarrow V_p(g_1,\dots,g_r) = \emptyset \\ &\Leftrightarrow \sqrt{(g_1,\dots,g_r)} = (1) \text{ or } \sqrt{(g_1,\dots,g_r)} = (x_0,\dots,x_n) \quad \text{(Proposition 6.22)} \\ &\Leftrightarrow \text{there are } k_i \in \mathbb{N} \text{ with } x_i^{k_i} \in (g_1,\dots,g_r) \text{ for all } i \\ &\Leftrightarrow K[x_0,\dots,x_n]_k \subset (g_1,\dots,g_r) \text{ for some } k \in \mathbb{N}, \end{aligned}$$

where as usual $K[x_0,\dots,x_n]_k$ denotes the homogeneous degree-$k$ part of the polynomial ring as in Definition 6.6, and the direction "$\Rightarrow$" of the last equivalence follows by setting $k = k_0 + \cdots + k_n$. Of course, the last condition can only be satisfied if $k \geq d$ and is equivalent to $K[x_0,\dots,x_n]_k = (g_1,\dots,g_r)_k$. As $(g_1,\dots,g_r) = \{h_1g_1 + \cdots + h_rg_r : h_1,\dots,h_r \in K[x_0,\dots,x_n]\}$ this is the same as saying that the $K$-linear map

$$F_k : (K[x_0,\dots,x_n]_{k-d})^r \to K[x_0,\dots,x_n]_k, \ \ (h_1,\dots,h_r) \mapsto h_1g_1 + \cdots h_rg_r$$

is surjective, i.e. has rank $\dim_K K[x_0,\dots,x_n]_k = \binom{n+k}{k}$ for some $k \geq d$. This in turn is the case if and only if at least one of the minors of size $\binom{n+k}{k}$ of a matrix for some $F_k$ is non-zero. But these minors are polynomials in the coefficients of $g$ and thus in the coordinates of $a$, and consequently the non-vanishing of one of these minors is an open condition in the Zariski topology of $\mathbb{P}^m$.

Hence the set of all $a \in \mathbb{P}^m$ with $a \notin \pi(Z)$ is open, which means that $\pi(Z)$ is closed. □

**Remark 7.19.** Let us look at Proposition 7.18 from an algebraic point of view. We start with some equations $f_1(x,y) = \cdots = f_r(x,y) = 0$ in two sets of variables $x = (x_0,\dots,x_n)$ and $y = (y_0,\dots,y_m)$ and ask for the image of their common zero locus under the projection map $(x,y) \mapsto x$. The equations satisfied on this image are precisely the equations in $x$ alone that can be derived from the given ones $f_1(x,y) = \cdots = f_r(x,y) = 0$ in $x$ and $y$. In other words, we want to *eliminate* the variables $y$ from the given system of equations. The statement of Proposition 7.18 is therefore sometimes called the *main theorem of elimination theory*.

**Corollary 7.20.** *The projection map $\pi : \mathbb{P}^n \times Y \to Y$ is closed for any variety $Y$.*

*Proof.* Let us first show the statement for an affine variety $Y \subset \mathbb{A}^m$. Then we can regard $Y$ as a locally closed subvariety of $\mathbb{P}^m$ via the embedding $\mathbb{A}^m \subset \mathbb{P}^m$. Now let $Z \subset \mathbb{P}^n \times Y$ be closed, and let

$\overline{Z}$ be its closure in $\mathbb{P}^n \times \mathbb{P}^m$. If $\pi : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^m$ is the projection map then $\pi(\overline{Z})$ is closed in $\mathbb{P}^m$ by Proposition 7.18, and thus

$$\pi(Z) = \pi(\overline{Z} \cap (\mathbb{P}^n \times Y)) = \pi(\overline{Z}) \cap Y$$

is closed in $Y$.

If $Y$ is any variety we can cover it by affine open subsets. As the condition that a subset is closed can be checked by restricting it to the elements of an open cover, the statement follows from the corresponding one for the affine open patches that we have just shown. □

It is in fact this property of Corollary 7.20 that captures the classical idea of compactness. Let us therefore give it a name:

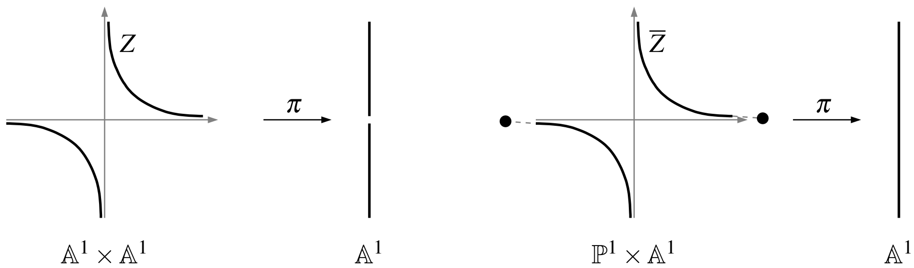
**Definition 7.21** (Complete varieties)**.** A variety $X$ is called **complete** if the projection map $\pi : X \times Y \to Y$ is closed for any variety $Y$.

**Example 7.22.**

(a) $\mathbb{P}^n$ is complete by Corollary 7.20.

(b) Any closed subvariety $X'$ of a complete variety $X$ is complete: if $Z \subset X' \times Y$ is closed then $Z$ is also closed in $X \times Y$, and hence its image under the second projection to $Y$ is closed as well. In particular, by (a) this means that every projective variety is complete.

(c) $\mathbb{A}^1$ is not complete: as in the picture below on the left, the image $\pi(Z)$ of the closed subset $Z = V(x_1x_2 - 1) \subset \mathbb{A}^1 \times \mathbb{A}^1$ under the second projection is $\mathbb{A}^1 \backslash \{0\}$, which is not closed.




The geometric reason for this is that $\mathbb{A}^1$ is missing a point at infinity: if we replace $\mathbb{A}^1$ by $\mathbb{P}^1$ as in the picture on the right there is an additional point in the closure $\overline{Z}$ of $Z \subset \mathbb{A}^1 \times \mathbb{A}^1$ in $\mathbb{P}^1 \times \mathbb{A}^1$; the image of this point under $\pi$ fills the gap and makes $\pi(\overline{Z})$ a closed set. Intuitively, one can think of the name "complete" as coming from the geometric idea that it contains all the "points at infinity" that are missing in affine varieties.

**Remark 7.23.** There are complete varieties that are not projective, but this is actually quite hard to show — we will certainly not meet such an example in this course. So for practical purposes you can usually assume that the terms "projective variety" and "complete variety" are synonymous.

In any case, complete varieties now have the property that we were aiming for:

**Corollary 7.24.** *Let $f : X \to Y$ be a morphism of varieties. If $X$ is complete then its image $f(X)$ is closed in $Y$.*

*Proof.* By Proposition 5.21 (a) the graph $\Gamma_f \subset X \times Y$ is closed. But then $f(X) = \pi(\Gamma_f)$ for the projection map $\pi : X \times Y \to Y$, which is closed again since $X$ is complete. □

Let us conclude this chapter with two applications of this property.

**Corollary 7.25.** *Let $X$ be a connected complete variety. Then $\mathscr{O}_X(X) = K$, i. e. every global regular function on $X$ is constant.*

*Proof.* A global regular function $\varphi \in \mathscr{O}_X(X)$ determines a morphism $\varphi : X \to \mathbb{A}^1$. By extension of the target we can consider this as a morphism $\varphi : X \to \mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$ whose image $\varphi(X) \subset \mathbb{P}^1$ does not contain the point $\infty$. But $\varphi(X)$ is also closed by Corollary 7.24 since $X$ is complete, and hence it must be a finite set since these are the only closed proper subsets of $\mathbb{P}^1$. Moreover, Exercise 2.21 (b) implies that $\varphi(X)$ is connected since $X$ is. Altogether this means that $\varphi(X)$ is a single point, i. e. that $\varphi$ is constant. □

**Remark 7.26.** Corollary 7.25 is another instance of a result that has a counterpart in complex analysis: it can be shown that every holomorphic function on a connected compact complex manifold is
12 constant.

**Construction 7.27** (Veronese embedding)**.** Choose $n,d \in \mathbb{N}_{>0}$, and let $f_0,\dots,f_N \in K[x_0,\dots,x_n]$ for $N = \binom{n+d}{n} - 1$ be the set of all monomials of degree $d$ in the variables $x_0,\dots,x_n$, in any order. Consider the map

$$F : \mathbb{P}^n \to \mathbb{P}^N, \;\; x \mapsto (f_0(x) : \cdots : f_N(x)).$$

By Lemma 7.5 this is a morphism (note that the monomials $x_0^d,\dots,x_n^d$, which cannot be simultaneously zero, are among the $f_0,\dots,f_N$). So by Corollary 7.24 the image $X = F(\mathbb{P}^n)$ is a projective variety.

We claim that $F : \mathbb{P}^n \to X$ is an isomorphism. All we have to do to prove this is to find an inverse morphism. This is not hard: we can do this on an affine open cover, so let us e. g. consider the open subset where $x_0 \neq 0$, i. e. $x_0^d \neq 0$. On this set we can pass to affine coordinates and set $x_0 = 1$. The inverse morphism is then given by $x_i = \frac{x_i x_0^{d-1}}{x_0^d}$ for $i = 1,\dots,n$, which is a quotient of two degree-$d$ monomials.

The morphism $F$ is therefore an isomorphism and thus realizes $\mathbb{P}^n$ as a subvariety $X$ of $\mathbb{P}^N$. It is usually called the degree-$d$ **Veronese embedding**; the coordinates on $\mathbb{P}^N$ are called **Veronese coordinates** of $\mathbb{P}^n \cong X$. Of course, this embedding can also be restricted to any projective variety $Y \subset \mathbb{P}^n$ and then gives an isomorphism by degree-$d$ polynomials between $Y$ and a projective variety in $\mathbb{P}^N$.

The importance of the Veronese embedding lies in the fact that degree-$d$ polynomials in the coordinates of $\mathbb{P}^n$ are translated into *linear* polynomials in the Veronese coordinates. An example where this is useful will be given in Corollary 7.30.

**Example 7.28.**

(a) For $d = 1$ the Veronese embedding of $\mathbb{P}^n$ is just the identity $\mathbb{P}^n \to \mathbb{P}^n$.

(b) For $n = 1$ the degree-$d$ Veronese embedding of $\mathbb{P}^1$ in $\mathbb{P}^d$ is

$$F : \mathbb{P}^1 \to \mathbb{P}^d, \;\; (x_0 : x_1) \mapsto (x_0^d : x_0^{d-1}x_1 : \cdots : x_0 x_1^{d-1} : x_1^d).$$

In the $d = 2$ case we have already seen in Example 7.6 (c) that this is an isomorphism.

**Exercise 7.29.** Let $F : \mathbb{P}^n \to \mathbb{P}^N$ be the degree-$d$ Veronese embedding as in Construction 7.27, with $N = \binom{n+d}{n} - 1$. By applying Corollary 7.24 we have seen already that the image $X = F(\mathbb{P}^n)$ is a projective variety. Find explicit equations describing $X$, i. e. generators for a homogeneous ideal $I$ such that $X = V(I)$.

**Corollary 7.30.** *Let $X \subset \mathbb{P}^n$ be a projective variety, and let $f \in S(X)$ be a non-zero homogeneous element. Then $X \backslash V(f)$ is an affine variety.*

*Proof.* If $f = x_0$ this is just Proposition 7.2. For a general linear polynomial $f$ the statement follows from this after a projective automorphism as in Example 7.6 (a) that takes $f$ to $x_0$, and if $f$ is of degree $d > 1$ we can reduce the claim to the linear case by first applying the degree-$d$ Veronese embedding of Construction 7.27. □

**Exercise 7.31.** Recall from Example 7.6 (d) that a conic in $\mathbb{P}^2$ over a field of characteristic not equal to 2 is the zero locus of an irreducible homogeneous polynomial of degree 2 in $K[x_0,x_1,x_2]$.

(a) Considering the coefficients of such polynomials, show that the set of all conics in $\mathbb{P}^2$ can be identified with an open subset $U$ of the projective space $\mathbb{P}^5$.

(b) Let $a \in \mathbb{P}^2$. Show that the subset of $U$ consisting of all conics passing through $a$ is the zero locus of a linear equation in the homogeneous coordinates of $U \subset \mathbb{P}^5$.

(c) Given 5 points in $\mathbb{P}^2$, no three of which lie on a line, show that there is a unique conic in $\mathbb{P}^5$ passing through all these points.

**Exercise 7.32.** Let $X \subset \mathbb{P}^3$ be the degree-3 Veronese embedding of $\mathbb{P}^1$, i. e. the image of the morphism

$$\mathbb{P}^1 \to \mathbb{P}^3, \ (x_0 : x_1) \mapsto (y_0 : y_1 : y_2 : y_3) = (x_0^3 : x_0^2 x_1 : x_0 x_1^2 : x_1^3).$$

Moreover, let $a = (0:0:1:0) \in \mathbb{P}^3$ and $L = V(y_2) \subset \mathbb{P}^3$, and consider the projection $f$ from $a$ to $L$ as in Example 7.6 (b).

(a) Show that $f$ is a morphism.

(b) Determine an equation of the curve $f(X)$ in $L \cong \mathbb{P}^2$.

(c) Is $f : X \to f(X)$ an isomorphism onto its image?

# 8. Grassmannians

After having introduced (projective) varieties — the main objects of study in algebraic geometry — let us now take a break in our discussion of the general theory to construct an interesting and useful class of examples of projective varieties. The idea behind this construction is simple: since the definition of projective spaces as the sets of 1-dimensional linear subspaces of $K^n$ turned out to be a very useful concept, let us now generalize this and consider instead the sets of $k$-dimensional linear subspaces of $K^n$ for an arbitrary $k = 0, \dots, n$.

**Definition 8.1** (Grassmannians)**.** Let $n \in \mathbb{N}_{>0}$, and let $k \in \mathbb{N}$ with $0 \le k \le n$. We denote by $G(k,n)$ the set of all $k$-dimensional linear subspaces of $K^n$. It is called the **Grassmannian** of $k$-planes in $K^n$.

**Remark 8.2.** By Example 6.12 (b) and Exercise 6.32 (a), the correspondence of Remark 6.17 shows that $k$-dimensional linear subspaces of $K^n$ are in natural one-to-one correspondence with $(k-1)$-dimensional linear subspaces of $\mathbb{P}^{n-1}$. We can therefore consider $G(k,n)$ alternatively as the set of such projective linear subspaces. As the dimensions $k$ and $n$ are reduced by 1 in this way, our Grassmannian $G(k,n)$ of Definition 8.1 is sometimes written in the literature as $G(k-1,n-1)$ instead.

Of course, as in the case of projective spaces our goal must again be to make the Grassmannian $G(k,n)$ into a variety — in fact, we will see that it is even a projective variety in a natural way. For this we need the algebraic concept of *alternating tensor products*, a kind of multilinear product on $K^n$ generalizing the well-known *cross product*

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}$$

on $K^3$ whose coordinates are all the $2 \times 2$ minors of the matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}.$$

If you have seen ordinary tensor products in commutative algebra already [G5, Chapter 5], you probably know that the best way to introduce these products is by a universal property similar to the one for products of varieties in Definition 5.16. Although the same is true for our alternating tensor products, we will follow a faster and more basic approach here, whose main disadvantage is that it is not coordinate-free. Of course, if you happen to know the "better" definition of alternating tensor products using their universal property already, you can use this definition as well and skip the following construction.

**Construction 8.3** (Alternating tensor products)**.** Let $(e_1, \dots, e_n)$ denote the standard basis of $K^n$. For $k \in \mathbb{N}$ we define $\Lambda^k K^n$ to be a $K$-vector space of dimension $\binom{n}{k}$ with basis vectors formally written as

$$e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k} \qquad (*)$$

for all multi-indices $(i_1, \dots, i_k)$ of natural numbers with $1 \le i_1 < i_2 < \cdots < i_k \le n$. Note that the set of these strictly increasing multi-indices is in natural bijection with the set of all $k$-element subsets $\{i_1, \dots, i_k\}$ of $\{1, \dots, n\}$, so that there are in fact exactly $\binom{n}{k}$ of these basis vectors. In particular, $\Lambda^k K^n$ is the zero vector space if $k > n$.

We extend the notation $(*)$ to arbitrary (i. e. not strictly increasing) multi-indices $(i_1, \dots, i_k)$ with $1 \le i_1, \dots, i_k \le n$ by setting $e_{i_1} \wedge \cdots \wedge e_{i_k} := 0$ if any two of the $i_1, \dots, i_k$ coincide, and

$$e_{i_1} \wedge \cdots \wedge e_{i_k} := \operatorname{sign} \sigma \cdot e_{i_{\sigma(1)}} \wedge \cdots \wedge e_{i_{\sigma(k)}}$$

if all $i_1,\dots,i_k$ are distinct, and $\sigma$ is the unique permutation of $\{1,\dots,k\}$ such that $i_{\sigma(1)} < \cdots < i_{\sigma(k)}$. We can then extend this notation multilinearly to a product $(K^n)^k \to \Lambda^k K^n$: for $v_1,\dots,v_k \in K^n$ with basis expansions $v_j = \sum_{i=1}^n a_{j,i}\, e_i$ for some $a_{j,i} \in K$ we define

$$v_1 \wedge \cdots \wedge v_k := \sum_{i_1,\dots,i_k} a_{1,i_1} \cdots a_{k,i_k} \cdot e_{i_1} \wedge \cdots \wedge e_{i_k} \quad \in \Lambda^k K^n.$$

More generally, we obtain bilinear and associative products $\Lambda^k K^n \times \Lambda^l K^n \to \Lambda^{k+l} K^n$ by a bilinear extension of

$$(e_{i_1} \wedge \cdots \wedge e_{i_k}) \wedge (e_{j_1} \wedge \cdots \wedge e_{j_l}) := e_{i_1} \wedge \cdots \wedge e_{i_k} \wedge e_{j_1} \wedge \cdots \wedge e_{j_l}.$$

The vector space $\Lambda^k K^n$ is usually called the $k$-fold **alternating** or **antisymmetric tensor product** of $K^n$, the elements of $\Lambda^k K^n$ are referred to as **alternating** or **antisymmetric tensors**.

**Example 8.4.**

(a) By definition we have $\Lambda^0 = K$ and $\Lambda^1 K^n = K^n$; a basis of $\Lambda^1 K^n$ is again $(e_1,\dots,e_n)$. We also have $\Lambda^n K^n \cong K$, with single basis vector $e_1 \wedge \cdots \wedge e_n$.

(b) As in (a), $\Lambda^2 K^2$ is isomorphic to $K$ with basis vector $e_1 \wedge e_2$. For two arbitrary vectors $v = a_1 e_1 + a_2 e_2$ and $w = b_1 e_1 + b_2 e_2$ of $K^2$ their alternating tensor product is

$$\begin{aligned} v \wedge w &= a_1 b_1\, e_1 \wedge e_1 + a_1 b_2\, e_1 \wedge e_2 + a_2 b_1\, e_2 \wedge e_1 + a_2 b_2\, e_2 \wedge e_2 \\ &= (a_1 b_2 - a_2 b_1)\, e_1 \wedge e_2, \end{aligned}$$

so under the isomorphism $\Lambda^2 K^2 \cong K$ it is just the determinant of the coefficient matrix of $v$ and $w$.

(c) Similarly, for $v = a_1 e_1 + a_2 e_2 + a_3 e_3$ and $w = b_1 e_1 + b_2 e_2 + b_3 e_3$ in $K^3$ we have

$$v \wedge w = (a_1 b_2 - b_2 a_1)\, e_1 \wedge e_2 + (a_1 b_3 - b_3 a_1)\, e_1 \wedge e_3 + (a_2 b_3 - a_3 b_2)\, e_2 \wedge e_3 \quad \in \Lambda^2 K^3 \cong K^3,$$

so (up to a simple change of basis) $v \wedge w$ is just the cross product $v \times w$ considered in the introduction to this chapter.

As we will see now, it is in fact a general phenomenon that the coordinates of alternating tensor products can be interpreted as determinants.

**Remark 8.5** (Alternating tensor products and determinants)**.** Let $0 \le k \le n$, and let $v_1,\dots,v_k \in K^n$ with basis expansions $v_j = \sum_i a_{j,i} e_i$ for $j = 1,\dots,k$. For a strictly increasing multi-index $(j_1,\dots,j_k)$ let us determine the coefficient of the basis vector $e_{j_1} \wedge \cdots \wedge e_{j_k}$ in the tensor product $v_1 \wedge \cdots \wedge v_k$. As in Construction 8.3 we have

$$v_1 \wedge \cdots \wedge v_k = \sum_{i_1,\dots,i_k} a_{1,i_1} \cdots a_{k,i_k} \cdot e_{i_1} \wedge \cdots \wedge e_{i_k}.$$

Note that the indices $i_1,\dots,i_k$ in the products $e_{i_1} \wedge \cdots \wedge e_{i_k}$ in the terms of this sum are not necessarily in strictly ascending order. So to figure out the coefficient of $e_{j_1} \wedge \cdots \wedge e_{j_k}$ in $v_1 \wedge \cdots \wedge v_k$ we have to sort the indices in each sum first; the resulting coefficient is then

$$\sum \operatorname{sign} \sigma \cdot a_{1,j_{\sigma(1)}} \cdots a_{k,j_{\sigma(k)}},$$

where the sum is taken over all permutations $\sigma$. By definition this is exactly the determinant of the maximal quadratic submatrix of the coefficient matrix $(a_{i,j})_{i,j}$ obtained by taking only the columns $j_1,\dots,j_k$. In other words, the coordinates of $v_1 \wedge \cdots \wedge v_k$ are just all the maximal minors of the matrix whose rows are $v_1,\dots,v_k$. So the alternating tensor product can be viewed as a convenient way to encode all these minors in a single object.

As a consequence, alternating tensor products can be used to encode the linear dependence and linear spans of vectors in a very elegant way.

**Lemma 8.6.** *Let $v_1,\dots,v_k \in K^n$ for some $k \le n$. Then $v_1 \wedge \cdots \wedge v_k = 0$ if and only if $v_1,\dots,v_k$ are linearly dependent.*

*Proof.* By Remark 8.5, we have $v_1 \wedge \cdots \wedge v_k = 0$ if and only if all maximal minors of the matrix with rows $v_1,\dots,v_k$ are zero. But this is the case if and only if this matrix does not have full rank [G2, Exercise 18.25], i. e. if and only if $v_1,\dots,v_k$ are linearly dependent. □

**Remark 8.7.**

(a) By construction, the alternating tensor product is antisymmetric in the sense that for all $v_1,\dots,v_k \in K^n$ and all permutations $\sigma$ we have

$$v_1 \wedge \cdots \wedge v_k = \operatorname{sign}\sigma \cdot v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(k)}.$$

Moreover, Lemma 8.6 tells us that $v_1 \wedge \cdots \wedge v_k = 0$ if two of the vectors $v_1,\dots,v_k$ coincide.

(b) We have constructed the alternating tensor product using a fixed basis $e_1,\dots,e_n$ of $K^n$. However, if $v_1,\dots,v_n$ is an arbitrary basis of $K^n$ it is easy to see that the alternating tensors $v_{i_1} \wedge \cdots \wedge v_{i_k}$ for strictly increasing multi-indices $(i_1,\dots,i_k)$ form a basis of $\Lambda^k K^n$ as well: there are $\binom{n}{k}$ of these vectors, and they generate $\Lambda^k K^n$ since every standard unit vector $e_i$ is a linear combination of $v_1,\dots,v_n$, and hence every $k$-fold alternating product $e_{i_1} \wedge \cdots \wedge e_{i_k}$ is a linear combination of $k$-fold alternating products of $v_1,\dots,v_n$ — which can be expressed by (a) in terms of such products with strictly increasing indices.

**Lemma 8.8.** *Let $v_1,\dots,v_k \in K^n$ and $w_1,\dots,w_k \in K^n$ both be linearly independent. Then $v_1 \wedge \cdots \wedge v_k$ and $w_1 \wedge \cdots \wedge w_k$ are linearly dependent in $\Lambda^k K^n$ if and only if* $\operatorname{Lin}(v_1,\dots,v_k) = \operatorname{Lin}(w_1,\dots,w_k)$.

*Proof.* As we have assumed both $v_1,\dots,v_k$ and $w_1,\dots,w_k$ to be linearly independent, we know by Lemma 8.6 that $v_1 \wedge \cdots \wedge v_k$ and $w_1 \wedge \cdots \wedge w_k$ are both non-zero.

"⇒" Assume that $\operatorname{Lin}(v_1,\dots,v_k) \neq \operatorname{Lin}(w_1,\dots,w_k)$, so without loss of generality that $w_1 \notin \operatorname{Lin}(v_1,\dots,v_k)$. Then $w_1,v_1,\dots,v_k$ are linearly independent, and thus $w_1 \wedge v_1 \wedge \cdots \wedge v_k \neq 0$ by Lemma 8.6. But by assumption we know that $v_1 \wedge \cdots \wedge v_k = \lambda\, w_1 \wedge \cdots \wedge w_k$ for some $\lambda \in K$, and hence

$$0 \neq w_1 \wedge v_1 \wedge \cdots \wedge v_k = \lambda\, w_1 \wedge w_1 \wedge \cdots \wedge w_k$$

in contradiction to Remark 8.7 (a).

"⇐" If $v_1,\dots,v_k$ and $w_1,\dots,w_k$ span the same subspace of $K^n$ then the basis $w_1,\dots,w_k$ of this subspace can be obtained from $v_1,\dots,v_k$ by a finite sequence of basis exchange operations $v_i \to v_i + \lambda\, v_j$ and $v_i \to \lambda\, v_i$ for $\lambda \in K$ and $i \neq j$. But both these operations change the alternating product of the vectors at most by a multiplicative scalar, since

$$v_1 \wedge \cdots \wedge v_{i-1} \wedge (v_i + \lambda\, v_j) \wedge v_{i+1} \wedge \cdots \wedge v_n = v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_n$$

and

$$v_1 \wedge \cdots \wedge (\lambda\, v_i) \wedge \cdots \wedge v_n = \lambda\, v_1 \wedge \cdots \wedge v_n$$

by multilinearity and Remark 8.7 (a). □

We can now use our results to realize the Grassmannian $G(k,n)$ as a subset of a projective space.

**Construction 8.9** (Plücker embedding)**.** Let $0 \le k \le n$, and consider the map $f : G(k,n) \to \mathbb{P}^{\binom{n}{k}-1}$ given by sending a linear subspace $\operatorname{Lin}(v_1,\dots,v_k) \in G(k,n)$ to the class of $v_1 \wedge \cdots \wedge v_k \in \Lambda^k K^n \cong K^{\binom{n}{k}}$ in $\mathbb{P}^{\binom{n}{k}-1}$. Note that this is well-defined: $v_1 \wedge \cdots \wedge v_k$ is non-zero by Lemma 8.6, and representing the same subspace by a different basis does not change the resulting point in $\mathbb{P}^{\binom{n}{k}-1}$ by the part "⇐" of Lemma 8.8. Moreover, the map $f$ is injective by the part "⇒" of Lemma 8.8. We call it the **Plücker embedding** of $G(k,n)$; for a $k$-dimensional linear subspace $L \in G(k,n)$ the (homogeneous) coordinates of $f(L)$ in $\mathbb{P}^{\binom{n}{k}-1}$ are the **Plücker coordinates** of $L$. By Remark 8.5, they are just all the maximal minors of the matrix whose rows are $v_1,\dots,v_k$.

In the following, we will always consider $G(k,n)$ as a subset of $\mathbb{P}^{\binom{n}{k}-1}$ using this Plücker embedding.

**Example 8.10.**

(a) The Plücker embedding of $G(1,n)$ simply maps a linear subspace $\operatorname{Lin}(a_1e_1 + \cdots + a_ne_n)$ to the point $(a_1 : \cdots : a_n) \in \mathbb{P}^{\binom{n}{1}-1} = \mathbb{P}^{n-1}$. Hence $G(1,n) = \mathbb{P}^{n-1}$ as expected.

(b) Consider the 2-dimensional subspace $L = \operatorname{Lin}(e_1+e_2, e_1+e_3) \in G(2,3)$ of $K^3$. As

$$(e_1+e_2)\wedge(e_1+e_3) = -e_1\wedge e_2 + e_1\wedge e_3 + e_2\wedge e_3,$$

the coefficients $(-1:1:1)$ of this vector are the Plücker coordinates of $L$ in $\mathbb{P}^{\binom{3}{2}-1} = \mathbb{P}^2$. Alternatively, these are the three maximal minors of the matrix

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

whose rows are the given spanning vectors $e_1+e_2$ and $e_1+e_3$ of $L$. Note that a change of these spanning vectors will just perform row operations on this matrix, which changes the maximal minors at most by a common constant factor. This shows again in this example that the homogeneous Plücker coordinates of $L$ are well-defined.

13

So far we have embedded the Grassmannian $G(k,n)$ into a projective space, but we still have to see that it is a closed subset, i. e. a projective variety. So by Construction 8.9 we have to find suitable equations describing the alternating tensors in $\Lambda^k K^n$ that can be written as a so-called *pure tensor*, i. e. as $v_1\wedge\cdots\wedge v_k$ for some $v_1,\dots,v_k \in K^n$ — and not just as a linear combination of such expressions. The key lemma to achieve this is the following.

**Lemma 8.11.** *For a fixed non-zero $\omega \in \Lambda^k K^n$ with $k<n$ consider the $K$-linear map*

$$f: K^n \to \Lambda^{k+1}K^n,\ \ v \mapsto v\wedge\omega.$$

*Then* $\operatorname{rk} f \ge n-k$*, with equality holding if and only if $\omega = v_1\wedge\cdots\wedge v_k$ for some $v_1,\dots,v_k \in K^n$.*

**Example 8.12.** Let $k=2$ and $n=4$.

(a) For $\omega = e_1\wedge e_2$ the map $f$ of Lemma 8.11 is given by

$$\begin{aligned} f(a_1e_1+a_2e_2+a_3e_3+a_4e_4) &= (a_1e_1+a_2e_2+a_3e_3+a_4e_4)\wedge e_1\wedge e_2 \\ &= a_3e_1\wedge e_2\wedge e_3 + a_4e_1\wedge e_2\wedge e_4, \end{aligned}$$

for $a_1,a_2,a_3,a_4 \in K$, and thus has rank $\operatorname{rk} f = 2 = n-k$ in accordance with the statement of the lemma.

(b) For $\omega = e_1\wedge e_2 + e_3\wedge e_4$ we get

$$\begin{aligned} &f(a_1e_1+a_2e_2+a_3e_3+a_4e_4) \\ &\quad = (a_1e_1+a_2e_2+a_3e_3+a_4e_4)\wedge(e_1\wedge e_2+e_3\wedge e_4) \\ &\quad = a_1e_1\wedge e_3\wedge e_4 + a_2e_2\wedge e_3\wedge e_4 + a_3e_1\wedge e_2\wedge e_3 + a_4e_1\wedge e_2\wedge e_4 \end{aligned}$$

instead, so that $\operatorname{rk} f = 4$. Hence Lemma 8.11 tells us that there is no way to write $\omega$ as a pure tensor $v_1\wedge v_2$ for some vectors $v_1,v_2 \in K^4$.

*Proof of Lemma 8.11.* Let $v_1,\dots,v_r$ be a basis of $\ker f$ (with $r = n-\operatorname{rk} f$), and extend it to a basis $v_1,\dots,v_n$ of $K^n$. By Remark 8.7 (b) the alternating tensors $v_{i_1}\wedge\cdots\wedge v_{i_k}$ with $1\le i_1<\cdots<i_k\le n$ then form a basis of $\Lambda^k K^n$, and so we can write

$$\omega = \sum_{i_1<\cdots<i_k} a_{i_1,\dots,i_k}\, v_{i_1}\wedge\cdots\wedge v_{i_k}$$

for suitable coefficients $a_{i_1,\dots,i_k} \in K$. Now for $i=1,\dots,r$ we know that $v_i \in \ker f$, and thus

$$0 = v_i\wedge\omega = \sum_{i_1<\cdots<i_k} a_{i_1,\dots,i_k}\, v_i\wedge v_{i_1}\wedge\cdots\wedge v_{i_k}. \tag{$*$}$$

Note that $v_i\wedge v_{i_1}\wedge\cdots\wedge v_{i_k} = 0$ if $i\in\{i_1,\dots,i_k\}$, and in the other cases these products are (up to sign) different basis vectors of $\Lambda^{k+1}K^n$. So the equation $(*)$ tells us that we must have $a_{i_1,\dots,i_k} = 0$ whenever $i\notin\{i_1,\dots,i_k\}$. As this holds for all $i=1,\dots,r$ we conclude that the coefficient $a_{i_1,\dots,i_k} = 0$ can only be non-zero if $\{1,\dots,r\}\subset\{i_1,\dots,i_k\}$.

But at least one of these coefficients has to be non-zero since $\omega \neq 0$ by assumption. This obviously requires that $r \leq k$, i.e. that $\operatorname{rk} f = n - r \geq n - k$. Moreover, if we have equality then only the coefficient $a_{1,\dots,k}$ can be non-zero, which means that $\omega$ is a scalar multiple of $v_1 \wedge \cdots \wedge v_k$.

Conversely, if $\omega = w_1 \wedge \cdots \wedge w_k$ for some (necessarily linearly independent) $w_1, \dots, w_k \in K^n$ then $w_1, \dots, w_k \in \ker f$. Hence in this case $\dim \ker f \geq k$, i.e. $\operatorname{rk} f \leq n-k$, and together with the above result $\operatorname{rk} f \geq n-k$ we have equality. □

**Corollary 8.13** ($G(k,n)$ as a projective variety)**.** *With the Plücker embedding of Construction 8.9, the Grassmannian $G(k,n)$ is a closed subset of $\mathbb{P}^{\binom{n}{k}-1}$. In particular, it is a projective variety.*

*Proof.* As $G(n,n)$ is just a single point (and hence clearly a variety) we may assume that $k < n$. Then by construction a point $\omega \in \mathbb{P}^{\binom{n}{k}-1}$ lies in $G(k,n)$ if and only if it is the class of a pure tensor $v_1 \wedge \cdots \wedge v_k$. Lemma 8.11 shows that this is the case if and only if the rank of the linear map $f: K^n \to \Lambda^{k+1} K^n$, $v \mapsto v \wedge \omega$ is $n-k$. As we also know that the rank of this map is always at least $n-k$, this condition can be checked by the vanishing of all $(n-k+1) \times (n-k+1)$ minors of the matrix corresponding to $f$ [G2, Exercise 18.25]. But these minors are polynomials in the entries of this matrix, and thus in the coordinates of $\omega$. Hence we see that the condition for $\omega$ to be in $G(k,n)$ is closed. □

**Example 8.14.** By the proof of Corollary 8.13, the Grassmannian $G(2,4)$ is given by the vanishing of all sixteen $3 \times 3$ minors of a $4 \times 4$ matrix corresponding to a linear map $K^4 \to \Lambda^3 K^4$, i.e. it is a subset of $\mathbb{P}^{\binom{4}{2}-1} = \mathbb{P}^5$ given by 16 cubic equations.

As you might expect, this is by no means the simplest set of equations describing $G(2,4)$ — in fact, we will see in Exercise 8.19 (a) that a single quadratic equation suffices to cut out $G(2,4)$ from $\mathbb{P}^5$. Our proof of Corollary 8.13 is just the easiest way to show that $G(k,n)$ is a variety; it is not suitable in practice to find a nice description of $G(k,n)$ as a zero locus of simple equations.

However, there is another useful description of the Grassmannian in terms of affine patches, as we will see now. This will then also allow us to easily read off the dimension of $G(k,n)$ — which would be very hard to compute from its equations as in Corollary 8.13.

**Construction 8.15** (Affine cover of the Grassmannian)**.** Let $U_0 \subset G(k,n) \subset \mathbb{P}^{\binom{n}{k}-1}$ be the affine open subset where the $e_1 \wedge \cdots \wedge e_k$-coordinate is non-zero. Then by Remark 8.5 a linear subspace $L = \operatorname{Lin}(v_1, \dots, v_k) \in G(k,n)$ is in $U_0$ if and only if the $k \times n$ matrix $A$ with rows $v_1, \dots, v_k$ is of the form $A = (B \,|\, C)$ for an invertible $k \times k$ matrix $B$ and an arbitrary $k \times (n-k)$ matrix $C$. This in turn is the case if and only if $A$ is equivalent by row transformations, i.e. by a change of basis for $L$, to a matrix of the form $(E_k \,|\, D)$, where $E_k$ denotes the $k \times k$ unit matrix and $D \in \operatorname{Mat}(k \times (n-k), K)$: namely by multiplying $A$ with $B^{-1}$ from the left to obtain $(E_k \,|\, D)$ with $D = B^{-1}C$. Note that this is in fact the only choice for $D$, so that we get a bijection

$$\begin{aligned} f: \mathbb{A}^{k(n-k)} = \operatorname{Mat}(k \times (n-k), K) &\to U_0, \\ D &\mapsto \text{the linear subspace spanned by the rows of } (E_k \,|\, D). \end{aligned}$$

As the Plücker coordinates of this subspace, i.e. the maximal minors of $(E_k \,|\, D)$, are clearly polynomial functions in the entries of $D$, we see that $f$ is a morphism. Conversely, the $(i,j)$-entry of $D$ can be reconstructed (up to sign) from $f(D)$ as the maximal minor of $(E_k \,|\, D)$ where we take all columns of $E_k$ except the $i$-th, together with the $j$-th column of $D$. Hence $f^{-1}$ is a morphism as well, showing that $f$ is an isomorphism and thus $U_0 \cong \mathbb{A}^{k(n-k)}$ is an affine *space* (and not just an affine *variety*, which is already clear from Proposition 7.2).

Of course, this argument holds in the same way for all other affine patches where one of the Plücker coordinates is non-zero. Hence we conclude:

**Corollary 8.16.** *$G(k,n)$ is an irreducible variety of dimension $k(n-k)$.*

*Proof.* We have just seen in Construction 8.15 that $G(k,n)$ has an open cover by affine spaces $\mathbb{A}^{k(n-k)}$. As any two of these patches have a non-empty intersection (it is in fact easy to write down a $k \times n$ matrix such that any two given maximal minors are non-zero), the result follows from Exercises 2.20 (b) and 2.33 (a). □

**Remark 8.17.** The argument of Construction 8.15 also shows an alternative description of the Grassmannian: it is the space of all full-rank $k \times n$ matrices modulo row transformations. As we know that every such matrix is equivalent modulo row transformations to a unique matrix in reduced row echelon form, we can also think of $G(k,n)$ as the set of full-rank $k \times n$ matrices in such a form. For example, in the case $k = 1$ and $n = 2$ (when $G(1,2) = \mathbb{P}^1$ by Example 8.10 (a)) the full-rank $1 \times 2$ matrices in reduced row echelon form are

$$\begin{array}{rll} & (\,1 \quad *\,) & \text{corresponding to } \mathbb{A}^1 \subset \mathbb{P}^1 \\ \text{and} & (\,0 \quad 1\,) & \text{corresponding to } \infty \in \mathbb{P}^1 \end{array}$$

as in the homogeneous coordinates of $\mathbb{P}^1$.

The affine cover of Construction 8.15 can also be used to show the following symmetry property of the Grassmannians.

**Proposition 8.18.** *For all $0 \le k \le n$ we have $G(k,n) \cong G(n-k,n)$.*

*Proof.* There is an obvious well-defined set-theoretic bijection $f : G(k,n) \to G(n-k,n)$ that sends a $k$-dimensional linear subspace $L$ of $K^n$ to its "orthogonal" complement

$$L^\perp = \{x \in K^n : \langle x,y \rangle = 0 \text{ for all } y \in L\},$$

where $\langle x,y \rangle = \sum_{i=1}^n x_i y_i$ denotes the standard bilinear form. It remains to be shown that $f$ (and analogously $f^{-1}$) is a morphism. By Lemma 4.6, we can do this on the affine coordinates of Construction 8.15. So let $L \in G(k,n)$ be described as the subspace spanned by the rows of a matrix $(E_k \,|\, D)$, where the entries of $D \in \mathrm{Mat}(k \times (n-k), K)$ are the affine coordinates of $L$. As

$$(E_k \,|\, D) \cdot \begin{pmatrix} -D \\ E_{n-k} \end{pmatrix} = 0,$$

we see that $L^\perp$ is the subspace spanned by the rows of $(-D^T \,|\, E_{n-k})$. But the maximal minors of this matrix, i. e. the Plücker coordinates of $L^\perp$, are clearly polynomials in the entries of $D$, and thus we conclude that $f$ is a morphism. □

**Exercise 8.19.** Let $G(2,4) \subset \mathbb{P}^5$ be the Grassmannian of lines in $\mathbb{P}^3$ (or of 2-dimensional linear subspaces of $K^4$). We denote the homogeneous Plücker coordinates of $G(2,4)$ in $\mathbb{P}^5$ by $x_{i,j}$ for $1 \le i < j \le 4$. Show:

(a) $G(2,4) = V(x_{1,2}x_{3,4} - x_{1,3}x_{2,4} + x_{1,4}x_{2,3})$.

(b) Let $L \subset \mathbb{P}^3$ be an arbitrary line. Show that the set of lines in $\mathbb{P}^3$ that intersect $L$, considered as a subset of $G(2,4) \subset \mathbb{P}^5$, is the zero locus of a homogeneous linear polynomial.

How many lines in $\mathbb{P}^3$ would you expect to intersect four general given lines?

**Exercise 8.20.** Show that the following sets are projective varieties:

(a) the *incidence correspondence*

$$\{(L,a) \in G(k,n) \times \mathbb{P}^{n-1} : L \subset \mathbb{P}^{n-1} \text{ a } (k-1)\text{-dimensional linear subspace and } a \in L\};$$

(b) the *join* of two disjoint varieties $X, Y \subset \mathbb{P}^n$, i. e. the union in $\mathbb{P}^n$ of all lines intersecting both $X$ and $Y$.

# 9. Birational Maps and Blowing Up

In the course of this class we have already seen many examples of varieties that are "almost the same" in the sense that they contain isomorphic dense open subsets (although the varieties are not isomorphic themselves). Let us quickly recall some of them.

**Example 9.1** (Irreducible varieties with isomorphic non-empty open subsets)**.**

(a) The affine space $\mathbb{A}^n$ and the projective space $\mathbb{P}^n$ have the common open subset $\mathbb{A}^n$ by Proposition 7.2. Consequently, $\mathbb{P}^m \times \mathbb{P}^n$ and $\mathbb{P}^{m+n}$ have the common open subset $\mathbb{A}^m \times \mathbb{A}^n = \mathbb{A}^{m+n}$ — but they are not isomorphic by Exercise 7.4.

(b) Similarly, the affine space $\mathbb{A}^{k(n-k)}$ and the Grassmannian $G(k,n)$ have the common open subset $\mathbb{A}^{k(n-k)}$ by Construction 8.15.

(c) The affine line $\mathbb{A}^1$ and the curve $X = V(x_1^2 - x_2^3) \subset \mathbb{A}^2$ of Example 4.9 have the isomorphic open subsets $\mathbb{A}^1 \backslash \{0\}$ resp. $X \backslash \{0\}$ — in fact, the morphism $f$ given there is an isomorphism after removing the origin from both the source and the target curve.

We now want to study this situation in more detail and present a very general construction — the so-called blow-ups — that gives rise to many examples of this type. But first of all we have to set up some notation to deal with morphisms that are defined on dense open subsets. For simplicity, we will do this only for the case of irreducible varieties, in which every non-empty open subset is automatically dense by Remark 2.18.

**Definition 9.2** (Rational maps)**.** Let $X$ and $Y$ be irreducible varieties. A **rational map** $f$ from $X$ to $Y$, written $f : X \dashrightarrow Y$, is a morphism $f : U \to Y$ (denoted by the same letter) from a non-empty open subset $U \subset X$ to $Y$. We say that two such rational maps $f_1 : U_1 \to Y$ and $f_2 : U_2 \to Y$ with $U_1, U_2 \subset X$ are the same if $f_1 = f_2$ on a non-empty open subset of $U_1 \cap U_2$.

**Remark 9.3.** Strictly speaking, Definition 9.2 means that a rational map $f : X \dashrightarrow Y$ is an equivalence class of morphisms from non-empty open subsets of $X$ to $Y$. Note that the given relation is in fact an equivalence relation: reflexivity and symmetry are obvious, and if $f_1 : U_1 \to Y$ agrees with $f_2 : U_2 \to Y$ on a non-empty open subset $U_{1,2}$ and $f_2$ with $f_3 : U_3 \to Y$ on a non-empty open subset $U_{2,3}$ then $f_1$ and $f_3$ agree on $U_{1,2} \cap U_{1,3}$, which is again non-empty by Remark 2.18 (a) since $X$ is irreducible. For the sake of readability it is customary however not to indicate these equivalence classes in the notation and to denote the rational map $f : X \dashrightarrow Y$ and the morphism $f : U \to Y$ by the same letter.

If we now want to consider "rational maps with an inverse", i. e. rational maps $f : X \dashrightarrow Y$ such that there is another rational map $g : Y \dashrightarrow X$ with $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, we run into problems: if e. g. $f$ is a constant map and $g$ is not defined at the point $f(X)$ then there is no meaningful way to compose it with $f$. So we need to impose a technical condition first to ensure that compositions are well-defined:

**Definition 9.4** (Birational maps)**.** Again let $X$ and $Y$ be irreducible varieties.

(a) A rational map $f : X \dashrightarrow Y$ is called **dominant** if its image contains a non-empty open subset $U$ of $Y$. In this case, if $g : Y \dashrightarrow Z$ is another rational map, defined on a non-empty open subset $V$ of $Y$, we can construct the composition $g \circ f : X \dashrightarrow Z$ as a rational map since we have such a composition of ordinary morphisms on the non-empty open subset $f^{-1}(U \cap V)$.

(b) A rational map $f : X \dashrightarrow Y$ is called **birational** if it is dominant, and if there is another dominant rational map $g : Y \dashrightarrow X$ with $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$.

(c) We say that $X$ and $Y$ are **birational** if there is a birational map $f : X \dashrightarrow Y$ between them.

**Remark 9.5.** By definition, two irreducible varieties are birational if and only if they contain isomorphic non-empty open subsets. In particular, Exercise 5.25 then implies that birational irreducible varieties have the same dimension.

14

An important case of rational maps is when the target space is just the ground field, i. e. if we consider regular functions on open subsets.

**Construction 9.6** (Rational functions and function fields)**.** Let $X$ be an irreducible variety.

A rational map $\varphi : X \dashrightarrow \mathbb{A}^1 = K$ is called a **rational function** on $X$. In other words, a rational function on $X$ is given by a regular function $\varphi \in \mathscr{O}_X(U)$ on some non-empty open subset $U \subset X$, with two such regular functions defining the same rational function if and only if they agree on a non-empty open subset. The set of all rational functions on $X$ will be denoted $K(X)$.

Note that $K(X)$ is a field: for $\varphi_1 \in \mathscr{O}_X(U_1)$ and $\varphi_2 \in \mathscr{O}_X(U_2)$ we can define $\varphi_1 + \varphi_2$ and $\varphi_1 \varphi_2$ on $U_1 \cap U_2 \neq \emptyset$, the additive inverse $-\varphi_1$ on $U_1$, and for $\varphi_1 \neq 0$ the multiplicative inverse $\varphi_1^{-1}$ on $U_1 \backslash V(\varphi_1)$. We call $K(X)$ the **function field** of $X$.

**Remark 9.7.** If $U \subset X$ is a non-empty open subset of an irreducible variety $X$ then $K(U) \cong K(X)$: an isomorphism is given by

$$\begin{aligned} K(U) &\to K(X) \\ \varphi \in \mathscr{O}_U(V) &\mapsto \varphi \in \mathscr{O}_X(V) \end{aligned} \quad \text{with inverse} \quad \begin{aligned} K(X) &\to K(U) \\ \varphi \in \mathscr{O}_X(V) &\mapsto \varphi|_{V\cap U} \in \mathscr{O}_U(V \cap U). \end{aligned}$$

In particular, birational irreducible varieties have isomorphic function fields.

**Exercise 9.8.** Let $X$ be an irreducible affine variety. Show:

(a) The function field $K(X)$ is isomorphic to the so-called *quotient field* of the coordinate ring $A(X)$, i. e. to the localization of the integral domain $A(X)$ at the multiplicatively closed subset $A(X)\backslash\{0\}$.

(b) Every local ring $\mathscr{O}_{X,a}$ for $a \in X$ is naturally a subring of $K(X)$.

**Exercise 9.9.** Let $X \subset \mathbb{P}^n$ be a *quadric*, i. e. an irreducible variety which is the zero locus of an irreducible homogeneous polynomial of degree 2. Show that $X$ is birational, but in general not isomorphic, to the projective space $\mathbb{P}^{n-1}$.

The main goal of this chapter is now to describe and study a general procedure to modify an irreducible variety to a birational one. In its original form, this construction depends on given polynomial functions $f_1, \dots, f_r$ on an affine variety $X$ — but we will see in Construction 9.17 that it can also be performed with a given ideal in $A(X)$ or subvariety of $X$ instead, and that it can be glued in order to work on arbitrary varieties.

**Construction 9.10** (Blowing up)**.** Let $X \subset \mathbb{A}^n$ be an affine variety. For some $r \in \mathbb{N}_{>0}$ let $f_1, \dots, f_r \in A(X)$ be polynomial functions on $X$, and set $U = X\backslash V(f_1, \dots, f_r)$. As $f_1, \dots, f_r$ then do not vanish simultaneously at any point of $U$, we obtain a well-defined morphism

$$f : U \to \mathbb{P}^{r-1}, \ x \mapsto (f_1(x) : \dots : f_r(x)).$$

We consider its graph

$$\Gamma_f = \{(x, f(x)) : x \in U\} \quad \subset U \times \mathbb{P}^{r-1}$$

which is isomorphic to $U$ (with inverse morphism the projection to the first factor). Note that $\Gamma_f$ is closed in $U \times \mathbb{P}^{r-1}$ by Proposition 5.21 (a), but in general not closed in $X \times \mathbb{P}^{r-1}$. The closure of $\Gamma_f$ in $X \times \mathbb{P}^{r-1}$ then contains $\Gamma_f$ as a dense open subset. It is called the **blow-up** of $X$ at $f_1, \dots, f_r$; we will usually denote it by $\tilde{X}$. Note that there is a natural projection morphism $\pi : \tilde{X} \to X$ to the first factor. Sometimes we will also say that this morphism $\pi$ is the blow-up of $X$ at $f_1, \dots, f_r$.

Before we give examples of blow-ups let us introduce some more notation and easy general results that will help us to deal with them.

**Remark 9.11** (Exceptional sets)**.** In construction 9.10, the graph $\Gamma_f$ is isomorphic to $U$, with $\pi|_{\Gamma_f} : \Gamma_f \to U$ being an isomorphism. By abuse of notation, one often uses this isomorphism to identify $\Gamma_f$ with $U$, so that $U$ becomes an open subset of $\tilde{X}$. Its complement $\tilde{X} \backslash U = \pi^{-1}(V(f_1,\dots,f_r))$, on which $\pi$ is usually not an isomorphism, is called the **exceptional set** of the blow-up.

If $X$ is irreducible and $f_1,\dots,f_r$ do not vanish simultaneously on all of $X$, then $U = X\backslash V(f_1,\dots,f_r)$ is a non-empty and hence dense open subset of $X$. So its closure in the blow-up, which is all of $\tilde{X}$ by definition, is also irreducible. We therefore conclude that $X$ and $\tilde{X}$ are birational in this case, with common dense open subset $U$.

**Remark 9.12** (Strict transforms and blow-ups of subvarieties)**.** In the notation of Construction 9.10, let $Y$ be a closed subvariety of $X$. Then we can blow up $Y$ at $f_1,\dots,f_r$ as well. By construction, the resulting space $\tilde{Y} \subset Y \times \mathbb{P}^{r-1} \subset X \times \mathbb{P}^{r-1}$ is then also a closed subvariety of $\tilde{X}$, in fact it is the closure of $Y \cap U$ in $\tilde{X}$ (using the isomorphism $\Gamma_f \cong U$ of Remark 9.11 to identify $Y \cap U$ with a subset of $\tilde{X}$). If we consider $\tilde{Y}$ as a subset of $\tilde{X}$ in this way it is often called the **strict transform** of $Y$ in the blow-up of $X$.

In particular, if $X = X_1 \cup \dots \cup X_m$ is the irreducible decomposition of $X$ then $\tilde{X}_i \subset \tilde{X}$ for $i = 1,\dots,m$. Moreover, since taking closures commutes with finite unions it is immediate from Construction 9.10 that

$$\tilde{X} = \tilde{X}_1 \cup \dots \cup \tilde{X}_m,$$

i. e. that for blowing up $X$ we just blow up its irreducible components individually. For many purposes it therefore suffices to consider blow-ups of irreducible varieties.

**Example 9.13** (Trivial cases of blow-ups)**.** Let $r = 1$ in the notation of Construction 9.10, i. e. consider the case when we blow up $X$ at only one function $f_1$. Then $\tilde{X} \subset X \times \mathbb{P}^0 \cong X$, and $\Gamma_f \cong U$. So $\tilde{X}$ is just the closure of $U$ in $X$ under this isomorphism. If we assume for simplicity that $X$ is irreducible we therefore obtain the following two cases:

(a) If $f_1 \neq 0$ then $U = X\backslash V(f_1)$ is a non-empty open subset of $X$, and hence $\tilde{X} = X$ by Remark 2.18 (b).

(b) If $f_1 = 0$ then $U = \emptyset$, and hence also $\tilde{X} = \emptyset$.

So in order to obtain interesting examples of blow-ups we will have to consider cases with $r \geq 2$.

In order to understand blow-ups better, one of our main tasks has to be to find an explicit description of them that does not refer to taking closures. The following inclusion is a first step in this direction.

**Lemma 9.14.** *The blow-up $\tilde{X}$ of an affine variety $X$ at $f_1,\dots,f_r \in A(X)$ satisfies*

$$\tilde{X} \subset \{(x,y) \in X \times \mathbb{P}^{r-1} : y_i f_j(x) = y_j f_i(x) \text{ for all } i,j = 1,\dots,r\}.$$

*Proof.* Let $U = X\backslash V(f_1,\dots,f_r)$. Then any point $(x,y) \in U \times \mathbb{P}^{r-1}$ on the graph $\Gamma_f$ of the function $f : U \to \mathbb{P}^{r-1}$, $x \mapsto (f_1(x) : \dots : f_r(x))$ satisfies $(y_1 : \dots : y_r) = (f_1(x) : \dots : f_r(x))$, and hence $y_i f_j(x) = y_j f_i(x)$ for all $i,j = 1,\dots,r$. As these equations then also have to hold on the closure $\tilde{X}$ of $\Gamma_f$, the lemma follows. □

**Example 9.15** (Blow-up of $\mathbb{A}^n$ at the coordinate functions)**.** Our first non-trivial (and in fact the most important) case of a blow-up is that of the affine space $\mathbb{A}^n$ at the coordinate functions $x_1,\dots,x_n$. This blow-up $\widetilde{\mathbb{A}^n}$ is then isomorphic to $\mathbb{A}^n$ on the open subset $U = \mathbb{A}^n\backslash V(x_1,\dots,x_n) = \mathbb{A}^n\backslash\{0\}$, and by Lemma 9.14 we have

$$\widetilde{\mathbb{A}^n} \subset \{(x,y) \in \mathbb{A}^n \times \mathbb{P}^{n-1} : y_i x_j = y_j x_i \text{ for all } i,j = 1,\dots,n\} =: Y. \tag{1}$$

We claim that this inclusion is in fact an equality. To see this, let us consider the open subset $U_1 = \{(x,y) \in Y : y_1 \neq 0\}$ with affine coordinates $x_1,\dots,x_n,y_2,\dots,y_n$ in which we set $y_1 = 1$. Note that for given $x_1,y_2,\dots,y_n$ the equations (1) for $Y$ then say exactly that $x_j = x_1 y_j$ for $j = 2,\dots,n$. Hence there is an isomorphism

$$\mathbb{A}^n \to U_1 \subset \mathbb{A}^n \times \mathbb{P}^{n-1}, \quad (x_1,y_2,\dots,y_n) \mapsto ((x_1,x_1y_2,\dots,x_1y_n),(1:y_2:\dots:y_n)). \tag{2}$$

Of course, the same holds for the open subsets $U_i$ of $Y$ where $y_i \neq 0$ for $i = 2,\dots,n$. Hence $Y$ can be covered by $n$-dimensional affine spaces. By Exercises 2.20 (b) and 2.33 (a) this means that $Y$ is irreducible of dimension $n$. But as $Y$ contains the closed subvariety $\widetilde{\mathbb{A}^n}$ which is also irreducible of dimension $n$ by Remarks 9.5 and 9.11, we conclude that we must already have $Y = \widetilde{\mathbb{A}^n}$.

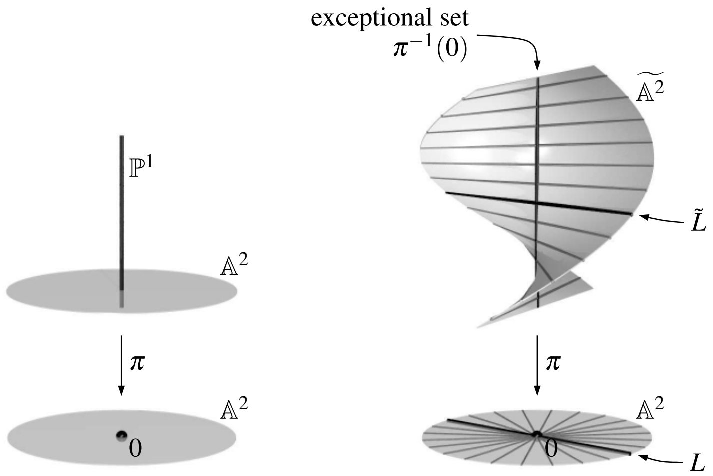
In fact, both the description (1) of $\widetilde{\mathbb{A}^n}$ (with equality, as we have just seen) and the affine coordinates of (2) are very useful in practice for explicit computations on this blow-up.

Let us now also study the blow-up (i. e. projection) morphism $\pi : \widetilde{\mathbb{A}^n} \to \mathbb{A}^n$ of Construction 9.10. We know already that this map is an isomorphism on $U = \mathbb{A}^n \backslash \{0\}$. In contrast, the exceptional set $\pi^{-1}(0)$ is given by setting $x_1,\dots,x_n$ to 0 in the description (1) above. As all defining equations $x_i y_j = x_j y_i$ become trivial in this case, we simply get

$$\pi^{-1}(0) = \{(0,y) \in \mathbb{A}^n \times \mathbb{P}^{n-1}\} \cong \mathbb{P}^{n-1}.$$

In other words, passing from $\mathbb{A}^n$ to $\widetilde{\mathbb{A}^n}$ leaves all points except 0 unchanged, whereas the origin is replaced by a projective space $\mathbb{P}^{n-1}$. This is the geometric reason why this construction is called blowing up — in fact, we will slightly extend our terminology in Construction 9.17 (a) so that we can then call the example above the blow-up of $\mathbb{A}^n$ *at the origin*, instead of at the functions $x_1,\dots,x_n$.

Because of this behavior of the inverse images of $\pi$ one might be tempted to think of $\widetilde{\mathbb{A}^n}$ as $\mathbb{A}^n$ with a projective space $\mathbb{P}^{n-1}$ attached at the origin, as in the picture below on the left. This is not correct however, as one can see already from the fact that this space would not be irreducible, whereas $\widetilde{\mathbb{A}^n}$ is. To get the true geometric picture for $\mathbb{A}^n$ let us consider the strict transform of a line $L \subset \mathbb{A}^n$ through the origin, i. e. the blow-up $\tilde{L}$ of $L$ at $x_1,\dots,x_n$ contained in $\widetilde{\mathbb{A}^n}$. We will give a general recipe to compute such strict transforms in Exercise 9.22, but in the case at hand this can also be done without much theory: by construction, over the complement of the origin every point $(x,y) \in \tilde{L} \subset L \times \mathbb{P}^{n-1}$ must have $y$ being equal to the projective point corresponding to $L \subset K^n$. Hence the same holds on the closure $\tilde{L}$, and thus the strict transform $\tilde{L}$ meets the exceptional set $\pi^{-1}(0) \cong \mathbb{P}^{n-1}$ above exactly in the point corresponding to $L$. In other words, the exceptional set parametrizes the directions in $\mathbb{A}^n$ at 0; two lines through the origin with distinct directions will become separated after the blow-up. The following picture on the right illustrates this in the case of the plane: we can imagine the blow-up $\widetilde{\mathbb{A}^2}$ as a helix winding around the central line $\pi^{-1}(0) \cong \mathbb{P}^1$ (in fact, it winds around this exceptional set once, so that one should think of the top of the helix as being glued to the bottom).




Wrong picture Correct picture

As already mentioned, the geometric interpretation of Example 9.15 suggests that we can think of this construction as the blow-up of $\mathbb{A}^n$ *at the origin* instead of *at the functions* $x_1,\dots,x_n$. To justify this notation let us now show that the blow-up construction does not actually depend on the chosen functions, but only on the ideal generated by them.

**Lemma 9.16.** *The blow-up of an affine variety $X$ at $f_1,\dots,f_r \in A(X)$ depends only on the ideal $(f_1,\dots,f_r) \trianglelefteq A(X)$.*

*More precisely, if $f_1',\dots,f_s' \in A(X)$ with $(f_1,\dots,f_r) = (f_1',\dots,f_s') \trianglelefteq A(X)$, and $\pi : \tilde{X} \to X$ and $\pi' : \tilde{X}' \to X$ are the corresponding blow-ups, there is an isomorphism $F : \tilde{X} \to \tilde{X}'$ with $\pi' \circ F = \pi$. In other words, we get a commutative diagram as in the picture on the right.*

$$\begin{tikzcd} \tilde{X} \arrow[rr, "F"] \arrow[dr, "\pi"'] & & \tilde{X}' \arrow[dl, "\pi'"] \\ & X & \end{tikzcd}$$

*Proof.* By assumption we have relations

$$f_i = \sum_{j=1}^{s} g_{i,j} f_j' \text{ for all } i = 1,\dots,r \quad \text{and} \quad f_j' = \sum_{k=1}^{r} h_{j,k} f_k \text{ for all } j = 1,\dots,s$$

in $A(X)$ for suitable $g_{i,j}, h_{j,k} \in A(X)$. We claim that then

$$F : \tilde{X} \to \tilde{X}', \ (x,y) \mapsto (x,y') := \Big(x, \Big(\sum_{k=1}^{r} h_{1,k}(x)\, y_k : \cdots : \sum_{k=1}^{r} h_{s,k}(x)\, y_k\Big)\Big)$$

is an isomorphism between $\tilde{X} \subset X \times \mathbb{P}^{r-1}$ and $\tilde{X}' \subset X \times \mathbb{P}^{s-1}$ as required. This is easy to check:

- The homogeneous coordinates of $y'$ are not simultaneously 0: note that by construction we have the relation $(y_1 : \cdots : y_r) = (f_1 : \cdots : f_r)$ on $U = X \backslash V(f_1,\dots,f_r) \subset \tilde{X} \subset X \times \mathbb{P}^{r-1}$, i.e. these two vectors are linearly dependent (and non-zero) at each point in this set. Hence the linear relations $f_i = \sum_{j,k} g_{i,j} h_{j,k} f_k$ in $f_1,\dots,f_r$ imply the corresponding relations $y_i = \sum_{j,k} g_{i,j} h_{j,k} y_k$ in $y_1,\dots,y_r$ on this set, and thus also on its closure $\tilde{X}$. So if we had $y_j' = \sum_k h_{j,k} y_k = 0$ for all $j$ then we would also have $y_i = \sum_j g_{i,j} y_j' = 0$ for all $i$, which is a contradiction.
- The image of $F$ lies in $\tilde{X}'$: by construction we have
$$F(x,y) = \Big(x, \Big(\sum_{k=1}^{r} h_{1,k}(x)\, f_k(x) : \cdots : \sum_{k=1}^{r} h_{s,k}(x)\, f_k(x)\Big)\Big) = \big(x, (f_1'(x) : \cdots : f_s'(x))\big) \quad \in \tilde{X}'$$
on the open subset $U$, and hence also on its closure $\tilde{X}$.
- $F$ is an isomorphism: by symmetry the same construction as above can also be done in the other direction and gives us an inverse morphism $F^{-1}$.
- It is obvious that $\pi' \circ F = \pi$. □

15

**Construction 9.17** (Generalizations of the blow-up construction)**.**

(a) Let $X$ be an affine variety. For an ideal $I \trianglelefteq A(X)$ we define the *blow-up of $X$ at $I$* to be the blow-up of $X$ at any set of generators of $I$ — which is well-defined up to isomorphisms by Lemma 9.16. If $Y \subset X$ is a closed subvariety the blow-up of $X$ at $I(Y) \trianglelefteq A(X)$ will also be called the *blow-up of $X$ at $Y$*. So in this language we can say that Example 9.15 describes the blow-up of $\mathbb{A}^n$ at the origin.

(b) Now let $X$ be an arbitrary variety, and let $Y \subset X$ be a closed subvariety. For an affine open cover $\{U_i : i \in I\}$ of $X$, let $\tilde{U}_i$ be the blow-up of $U_i$ at the closed subvariety $U_i \cap Y$. It is then easy to check that these blow-ups $\tilde{U}_i$ can be glued together to a variety $\tilde{X}$. We will call it again the *blow-up of $X$ at $Y$*.

In the following, we will probably only need this in the case of the blow-up of a point, where the construction is even easier as it is local around the blown-up point: let $X$ be a variety, and let $a \in X$ be a point. Choose an affine open neighborhood $U \subset X$ of $a$, and let $\tilde{U}$ be the blow-up of $U$ at $a$. Then we obtain $\tilde{X}$ by gluing $X \backslash \{a\}$ to $\tilde{U}$ along the common open subset $U \backslash \{a\}$.

(c) With our current techniques the gluing procedure of (b) only works for blow-ups at subvarieties — for the general construction of blowing up ideals we would need a way to patch ideals. This is in fact possible and leads to the notion of a *sheaf of ideals*, but we will not do this in this class.

Note however that *blow-ups of a projective variety* $X$ can be defined in essentially the same way as for affine varieties: if $f_1,\dots,f_r \in S(X)$ are homogeneous of the same degree the blow-up of $X$ at $f_1,\dots,f_r$ is defined as the closure of the graph

$$\Gamma = \{(x,(f_1(x):\cdots:f_r(x))) : x \in U\} \quad \subset U \times \mathbb{P}^{r-1}$$

(for $U = X \backslash V(f_1,\dots,f_r)$) in $X \times \mathbb{P}^{r-1}$; by the Segre embedding as in Remark 7.14 it is again a projective variety.

**Exercise 9.18.** Let $\widetilde{\mathbb{A}^3}$ be the blow-up of $\mathbb{A}^3$ at the line $V(x_1,x_2) \cong \mathbb{A}^1$. Show that its exceptional set is isomorphic to $\mathbb{A}^1 \times \mathbb{P}^1$. When do the strict transforms of two lines in $\mathbb{A}^3$ through $V(x_1,x_2)$ intersect in the blow-up? What is therefore the geometric meaning of the points in the exceptional set (corresponding to Example 9.15 in which the points of the exceptional set correspond to the directions through the blown-up point)?

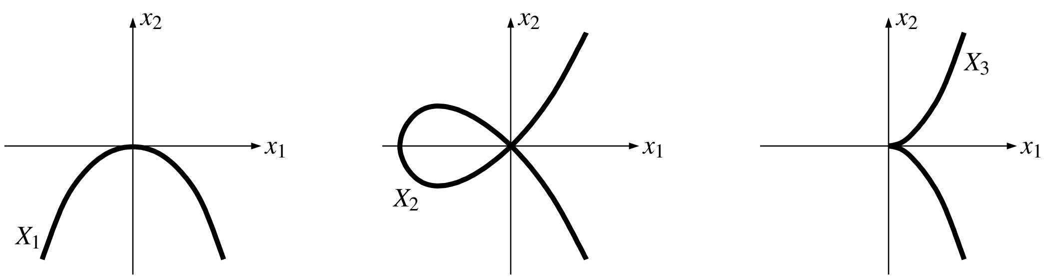
**Exercise 9.19.** Let $X \subset \mathbb{A}^n$ be an affine variety, and let $Y_1, Y_2 \subsetneq X$ be irreducible, closed subsets, no-one contained in the other. Moreover, let $\tilde{X}$ be the blow-up of $X$ at the ideal $I(Y_1)+I(Y_2)$. Show that the strict transforms of $Y_1$ and $Y_2$ in $\tilde{X}$ are disjoint.

One of the main applications of blow-ups is the local study of varieties. We have seen already in Example 9.15 that the exceptional set of the blow-up of $\mathbb{A}^n$ at the origin parametrizes the directions of lines at this point. It should therefore not come as a surprise that the exceptional set of the blow-up of a general variety $X$ at a point $a \in X$ parametrizes the tangent directions of $X$ at $a$.

**Construction 9.20** (Tangent cones)**.** Let $a$ be a point on a variety $X$. Consider the blow-up $\pi : \tilde{X} \to X$ of $X$ at $a$; its exceptional set $\pi^{-1}(a)$ is a projective variety (e. g. by choosing an affine open neighborhood $U \subset \mathbb{A}^n$ of $a = (a_1,\dots,a_n)$ in $X$ and blowing up $U$ at $x_1 - a_1,\dots,x_n - a_n$; the exceptional set is then contained in the projective space $\{a\} \times \mathbb{P}^{n-1} \subset U \times \mathbb{P}^{n-1}$).

The cone over this exceptional set $\pi^{-1}(a)$ (as in Definition 6.15 (c)) is called the **tangent cone** $C_aX$ of $X$ at $a$. Note that it is well-defined up to isomorphisms by Lemma 9.16. In the special case (of an affine patch) when $X \subset \mathbb{A}^n$ and $a \in X$ is the origin, we will also consider $C_aX \subset C(\mathbb{P}^{n-1}) = \mathbb{A}^n$ as a closed subvariety of the same ambient affine space as for $X$ by blowing up at $x_1,\dots,x_n$.

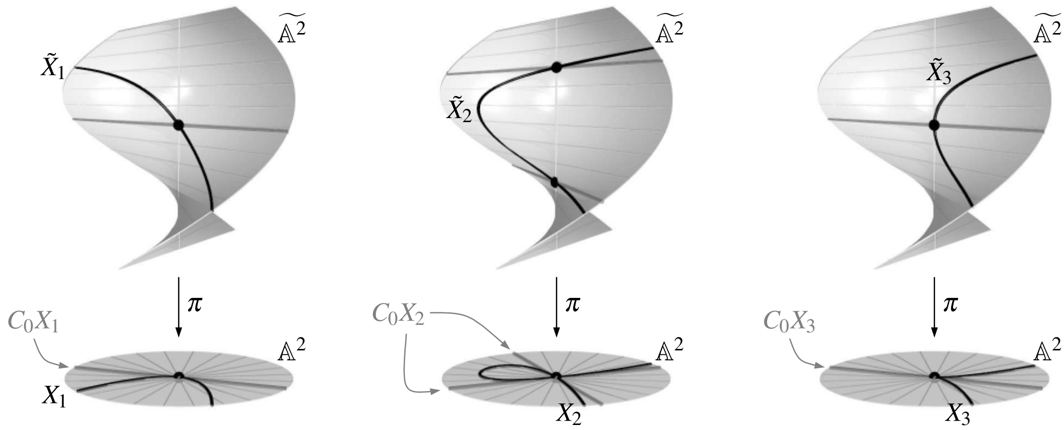
**Example 9.21.** Consider the three complex affine curves $X_1, X_2, X_3 \subset \mathbb{A}^2_{\mathbb{C}}$ with real parts as in the picture below.




$X_1 = V(x_2 + x_1^2)$ $\quad$ $X_2 = V(x_2^2 - x_1^2 - x_1^3)$ $\quad$ $X_3 = V(x_2^2 - x_1^3)$

Note that by Remark 9.12 the blow-ups $\tilde{X}_i$ of these curves at the origin (for $i = 1,2,3$) are contained as strict transforms in the blow-up $\widetilde{\mathbb{A}^2}$ of the affine plane at the origin as in Example 9.15. They can thus be obtained geometrically as in the following picture by lifting the curves $X_i \backslash \{0\}$ by the map $\pi : \widetilde{\mathbb{A}^2} \to \mathbb{A}^2$ and taking the closure in $\widetilde{\mathbb{A}^2}$. The additional points in these closures (drawn as dots in the picture below) are the exceptional sets of the blow-ups. By definition, the tangent cones $C_0X_i$

then consist of the lines corresponding to these points, as shown in gray below. They can be thought of as the cones, i. e. unions of lines, that approximate $X_i$ best around the origin.




Let us now study how these tangent cones can be computed rigorously. For example, for a point $((x_1,x_2),(y_1:y_2)) \in \tilde{X}_2 \subset \widetilde{\mathbb{A}^2} \subset \mathbb{A}^2 \times \mathbb{P}^1$ we have $x_2^2 - x_1^2 - x_1^3 = 0$ (as the equation of the curve) and $y_1x_2 - y_2x_1 = 0$ by Lemma 9.14. The latter means that the vectors $(x_1,x_2)$ and $(y_1,y_2)$ are linearly dependent, i. e. that $y_1 = \lambda x_1$ and $y_2 = \lambda x_2$ away from the origin for some non-zero $\lambda \in K$. Multiplying the equation of the curve with $\lambda^2$ thus yields

$$\lambda^2 (x_2^2 - x_1^2 - x_1^3) = 0 \quad \Rightarrow \quad y_2^2 - y_1^2 - y_1^2 x_1 = 0$$

on $\tilde{X}_2 \backslash \pi^{-1}(0)$, and thus also on its closure $\tilde{X}_2$. On $\pi^{-1}(0)$, i. e. if $x_1 = x_2 = 0$, this implies

$$y_2^2 - y_1^2 = 0 \quad \Rightarrow \quad (y_2 - y_1)(y_2 + y_1) = 0,$$

so that the exceptional set consists of the two points with $(y_1:y_2) \in \mathbb{P}^1$ equal to $(1:1)$ or $(1:-1)$. Consequently, the tangent cone $C_0X_2$ is the cone in $\mathbb{A}^2$ with the same equation

$$(x_2 - x_1)(x_2 + x_1) = 0,$$

i. e. the union of the two diagonals in $\mathbb{A}^2$ as in the picture above.

Note that the effect of this computation was exactly to pick out the terms of minimal degree of the defining equation $x_2^2 - x_1^2 - x_1^3 = 0$ — in this case of degree 2 — to obtain the equation $x_2^2 - x_1^2 = 0$ of the tangent cone at the origin. This obviously yields a homogeneous polynomial (so that its affine zero locus is a cone), and it fits well with the intuitive idea that for small values of $x_1$ and $x_2$ the higher powers of the coordinates are much smaller, so that we get a good approximation for the curve around the origin when we neglect them.

In fact, the following exercise (which is similar in style to proposition 6.33) shows that taking the terms of smallest degree of the defining equations is the general way to compute tangent cones explicitly after the coordinates have been shifted so that the point under consideration is the origin.

**Exercise 9.22** (Computation of tangent cones)**.** Let $I \trianglelefteq K[x_1,\dots,x_n]$ be an ideal, and assume that the corresponding affine variety $X = V(I) \subset \mathbb{A}^n$ contains the origin. Consider the blow-up $\tilde{X} \subset \widetilde{\mathbb{A}^n} \subset \mathbb{A}^n \times \mathbb{P}^{n-1}$ at $x_1,\dots,x_n$, and denote the homogeneous coordinates of $\mathbb{P}^{n-1}$ by $y_1,\dots,y_n$.

(a) By Example 9.15 we know that $\widetilde{\mathbb{A}^n}$ can be covered by affine spaces, with one coordinate patch being

$$\mathbb{A}^n \to \widetilde{\mathbb{A}^n} \subset \mathbb{A}^n \times \mathbb{P}^{n-1},$$
$$(x_1,y_2,\dots,y_n) \mapsto ((x_1,x_1y_2,\dots,x_1y_n),(1:y_2:\dots:y_n)).$$

Prove that on this coordinate patch the blow-up $\tilde{X}$ is given as the zero locus of the polynomials

$$\frac{f(x_1, x_1 y_2, \dots, x_1 y_n)}{x_1^{\min \deg f}}$$

for all non-zero $f \in I$, where $\min \deg f$ denotes the smallest degree of a monomial in $f$.

(b) Prove that the exceptional hypersurface of $\tilde{X}$ is

$$V_p(f^{in} : f \in I) \quad \subset \{0\} \times \mathbb{P}^{n-1},$$

where $f^{in}$ is the *initial term* of $f$, i.e. the sum of all monomials in $f$ of smallest degree. Consequently, the tangent cone of $X$ at the origin is

$$C_0 X = V_a(f^{in} : f \in I) \quad \subset \mathbb{A}^n.$$

(c) If $I = (f)$ is a principal ideal prove that $C_0 X = V_a(f^{in})$. However, for a general ideal $I$ show that $C_0 X$ is in general not the zero locus of the initial terms of a set of generators for $I$.

In Example 9.15 above, blowing up the $n$-dimensional variety $\mathbb{A}^n$ at $(x_1, \dots, x_n)$ has replaced the origin by a variety $\mathbb{P}^{n-1}$ of codimension 1 in $\widetilde{\mathbb{A}^n}$. We will now see that this is in fact a general phenomenon.

**Proposition 9.23** (Dimension of the exceptional set)**.** *Let $\pi : \tilde{X} \to X$ be the blow-up of an irreducible affine variety $X$ at $f_1, \dots, f_r \in A(X)$. Then every irreducible component of the exceptional set $\pi^{-1}(V(f_1, \dots, f_r))$ has codimension 1 in $\tilde{X}$. It is therefore often called the **exceptional hypersurface** of the blow-up.*

*Proof.* It is enough to prove the statement on the affine open subsets $U_i \subset \tilde{X} \subset X \times \mathbb{P}^{r-1}$ for $i = 1, \dots, r$ where the $i$-th projective coordinate $y_i$ is non-zero, since these open subsets cover $\tilde{X}$. But note that for $a \in U_i$ the condition $f_i(a) = 0$ implies $f_j(a) = 0$ for all $j$ by Lemma 9.14. So the exceptional set is given by one equation $f_i = 0$ on $U_i$. Moreover, if $U_i$ is non-empty then this polynomial $f_i$ is not identically zero on $U_i$: otherwise $U_i$, and thus also its closure $\tilde{X}$, would be contained in the exceptional set — which is a contradiction since this implies $U = \emptyset$ and thus $\tilde{X} = \emptyset$. The statement of the lemma thus follows from Proposition 2.25 (c). $\square$

**Corollary 9.24** (Dimension of tangent cones)**.** *Let $a$ be a point on a variety $X$. Then the dimension $\dim C_a X$ of the tangent cone of $X$ at $a$ is the local dimension $\operatorname{codim}_X\{a\}$ of $X$ at $a$.*

*Proof.* Note that both $\dim C_a X$ and $\operatorname{codim}_X\{a\}$ are local around the point $a$. By passing to an open neighborhood of $a$ we can therefore assume that every irreducible component of $X$ meets $a$, and that $X \subset \mathbb{A}^n$ is affine. We may also assume that $X$ is not just the one-point set $\{a\}$, since otherwise the statement of the corollary is trivial.

Now let $X = X_1 \cup \dots \cup X_m$ be the irreducible decomposition of $X$. Note that $X \neq \{a\}$ implies that all of these components have dimension at least 1. By Proposition 9.23 every irreducible component of the exceptional set of the blow-up $\tilde{X}_i$ of $X_i$ at $a$ has dimension $\dim X_i - 1$, and so by Exercise 6.32 (a) every irreducible component of the tangent cone $C_a X_i$ has dimension $\dim X_i$. As the maximum of these dimensions is just the local dimension $\operatorname{codim}_X\{a\}$ (see Exercise 5.11 (b)) it therefore suffices to show that all these exceptional sets (and hence also the tangent cones) are non-empty.

Assume the contrary, i.e. that the exceptional set of $\tilde{X}_i$ is empty for some $i$. Extending this to the projective closure $\mathbb{P}^n$ of $\mathbb{A}^n$ we obtain an irreducible variety $\overline{X_i} \subset \mathbb{P}^n$ containing $a$ whose blow-up $\widetilde{\overline{X_i}}$ in $\widetilde{\mathbb{P}^n}$ has an empty exceptional set. This means that $\pi\left(\widetilde{\overline{X_i}}\right) = \overline{X_i} \backslash \{a\}$, where $\pi : \widetilde{\mathbb{P}^n} \to \mathbb{P}^n$ is the blow-up map. As $\widetilde{\overline{X_i}}$ is a projective (and hence complete) variety by Construction 9.17 (c) this is a contradiction to Corollary 7.24 since $\overline{X_i} \backslash \{a\}$ is not closed (recall that $\overline{X_i}$ has dimension at least 1, so that $\overline{X_i} \backslash \{a\} \neq \emptyset$). $\square$

**Exercise 9.25.** Let $X = V(x_2^2 - x_1^2 - x_1^3) \subset \mathbb{A}^2$. Show that $X$ is not isomorphic to $\mathbb{A}^1$, but that the blow-up of $X$ at the origin is.

Can you interpret this result geometrically?

**Exercise 9.26.**

(a) Show that the blow-up of $\mathbb{A}^2$ at the ideal $(x_1^2, x_1x_2, x_2^2)$ is isomorphic to the blow-up of $\mathbb{A}^2$ at the ideal $(x_1, x_2)$.

(b) Let $X$ be an affine variety, and let $I \trianglelefteq A(X)$ be an ideal. Is it true in general that the blow-up of $X$ at $I$ is isomorphic to the blow-up of $X$ at $\sqrt{I}$?

We will now discuss another important application of blow-ups that follows more or less directly from the definitions: they can be used to extend morphisms defined only on an open subset of a variety.

**Remark 9.27** (Blowing up to extend morphisms)**.** Let $X \subset \mathbb{A}^n$ be an affine variety, and let $f_1, \dots, f_r$ be polynomial functions on $X$. Note that the morphism $f : x \mapsto (f_1(x) : \cdots : f_r(x))$ to $\mathbb{P}^{r-1}$ is only well-defined on the open subset $U = X \backslash V(f_1, \dots, f_r)$ of $X$. In general, we can not expect that this morphism can be extended to a morphism on all of $X$. But we can always extend it "after blowing up the ideal $(f_1, \dots, f_r)$ of the indeterminacy locus": there is an extension $\tilde{f} : \tilde{X} \to \mathbb{P}^{r-1}$ of $f$ (that agrees with $f$ on $U$), namely just the projection from $\tilde{X} \subset X \times \mathbb{P}^{r-1}$ to the second factor $\mathbb{P}^{r-1}$. So blowing up is a way to extend morphisms to bigger sets on which they would otherwise be ill-defined. Let us consider a concrete example of this idea in the next lemma and the following remark.

**Lemma 9.28.** *$\mathbb{P}^1 \times \mathbb{P}^1$ blown up in one point is isomorphic to $\mathbb{P}^2$ blown up in two points.*

*Proof.* We know from Example 7.12 that $\mathbb{P}^1 \times \mathbb{P}^1$ is isomorphic to the quadric surface

$$X = \{(x_0 : x_1 : x_2 : x_3) : x_0x_3 = x_1x_2\} \quad \subset \mathbb{P}^3.$$

Let $\tilde{X}$ be blow-up of $X$ at $a = (0:0:0:1) \in X$, which can be realized as in Construction 9.17 (c) as the blow-up $\tilde{X} \subset \mathbb{P}^3 \times \mathbb{P}^2$ of $X$ at $x_0, x_1, x_2$.

On the other hand, let $b = (0:1:0), c = (0:0:1) \in \mathbb{P}^2$, and let $\widetilde{\mathbb{P}^2} \subset \mathbb{P}^2 \times \mathbb{P}^3$ be the blow-up of $\mathbb{P}^2$ at $y_0^2, y_0y_1, y_0y_2, y_1y_2$. Note that these polynomials do not generate the ideal $I(\{b,c\}) = (y_0, y_1y_2)$, but this does not matter: the blow-up is a local construction, so let us check that we are locally just blowing up $b$, and similarly $c$. There is an open affine neighborhood around $b$ given by $y_1 \neq 0$, where we can set $y_1 = 1$, and on this neighborhood the given functions $y_0^2, y_0, y_0y_2, y_2$ generate the ideal $(y_0, y_2)$ of $b$. So $\widetilde{\mathbb{P}^2}$ is actually the blow-up of $\mathbb{P}^2$ at $b$ and $c$.

Now we claim that an isomorphism is given by

$$f : \tilde{X} \mapsto \widetilde{\mathbb{P}^2}, \ \ ((x_0 : x_1 : x_2 : x_3), (y_0 : y_1 : y_2)) \mapsto ((y_0 : y_1 : y_2), (x_0 : x_1 : x_2 : x_3)).$$

In fact, this is easy to prove: obviously, $f$ is an isomorphism from $\mathbb{P}^3 \times \mathbb{P}^2$ to $\mathbb{P}^2 \times \mathbb{P}^3$, so we only have to show that $f$ maps $\tilde{X}$ to $\widetilde{\mathbb{P}^2}$, and that $f^{-1}$ maps $\widetilde{\mathbb{P}^2}$ to $\tilde{X}$. Note that it suffices to check this on a dense open subset. But this is easy: on the complement of the exceptional set in $\tilde{X}$ we have $x_0x_3 = x_1x_2$ and $(y_0 : y_1 : y_2) = (x_0 : x_1 : x_2)$, so on the (smaller) complement of $V(x_0)$ we get the correct equations

$$(x_0 : x_1 : x_2 : x_3) = (x_0^2 : x_0x_1 : x_0x_2 : x_0x_3) = (x_0^2 : x_0x_1 : x_0x_2 : x_1x_2) = (y_0^2 : y_0y_1 : y_0y_2 : y_1y_2)$$

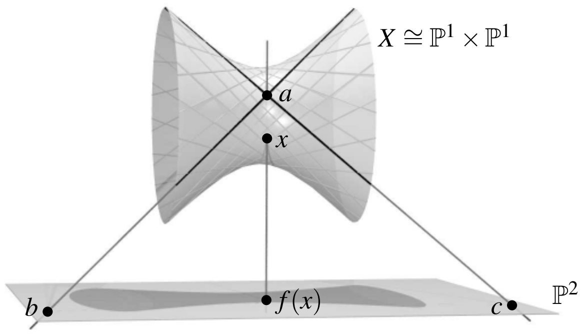
for the image point under $f$ to lie in $\widetilde{\mathbb{P}^2}$. Conversely, on the complement of the exceptional set in $\widetilde{\mathbb{P}^2}$ we have $(x_0 : x_1 : x_2 : x_3) = (y_0^2 : y_0y_1 : y_0y_2 : y_1y_2)$, so we conclude that $x_0x_3 = x_1x_2$ and $(y_0 : y_1 : y_2) = (x_0 : x_1 : x_2)$ where $y_0 \neq 0$. $\square$

16

**Remark 9.29.** The proof of Lemma 9.28 is short and elegant, but not very insightful. So let us try to understand geometrically what is going on. As in the proof above, we think of $\mathbb{P}^1 \times \mathbb{P}^1$ as the quadric surface

$$X = \{(x_0 : x_1 : x_2 : x_3) : x_0x_3 = x_1x_2\} \quad \subset \mathbb{P}^3.$$

Let us project $X$ from $a = (0:0:0:1) \in X$ to $V_p(x_3) \cong \mathbb{P}^2$. The corresponding morphism $f$ is shown in the picture below; as in Example 7.6 (b) it is given by $f(x_0:x_1:x_2:x_3) = (x_0:x_1:x_2)$ and well-defined away from $a$.




Recall that, in the corresponding case of the projection of a quadric *curve* in Example 7.6 (c), the morphism $f$ could be extended to the point $a$. This is now no longer the case for our quadric *surface* $X$: to construct $f(a)$ we would have to take the limit of the points $f(x)$ for $x$ approaching $a$, i. e. consider lines through $a$ and $x$ for $x \to a$. These lines will then become tangent lines to $X$ at $a$ — but $X$, being two-dimensional, has a one-parameter family of such tangent lines. This is why $f(a)$ is ill-defined. But we also see from this discussion that blowing up $a$ on $X$, i. e. replacing it by the set of all tangent lines through $a$, will exactly resolve this indeterminacy. Hence $f$ becomes a well-defined morphism from $\tilde{X}$ to $V_p(x_3) \cong \mathbb{P}^2$.

Let us now check if there is an inverse morphism. By construction, it is easy to see what it would have to look like: the points of $X \backslash \{a\}$ mapped to a point $y \in V_p(x_3)$ are exactly those on the line $\overline{ay}$ through $a$ and $y$. In general, this line intersects $X$ in two points, one of which is $a$. So there is then exactly one point on $X$ which maps to $y$, leading to an inverse morphism $f^{-1}$. This reasoning is only false if the whole line $\overline{ay}$ lies in $X$. Then this whole line would be mapped to $y$, so that we cannot have an inverse $f^{-1}$ there. But of course we expect again that this problem can be taken care of by blowing up $y$ in $\mathbb{P}^2$, so that it is replaced by a $\mathbb{P}^1$ that can then be mapped bijectively to $\overline{ay}$.

There are obviously two such lines $\overline{ab}$ and $\overline{ac}$, given by $b = (0:1:0)$ and $c = (0:0:1)$. If you think of $X$ as $\mathbb{P}^1 \times \mathbb{P}^1$ again, these lines are precisely the "horizontal" and "vertical" lines passing through $a$ where the coordinate in one of the two factors is constant. So we would expect that $f$ can be made into an isomorphism after blowing up $b$ and $c$, which is exactly what we have shown in Lemma 9.28.

**Exercise 9.30** (Cremona transformation)**.** Let $a = (1:0:0)$, $b = (0:1:0)$, and $c = (0:0:1)$ be the three coordinate points of $\mathbb{P}^2$, and let $U = \mathbb{P}^2 \backslash \{a,b,c\}$. Consider the morphism

$$f : U \to \mathbb{P}^2, \ (x_0:x_1:x_2) \mapsto (x_1x_2 : x_0x_2 : x_0x_1).$$

(a) Show that there is no morphism $\mathbb{P}^2 \to \mathbb{P}^2$ extending $f$.

(b) Let $\widetilde{\mathbb{P}^2}$ be the blow-up of $\mathbb{P}^2$ at $\{a,b,c\}$. Show that $f$ can be extended to an isomorphism $\tilde{f} : \widetilde{\mathbb{P}^2} \to \widetilde{\mathbb{P}^2}$. This isomorphism is called the *Cremona transformation*.

# 10. Smooth Varieties

Let $a$ be a point on a variety $X$. In the last chapter we have introduced the tangent cone $C_aX$ as a way to study $X$ locally around $a$ (see Construction 9.20). It is a cone whose dimension is the local dimension $\operatorname{codim}_X\{a\}$ (Corollary 9.24), and we can think of it as the cone that best approximates $X$ around $a$. In an affine open chart where $a$ is the origin, we can compute $C_aX$ by choosing an ideal with zero locus $X$ and replacing each polynomial in this ideal by its initial term (Exercise 9.22 (b)).

However, in practice one often wants to approximate a given variety by a linear space rather than by a cone. We will therefore study now to what extent this is possible, and how the result compares to the tangent cones that we already know. Of course, the idea to construct this is just to take the *linear terms* instead of the *initial terms* of the defining polynomials when considering the origin in an affine variety. For simplicity, let us therefore assume for a moment that we have chosen an affine neighborhood of the point $a$ such that $a = 0$ — we will see in Lemma 10.5 that the following construction actually does not depend on this choice.

**Definition 10.1** (Tangent spaces)**.** Let $a$ be a point on a variety $X$. By choosing an affine neighborhood of $a$ we assume that $X \subset \mathbb{A}^n$ and that $a = 0$ is the origin. Then

$$T_aX := V(f_1 : f \in I(X)) \quad \subset \mathbb{A}^n$$

is called the **tangent space** of $X$ at $a$, where $f_1 \in K[x_1,\dots,x_n]$ denotes the linear term of a polynomial $f \in K[x_1,\dots,x_n]$ as in Definition 6.6 (a).

As in the case of tangent cones, we can consider $T_aX$ either as an abstract variety (leaving its dimension as the only invariant since it is a linear space) or as a subspace of $\mathbb{A}^n$.

**Remark 10.2.**

(a) In contrast to the case of tangent cones in Exercise 9.22 (c), it always suffices in Definition 10.1 to take the zero locus only of the linear parts of a set $S$ of generators for $I(X)$: if $f, g \in S$ are polynomials such that $f_1$ and $g_1$ vanish at a point $x \in \mathbb{A}^n$ then

$$(f+g)_1(x) = f_1(x) + g_1(x) = 0$$

$$\text{and} \qquad (hf)_1(x) = h(0)\,f_1(x) + f(0)\,h_1(x) = h(0)\cdot 0 + 0 \cdot h_1(x) = 0$$

for an arbitrary polynomial $h \in K[x_1,\dots,x_n]$, and hence $x \in T_aX$.

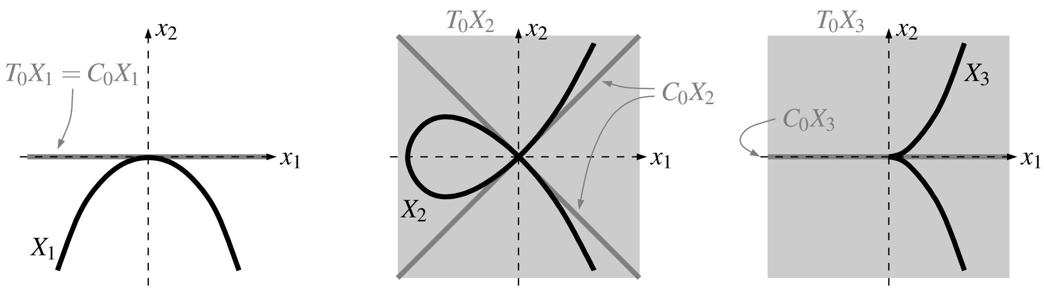
(b) However, again in contrast to the case of tangent cones in Exercise 9.22 it is crucial in Definition 10.1 that we take the radical ideal of $X$ and not just any ideal with zero locus $X$: the ideals $(x)$ and $(x^2)$ in $K[x]$ have the same zero locus $\{0\}$ in $\mathbb{A}^1$, but the zero locus of the linear term of $x$ is the origin again, whereas the zero locus of the linear term of $x^2$ is all of $\mathbb{A}^1$.

(c) For polynomials vanishing at the origin, a non-vanishing linear term is clearly always initial. Hence by Exercise 9.22 (b) it follows that $C_aX \subset T_aX$, i. e. that the tangent space always contains the tangent cone. In particular, this means by Corollary 9.24 that $\dim T_aX \geq \operatorname{codim}_X\{a\}$.

**Example 10.3.** Consider again the three curves $X_1, X_2, X_3$ of Example 9.21. By taking the initial resp. linear term of the defining polynomials we can compute the tangent cones and spaces of these curves at the origin:

- $X_1 = V(x_2 + x_1^2)$: $C_0X_1 = T_0X_1 = V(x_2)$;
- $X_2 = V(x_2^2 - x_1^2 - x_1^3)$: $C_0X_2 = V(x_2^2 - x_1^2)$, $T_0X_2 = V(0) = \mathbb{A}^2$;
- $X_3 = V(x_2^2 - x_1^3)$: $C_0X_3 = V(x_2^2) = V(x_2)$, $T_0X_3 = V(0) = \mathbb{A}^2$.

The following picture shows these curves together with their tangent cones and spaces. Note that for the curve $X_1$ the tangent cone is already a linear space, and the notions of tangent space and tangent cone agree. In contrast, the tangent cone of $X_2$ at the origin is not linear. By Remark 10.2 (c), the tangent space $T_0X_2$ must be a linear space containing $C_0X_2$, and hence it is necessarily all of $\mathbb{A}^2$. However, the curve $X_3$ shows that the tangent space is not always the linear space spanned by the tangent cone.




$X_1 = V(x_2 + x_1^2)$ $\quad$ $X_2 = V(x_2^2 - x_1^2 - x_1^3)$ $\quad$ $X_3 = V(x_2^2 - x_1^3)$

Before we study the relation between tangent spaces and cones in more detail, let us show first of all that the (dimension of the) tangent space is actually an intrinsic local invariant of a variety around a point, i. e. that it does not depend on a choice of affine open subset or coordinates around the point. We will do this by establishing an alternative description of the tangent space that does not need any such choices. The key observation needed for this is the isomorphism of the following lemma.

**Lemma 10.4.** *Let $X \subset \mathbb{A}^n$ be an affine variety containing the origin. Moreover, let us denote by $M := (\overline{x_1}, \dots, \overline{x_n}) = I(0) \trianglelefteq A(X)$ the ideal of the origin in $X$. Then there is a natural vector space isomorphism*

$$M/M^2 \cong \operatorname{Hom}_K(T_0X, K).$$

*In other words, the tangent space $T_0X$ is naturally the vector space dual to $M/M^2$.*

*Proof.* Consider the $K$-linear map

$$\varphi : M \to \operatorname{Hom}_K(T_0X, K), \ \overline{f} \mapsto f_1|_{T_0X}$$

sending the class of a polynomial modulo $I(X)$ to its linear term, regarded as a map restricted to the tangent space. By definition of the tangent space, this map is well-defined. Moreover, note that $\varphi$ is surjective since any linear map on $T_0X$ can be extended to a linear map on $\mathbb{A}^n$. So by the homomorphism theorem it suffices to prove that $\ker \varphi = M^2$:

"$\subset$" Consider the vector subspace $W = \{g_1 : g \in I(X)\}$ of $K[x_1, \dots, x_n]$, and let $k$ be its dimension. Then its zero locus $T_0X$ has dimension $n-k$, and hence the space of linear forms vanishing on $T_0X$ has dimension $k$ again. As it clearly contains $W$, we conclude that $W$ must be equal to the space of linear forms vanishing on $T_0X$.

So if $\overline{f} \in \ker \varphi$, i. e. the linear term of $f$ vanishes on $T_0X$, we know that there is a polynomial $g \in I(X)$ with $g_1 = f_1$. But then $f - g$ has no constant or linear term, and hence $\overline{f} = \overline{f-g} \in M^2$.

"$\supset$" If $\overline{f}, \overline{g} \in M$ then $(fg)_1 = f(0)\,g_1 + g(0)\,f_1 = 0 \cdot g_1 + 0 \cdot f_1 = 0$, and hence $\varphi(\overline{fg}) = 0$. □

In order to make Lemma 10.4 into an intrinsic description of the tangent space we need to transfer it from the affine coordinate ring $A(X)$ (which for a general variety would require the choice of an affine coordinate chart that sends the given point $a$ to the origin) to the local ring $\mathscr{O}_{X,a}$ (which is independent of any choices). To do this, recall by Lemma 3.21 that with the notations from above we have $\mathscr{O}_{X,a} \cong S^{-1}A(X)$, where $S = A(X) \backslash M = \{f \in A(X) : f(a) \neq 0\}$ is the multiplicatively closed subset of polynomial functions that are non-zero at the point $a$. In this ring

$$S^{-1}M = \left\{ \frac{g}{f} : g, f \in A(X) \text{ with } g(a) = 0 \text{ and } f(a) \neq 0 \right\}$$

is just the maximal ideal $I_a$ of all local functions vanishing at $a$ as in Definition 3.22. Using these constructions we obtain the following result.

**Lemma 10.5.** *With notations as above we have*

$$M/M^2 \cong (S^{-1}M)/(S^{-1}M)^2.$$

*In particular, if $a$ is a point on a variety $X$ and $I_a = \{\varphi \in \mathscr{O}_{X,a} : \varphi(a) = 0\}$ is the maximal ideal of local functions in $\mathscr{O}_{X,a}$ vanishing at $a$, then $T_aX$ is naturally isomorphic to the vector space dual to $I_a/I_a^2$, and thus independent of any choices.*

*Proof.* This time we consider the vector space homomorphism

$$\varphi : M \to (S^{-1}M)/(S^{-1}M)^2, \;\; g \mapsto \overline{\left(\frac{g}{1}\right)}$$

where the bar denotes classes modulo $(S^{-1}M)^2$. In order to deduce the lemma from the homomorphism theorem we have to show the following three statements:

- $\varphi$ is surjective: Let $\frac{g}{f} \in S^{-1}M$. Then $\frac{g}{f(0)} \in M$ is an inverse image of the class of this fraction under $\varphi$ since
$$\frac{g}{f} - \frac{g}{f(0)} = \frac{g}{f} \cdot \frac{f(0)-f}{f(0)} \quad \in (S^{-1}M)^2$$
(note that $f(0)-f$ lies in $M$ as it does not contain a constant term).
- $\ker\varphi \subset M^2$: Let $g \in \ker\varphi$, i. e. $\frac{g}{1} \in (S^{-1}M)^2$. This means that
$$\frac{g}{1} = \sum_i \frac{h_i k_i}{f_i} \tag{$*$}$$
for a finite sum with elements $h_i, k_i \in M$ and $f_i \in S$. By bringing this to a common denominator we can assume that all $f_i$ are equal, say to $f$. The equation $(*)$ in $\mathscr{O}_{X,a}$ then means $\tilde{f}(fg - \sum_i h_i k_i) = 0$ in $A(X)$ for some $\tilde{f} \in S$ by Construction 3.12. This implies $\tilde{f} f g \in M^2$. But $((\tilde{f} f)(0) - \tilde{f} f)\, g \in M^2$ as well, and hence $(\tilde{f} f)(0)\, g \in M^2$, which implies $g \in M^2$ since $(\tilde{f} f)(0) \in K^*$.
- $M^2 \subset \ker\varphi$ is trivial. □

**Exercise 10.6.** Let $f : X \to Y$ be a morphism of varieties, and let $a \in X$. Show that $f$ induces a linear map $T_aX \to T_{f(a)}Y$ between tangent spaces.

We have now constructed two objects associated to the local structure of a variety $X$ at a point $a \in X$:

- the *tangent cone* $C_aX$, which is a cone of dimension $\operatorname{codim}_X\{a\}$, but in general not a linear space; and
- the *tangent space* $T_aX$, which is a linear space, but whose dimension might be bigger than $\operatorname{codim}_X\{a\}$.

Of course, we should give special attention to the case when these two notions agree, i. e. when $X$ can be approximated around $a$ by a linear space whose dimension is the local dimension of $X$ at $a$.

**Definition 10.7** (Smooth and singular varieties)**.** Let $X$ be a variety.

(a) A point $a \in X$ is called **smooth**, **regular**, or **non-singular** if $T_aX = C_aX$. Otherwise it is called a **singular** point of $X$.

(b) If $X$ has a singular point we say that $X$ is singular. Otherwise $X$ is called smooth, regular, or non-singular.

**Example 10.8.** Of the three curves of Example 10.3, exactly the first one is smooth at the origin. As in our original motivation for the definition of tangent spaces, this is just the statement that $X_1$ can be approximated around the origin by a straight line — in contrast to $X_2$ and $X_3$, which have a "multiple point" resp. a "corner" there. A more precise geometric interpretation of smoothness can be obtained by comparing our algebraic situation with the Implicit Function Theorem from analysis, see Remark 10.14.

**Lemma 10.9.** *Let $X$ be a variety, and let $a \in X$ be a point. The following statements are equivalent:*

(a) *The point $a$ is smooth on $X$.*

(b) $\dim T_a X = \operatorname{codim}_X \{a\}$.

(c) $\dim T_a X \leq \operatorname{codim}_X \{a\}$.

*Proof.* The implication (a) $\Rightarrow$ (b) follows immediately from Corollary 9.24, and (b) $\Rightarrow$ (c) is trivial. To prove (c) $\Rightarrow$ (a), note first that (c) together with Remark 10.2 (c) implies $\dim T_a X = \operatorname{codim}_X \{a\}$. But again by Remark 10.2 (c), the tangent space $T_a X$ contains the tangent cone $C_a X$, which is of the same dimension by Corollary 9.24. As $T_a X$ is irreducible (since it is a linear space), this is only possible if $T_a X = C_a X$, i. e. if $a$ is a smooth point of $X$. □ 17

**Remark 10.10** (Smoothness in commutative algebra)**.** Let $a$ be a point on a variety $X$.

(a) Let $I_a \trianglelefteq \mathscr{O}_{X,a}$ be the maximal ideal of local functions vanishing at $a$ as in Definition 3.22. Combining Lemma 10.5 with Lemma 10.9 we see that $a$ is a smooth point of $X$ if and only if the vector space dimension of $I_a/I_a^2$ is equal to the local dimension $\operatorname{codim}_X \{a\}$ of $X$ at $a$. This is a property of the local ring $\mathscr{O}_{X,a}$ alone, and one can therefore study it with methods from commutative algebra. A ring with these properties is usually called a *regular local ring* [G5, Definition 11.38], which is also the reason for the name "regular point" in Definition 10.7 (a).

(b) It is a result of commutative algebra that a regular local ring as in (a) is always an integral domain [G5, Proposition 11.40]. Translating this into geometry as in Proposition 2.9, this yields the intuitively obvious statement that a variety is *locally irreducible* at every smooth point $a$, i. e. that $X$ has only one irreducible component meeting $a$. Equivalently, any point on a variety at which two irreducible components meet is necessarily a singular point.

The good thing about smoothness is that it is very easy to check using (formal) partial derivatives:

**Proposition 10.11 (Affine Jacobi criterion).** *Let $X \subset \mathbb{A}^n$ be an affine variety with ideal $I(X) = (f_1, \dots, f_r)$, and let $a \in X$ be a point. Then $X$ is smooth at $a$ if and only if the rank of the $r \times n$* ***Jacobian matrix***

$$\left( \frac{\partial f_i}{\partial x_j}(a) \right)_{i,j}$$

*is at least $n - \operatorname{codim}_X \{a\}$.*

*Proof.* Let $x = (x_1, \dots, x_n)$ be the coordinates of $\mathbb{A}^n$, and let $y := x - a$ be the shifted coordinates in which the point $a$ becomes the origin. By a formal Taylor expansion, the linear term of the polynomial $f_i$ in these coordinates $y$ is $\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(a) \cdot y_j$. Hence the tangent space $T_a X$ is by Definition 10.1 and Remark 10.2 (a) the zero locus of these linear terms, i. e. the kernel of the Jacobian matrix $J = \left( \frac{\partial f_i}{\partial x_j}(a) \right)_{i,j}$. So by Lemma 10.9 the point $a$ is smooth if and only if $\dim \ker J \leq \operatorname{codim}_X \{a\}$, which is equivalent to $\operatorname{rk} J \geq n - \operatorname{codim}_X \{a\}$. □

To check smoothness for a point on a projective variety, we can of course restrict to an affine open subset of the point. However, the following exercise shows that there is also a projective version of the Jacobi criterion that does not need these affine patches and works directly with the homogeneous coordinates instead.

**Exercise 10.12.**

(a) Show that
$$\sum_{i=0}^{n} x_i \cdot \frac{\partial f}{\partial x_i} = d \cdot f$$
for any homogeneous polynomial $f \in K[x_0, \dots, x_n]$ of degree $d$.

(b) (**Projective Jacobi criterion**) Let $X \subset \mathbb{P}^n$ be a projective variety with homogeneous ideal $I(X) = (f_1, \dots, f_r)$, and let $a \in X$. Prove that $X$ is smooth at $a$ if and only if the rank of the $r \times (n+1)$ Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j}$ is at least $n - \operatorname{codim}_X \{a\}$.

In this criterion, note that the entries $\frac{\partial f_i}{\partial x_j}(a)$ of the Jacobian matrix are not well-defined: multiplying the coordinates of $a$ by a scalar $\lambda \in K^*$ will multiply $\frac{\partial f_i}{\partial x_j}(a)$ by $\lambda^{d_i - 1}$, where $d_i$ is the degree of $f_i$. However, these are just row transformations of the Jacobian matrix, which do not affect its rank. Hence the condition in the projective Jacobi criterion is well-defined.

**Remark 10.13** (Variants of the Jacobi criterion)**.** There are a few ways to extend the Jacobi criterion even further. For simplicity, we will discuss this here in the case of an affine variety $X$ as in Proposition 10.11, but it is easy to see that the corresponding statements hold in the projective setting of Exercise 10.12 (b) as well.

(a) If $X$ is irreducible then $\operatorname{codim}_X \{a\} = \dim X$ for all $a \in X$ by Proposition 2.25 (b). So in this case $a$ is a smooth point of $X$ if and only if the rank of the Jacobian matrix is at least $n - \dim X = \operatorname{codim} X$.

(b) Let $f_1, \dots, f_r \in K[x_1, \dots, x_n]$ be polynomials such that $X = V(f_1, \dots, f_r)$, but that do not necessarily generate the ideal of $X$ (as required in Proposition 10.11). Then the Jacobi criterion still holds in one direction: assume that the rank of the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j}$ is at least $n - \operatorname{codim}_X \{a\}$. The proof of the affine Jacobi criterion then shows that the zero locus of all linear terms of the elements of $(f_1, \dots, f_r)$ has dimension at most $\operatorname{codim}_X \{a\}$. The same then necessarily holds for the zero locus of all linear terms of the elements of $\sqrt{(f_1, \dots, f_r)} = I(X)$ (which might only be smaller). By Proposition 10.11 this means that $a$ is a smooth point of $X$.

The converse is in general false, as we have already seen in Remark 10.2 (b).

(c) Again let $f_1, \dots, f_r \in K[x_1, \dots, x_n]$ be polynomials with $X = V(f_1, \dots, f_r)$. This time assume that the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j}$ has maximal row rank, i.e. that its rank is equal to $r$. As every irreducible component of $X$ has dimension at least $n - r$ by Proposition 2.25 (c) we know moreover that $\operatorname{codim}_X \{a\} \geq n - r$. Hence the rank of the Jacobian matrix is $r \geq n - \operatorname{codim}_X \{a\}$, so $X$ is smooth at $a$ by (b).

**Remark 10.14** (Relation to the Implicit Function Theorem)**.** The version of the Jacobi criterion of Remark 10.13 (c) is closely related to the *Implicit Function Theorem* from analysis. Given real polynomials $f_1, \dots, f_r \in \mathbb{R}[x_1, \dots, x_n]$ (or more generally continuously differentiable functions on an open subset of $\mathbb{R}^n$) and a point $a$ in their common zero locus $X = V(f_1, \dots, f_r)$ such that the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j}$ has rank $r$, this theorem states roughly that $X$ is locally around $a$ the graph of a continuously differentiable function [G2, Proposition 27.9] — so that in particular it does not have any "corners". It can be shown that the same result holds over the complex numbers as well. So in the case $K = \mathbb{C}$ the statement of Remark 10.13 (c) that $X$ is smooth at $a$ can also be interpreted in this geometric way.

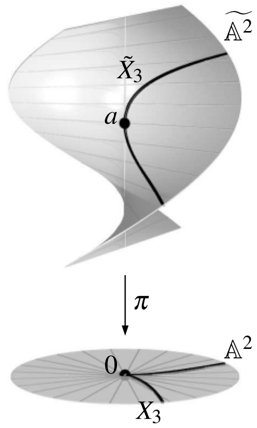
Note however that there is no algebraic analogue of the Implicit Function Theorem itself: for example, the polynomial equation $f(x_1, x_2) := x_2 - x_1^2 = 0$ cannot be solved for $x_1$ by a *regular* function locally around the point $(1,1)$, although $\frac{\partial f}{\partial x_1}(1,1) = -2 \neq 0$ — it can only be solved by a continuously differentiable function $x_1 = \sqrt{x_2}$.

**Example 10.15.** Consider again the curve $X_3 = V(x_2^2 - x_1^3) \subset \mathbb{A}^2_{\mathbb{C}}$ of Examples 9.21 and 10.3. The Jacobian matrix of the single polynomial $f = x_2^2 - x_1^3$ is

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{pmatrix} = (-3x_1^2 \;\; 2x_2),$$

so it has rank (at least) $2 - \dim X = 1$ exactly if $(x_1, x_2) \in \mathbb{A}^2 \backslash \{0\}$. Hence the Jacobi criterion as in Remark 10.13 (a) does not only reprove our observation from Example 10.3 that the origin is a singular point of $X_3$, but also shows simultaneously that all other points of $X_3$ are smooth.

In the picture on the right we have also drawn the blow-up $\tilde{X}_3$ of $X_3$ at its singular point again. We have seen already that its exceptional set consists of only one point $a \in \tilde{X}_3$. Let us now check that this is a smooth point of $\tilde{X}_3$ — as we would expect from the picture.




In the coordinates $((x_1, x_2), (y_1 : y_2))$ of $\tilde{X}_3 \subset \widetilde{\mathbb{A}^2} \subset \mathbb{A}^2 \times \mathbb{P}^1$, the point $a$ is given as $((0,0),(1:0))$. So around $a$ we can use the affine open chart $U_1 = \{((x_1,x_2),(y_1:y_2)) : y_1 \neq 0\}$ with affine coordinates $x_1$ and $y_2$ as in Example 9.15. By Exercise 9.22 (a), the blow-up $\tilde{X}_3$ is given in these coordinates by

$$\frac{(x_1 y_2)^2 - x_1^3}{x_1^2} = 0, \quad \text{i. e. } g(x_1, y_2) := y_2^2 - x_1 = 0.$$

As the Jacobian matrix

$$\begin{pmatrix} \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial y_2} \end{pmatrix} = (-1 \;\; 2y_2)$$

of this polynomial has rank 1 at every point, the Jacobi criterion tells us that $\tilde{X}_3$ is smooth. In fact, from the defining equation $y_2^2 - x_1 = 0$ we see that on the open subset $U_1$ the curve $\tilde{X}_3$ is just the "standard parabola" tangent to the exceptional set of $\widetilde{\mathbb{A}^2}$ (which is given on $U_1$ by the equation $x_1 = 0$ by the proof of Proposition 9.23).

It is actually a general statement that blowing up makes singular points "nicer", and that successive blow-ups will eventually make all singular points smooth. This process is called *resolution of singularities*. We will not discuss this here in detail, but the following exercise shows an example of this process.

**Exercise 10.16.** For $k \in \mathbb{N}$ let $X_k$ be the affine curve $X_k := V(x_2^2 - x_1^{2k+1}) \subset \mathbb{A}^2$. Show that $X_k$ is not isomorphic to $X_l$ if $k \neq l$.

(Hint: Consider the blow-up of $X_k$ at the origin.)

**Exercise 10.17.** Let $X \subset \mathbb{P}^3$ be the degree-3 Veronese embedding of $\mathbb{P}^1$ as in Exercise 7.32. Of course, $X$ must be smooth since it is isomorphic to $\mathbb{P}^1$. Verify this directly using the projective Jacobi criterion of Exercise 10.12 (b).

**Corollary 10.18.** *The set of singular points of a variety is closed.*

*Proof.* It suffices to prove the statement in the case of an affine variety $X$. We show that the subset $U \subset X$ of smooth points is open. So let $a \in U$. By possibly restricting to a smaller affine subset, we may assume by Remark 10.10 (b) that $X$ is irreducible. Then by Remark 10.13 (a) we know that $U$ is exactly the set of points at which the rank of the Jacobian matrix of generators of $I(X)$ is at least $\operatorname{codim} X$. As this is an open condition (given by the non-vanishing of at least one minor of size $\operatorname{codim} X$), the result follows. □

As the set of smooth points of a variety is open in the Zariski topology by Corollary 10.18, it is very "big" — unless it is empty, of course. Let us quickly study whether this might happen.

**Remark 10.19** (Generic smoothness)**.** Let $f \in K[x_1,\dots,x_n]$ be a non-constant irreducible polynomial, and let $X = V(f) \subset \mathbb{A}^n$. We claim that $X$ has a smooth point, so that the set of smooth points of $X$ is a non-empty open subset by Corollary 10.18, and thus dense by Remark 2.18.

Assume the contrary, i. e. that all points of $X$ are singular. Then by Remark 10.13 (a) the Jacobian matrix of $f$ must have rank 0 at every point, which means that $\frac{\partial f}{\partial x_i}(a) = 0$ for all $a \in X$ and $i = 1,\dots,n$. Hence $\frac{\partial f}{\partial x_i} \in I(V(f)) = (f)$ by the Nullstellensatz. But since $f$ is irreducible and the polynomial $\frac{\partial f}{\partial x_i}$ has smaller degree than $f$ this is only possible if $\frac{\partial f}{\partial x_i} = 0$ for all $i$.

In the case $\operatorname{char} K = 0$ this is already a contradiction to $f$ being non-constant. If $\operatorname{char} K = p$ is positive, then $f$ must be a polynomial in $x_1^p,\dots,x_n^p$, and so

$$f = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} x_1^{pi_1} \cdot \dots \cdot x_n^{pi_n} = \left( \sum_{i_1,\dots,i_n} b_{i_1,\dots,i_n} x_1^{i_1} \cdot \dots \cdot x_n^{i_n} \right)^p,$$

for $p$-th roots $b_{i_1,\dots,i_n}$ of $a_{i_1,\dots,i_n}$. This is a contradiction since $f$ was assumed to be irreducible.

In fact, one can show this "*generic smoothness*" statement for any variety $X$: the set of smooth points of $X$ is dense in $X$. A proof of this result can be found in [H, Theorem I.5.3].

**Example 10.20** (Fermat hypersurfaces)**.** For given $n,d \in \mathbb{N}_{>0}$ consider the *Fermat hypersurface*

$$X := V_p(x_0^d + \dots + x_n^d) \quad \subset \mathbb{P}^n.$$

We want to show that $X$ is smooth for all choices of $n$, $d$, and $K$. For this we use the Jacobian matrix $(dx_0^{d-1} \;\; \cdots \;\; dx_n^{d-1})$ of the given polynomial:

(a) If $\operatorname{char} K \nmid d$ the Jacobian matrix has rank 1 at every point, so $X$ is smooth by Exercise 10.12 (b).

(b) If $p = \operatorname{char} K \,|\, d$ we can write $d = k\,p^r$ for some $r \in \mathbb{N}_{>0}$ and $p \nmid k$. Since

$$x_0^d + \dots + x_n^d = (x_0^k + \dots + x_n^k)^{p^r},$$

we see again that $X = V_p(x_0^k + \dots + x_n^k)$ is smooth by (a).

**Exercise 10.21.** Let $X$ be a projective variety of dimension $n$. Prove:

(a) There is an injective morphism $X \to \mathbb{P}^{2n+1}$.

(b) There is in general no such morphism that is an isomorphism onto its image.

**Exercise 10.22.** Let $n \geq 2$. Prove:

(a) Every smooth hypersurface in $\mathbb{P}^n$ is irreducible.

(b) A general hypersurface in $\mathbb{P}^n_{\mathbb{C}}$ is smooth (and thus by (a) irreducible). More precisely, for a given $d \in \mathbb{N}_{>0}$ the vector space $\mathbb{C}[x_0,\dots,x_n]_d$ has dimension $\binom{n+d}{n}$, and so the space of all homogeneous degree-$d$ polynomials in $x_0,\dots,x_n$ modulo scalars can be identified with the projective space $\mathbb{P}^{\binom{n+d}{n}-1}$. Show that the subset of this projective space of all (classes of) polynomials $f$ such that $f$ is irreducible and $V_p(f)$ is smooth is dense and open.

**Exercise 10.23** (Dual curves)**.** Assume that $\operatorname{char} K \neq 2$, and let $f \in K[x_0,x_1,x_2]$ be a homogeneous polynomial whose partial derivatives $\frac{\partial f}{\partial x_i}$ for $i = 0,1,2$ do not vanish simultaneously at any point of $X = V_p(f) \subset \mathbb{P}^2$. Then the image of the morphism

$$F : X \to \mathbb{P}^2, \;\; a \mapsto \left( \frac{\partial f}{\partial x_0}(a) : \frac{\partial f}{\partial x_1}(a) : \frac{\partial f}{\partial x_2}(a) \right)$$

is called the *dual curve* to $X$.

(a) Find a geometric description of $F$. What does it mean geometrically if $F(a) = F(b)$ for two distinct points $a,b \in X$?

(b) If $X$ is a conic, prove that its dual $F(X)$ is also a conic.

(c) For any five lines in $\mathbb{P}^2$ in general position (what does this mean?) show that there is a unique conic in $\mathbb{P}^2$ that is tangent to all of them.

# 11. The 27 Lines on a Smooth Cubic Surface

As an application of the theory that we have developed so far, we now want to study lines on cubic surfaces in $\mathbb{P}^3$. In Example 0.9, we have already mentioned that every smooth cubic surface has exactly 27 lines on it. Our goal is now to show this, to study the configuration of these lines, and to prove that every smooth cubic surface is birational (but not isomorphic) to $\mathbb{P}^2$. All these results are classical, dating back to the 19th century. They can be regarded historically as being among the first non-trivial statements in projective algebraic geometry.

The results of this chapter will not be needed later on. Most proofs will therefore not be given in every detail here. The aim of this chapter is rather to give an idea of what can be done with our current methods.

For simplicity, we will restrict ourselves to the case of the ground field $K = \mathbb{C}$. By a smooth cubic surface we will always mean a smooth hypersurface in $\mathbb{P}^3$ that can be written as the zero locus of an irreducible homogeneous polynomial of degree 3. Let us start with the discussion of a special case of such a cubic surface: the Fermat cubic $V_p(x_0^3+x_1^3+x_2^3+x_3^3) \subset \mathbb{P}^3$ as in Example 10.20.

**Lemma 11.1.** *The Fermat cubic $X = V_p(x_0^3+x_1^3+x_2^3+x_3^3) \subset \mathbb{P}^3$ contains exactly 27 lines.*

*Proof.* Up to a permutation of coordinates, every line in $\mathbb{P}^3$ is given by two linear equations of the form $x_0 = a_2x_2 + a_3x_3$ and $x_1 = b_2x_2 + b_3x_3$ for suitable $a_2, a_3, b_2, b_3 \in \mathbb{C}$. Such a line lies in $X$ if and only if

$$(a_2x_2+a_3x_3)^3 + (b_2x_2+b_3x_3)^3 + x_2^3 + x_3^3 = 0$$

as a polynomial in $\mathbb{C}[x_2,x_3]$, so by comparing coefficients if and only if

$$\begin{aligned} a_2^3+b_2^3 &= -1, && (1)\\ a_3^3+b_3^3 &= -1, && (2)\\ a_2^2a_3 &= -b_2^2b_3, && (3)\\ a_2a_3^2 &= -b_2b_3^2. && (4) \end{aligned}$$

If $a_2, a_3, b_2, b_3$ are all non-zero, then $(3)^2/(4)$ gives $a_2^3 = -b_2^3$, in contradiction to (1). Hence for a line in the cubic at least one of these numbers must be zero. Again after possibly renumbering the coordinates we may assume that $a_2 = 0$. Then $b_2^3 = -1$ by (1), $b_3 = 0$ by (3), and $a_3^3 = -1$ by (2). Conversely, for such values of $a_2, a_3, b_2, b_3$ the above equations all hold, so that we really obtain a line in the cubic.

We thus obtain 9 lines in $X$ by setting $b_2 = -\omega^j$ and $a_3 = -\omega^k$ for $0 \le j,k \le 2$ and $\omega = \exp(\frac{2\pi i}{3})$ a primitive third root of unity. So by finally allowing permutations of the coordinates we find that there are exactly the following 27 lines on $X$:

$$\begin{aligned} x_0 + x_3\omega^k &= x_1 + x_2\omega^j = 0, \quad 0 \le j,k \le 2,\\ x_0 + x_2\omega^k &= x_3 + x_1\omega^j = 0, \quad 0 \le j,k \le 2,\\ x_0 + x_1\omega^k &= x_3 + x_2\omega^j = 0, \quad 0 \le j,k \le 2. \end{aligned}$$

□

18

**Corollary 11.2.** *Let $X \subset \mathbb{P}^3$ again be the Fermat cubic as in Lemma 11.1.*

(a) *Given any line $L$ in $X$, there are exactly* 10 *other lines in $X$ that intersect $L$.*

(b) *Given any two disjoint lines $L_1, L_2$ in $X$, there are exactly* 5 *other lines in $X$ meeting both $L_1$ and $L_2$.*

*Proof.* As we know all the lines in $X$ by the proof of Lemma 11.1, this is just simple checking. For example, to prove (a) we may assume by permuting coordinates and multiplying them with suitable third roots of unity that $L$ is given by $x_0+x_3 = x_1+x_2 = 0$. The other lines meeting $L$ are then exactly the following:

$$\begin{aligned} &\text{4 lines of the form} \quad x_0+x_3\omega^k = x_1+x_2\omega^j = 0, \quad (j,k) = (1,0),(2,0),(0,1),(0,2), \\ &\text{3 lines of the form} \quad x_0+x_2\omega^j = x_3+x_1\omega^j = 0, \quad 0 \le j \le 2, \\ &\text{3 lines of the form} \quad x_0+x_1\omega^j = x_3+x_2\omega^j = 0, \quad 0 \le j \le 2. \end{aligned}$$

The proof of part (b) is analogous. □

Let us now transfer these results to an arbitrary smooth cubic surface. This is where it gets interesting, since the equations determining the lines lying in the cubic as in the proof of Lemma 11.1 will in general be too complicated to solve them directly. Instead, we will only show that the number of lines in a smooth cubic must be the same for all cubics, so that we can then conclude by Lemma 11.1 that this number must be 27. In other words, we have to consider all smooth cubic surfaces at once.

**Construction 11.3** (The incidence correspondence of lines in smooth cubic surfaces)**.** As in Exercise 10.22 (b), let $\mathbb{P}^{19} = \mathbb{P}^{\binom{3+3}{3}-1}$ be the projective space of all homogeneous degree-3 polynomials in $x_0,x_1,x_2,x_3$ modulo scalars, so that the space of smooth cubic surfaces is a dense open subset $U$ of $\mathbb{P}^{19}$. More precisely, a smooth cubic surface can be given as the zero locus of an irreducible polynomial $f_c := \sum_\alpha c_\alpha x^\alpha = 0$ in multi-index notation, i. e. $\alpha$ runs over all quadruples of non-negative indices $(\alpha_0,\alpha_1,\alpha_2,\alpha_3)$ with $\sum_i \alpha_i = 3$. The corresponding point in $U \subset \mathbb{P}^{19}$ is then the one with homogeneous coordinates $c = (c_\alpha)_\alpha$.

Moreover, we know already that the lines in $\mathbb{P}^3$ are parametrized by the Grassmannian $G(2,4)$ introduced in Chapter 8. We can therefore consider the *incidence correspondence*

$$M := \{(X,L) : L \text{ is a line contained in the smooth cubic } X\} \quad \subset U \times G(2,4).$$

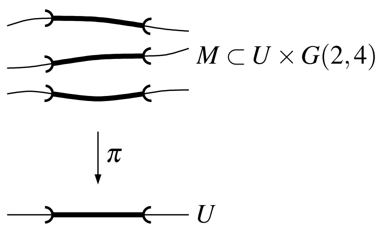
Note that it comes with a natural projection map $\pi : M \to U$ sending a pair $(X,L)$ to $X$, and that the number of lines in a cubic surface is just its number of inverse images under $\pi$.

To show that this number of inverse images is constant on $U$, we will pass from the algebraic to the analytic category and prove the following statement.

**Lemma 11.4.** *With notations as in Construction 11.3, the incidence correspondence $M$ is...*

(a) *closed in the Zariski topology of $U \times G(2,4)$;*

(b) *locally in the classical topology the graph of a continuously differentiable function $U \to G(2,4)$, as shown in the picture on the right.*




*Proof.* Let $(X,L) \in M$. By a linear change of coordinates we can assume that $L$ is given by the equations $x_2 = x_3 = 0$. Locally around this point $L \in G(2,4)$ in the Zariski topology we can use the affine coordinates on the Grassmannian as in Construction 8.15, namely $a_2,a_3,b_2,b_3 \in \mathbb{C}$ corresponding to the line in $\mathbb{P}^3$ spanned by the rows of the matrix

$$\begin{pmatrix} 1 & 0 & a_2 & a_3 \\ 0 & 1 & b_2 & b_3 \end{pmatrix},$$

with the point $(a_2,a_3,b_2,b_3) = (0,0,0,0)$ corresponding to $L$. On the space $U$ of smooth cubic surfaces we use the coordinates $(c_\alpha)_\alpha$ as in Construction 11.3. In these coordinates $(a,b,c) =$

$(a_2,a_3,b_2,b_3,c_\alpha)$ on $U \times G(2,4)$, the incidence correspondence $M$ is then given by

$$\begin{aligned}
(a,b,c) \in M &\iff f_c(s(1,0,a_2,a_3)+t(0,1,b_2,b_3)) = 0 \text{ for all } s,t \\
&\iff \sum_\alpha c_\alpha s^{\alpha_0} t^{\alpha_1} (sa_2+tb_2)^{\alpha_2}(sa_3+tb_3)^{\alpha_3} = 0 \text{ for all } s,t \\
&\iff: \sum_i s^i t^{3-i} F_i(a,b,c) = 0 \text{ for all } s,t \\
&\iff F_i(a,b,c) = 0 \text{ for } 0 \le i \le 3.
\end{aligned}$$

This shows (a), since $F_0,\dots,F_3$ are polynomial functions in $a,b,c$. The claim of (b) is that these four equations determine $(a_2,a_3,b_2,b_3)$ locally around the origin in the classical topology in terms of $c$. Of course, we will prove this with (the complex version of) the Implicit Function Theorem [G2, Proposition 27.9]. All we have to show is therefore that the Jacobian matrix $J := \frac{\partial(F_0,F_1,F_2,F_3)}{\partial(a_2,a_3,b_2,b_3)}$ is invertible at $a = b = 0$.

So let us compute this Jacobian matrix. Note that

$$\begin{aligned}
\frac{\partial}{\partial a_2}\Big(\sum_i s^i t^{3-i} F_i\Big)\Big|_{a=b=0} &= \frac{\partial}{\partial a_2} f_c(s,t,sa_2+tb_2,sa_3+tb_3)\Big|_{a=b=0} \\
&= s\,\frac{\partial f_c}{\partial x_2}(s,t,0,0).
\end{aligned}$$

The $(s,t)$-coefficients of this polynomial are the first column in the matrix $J$. Similarly, the other columns are obviously $s\,\frac{\partial f_c}{\partial x_3}(s,t,0,0)$, $t\,\frac{\partial f_c}{\partial x_2}(s,t,0,0)$, and $t\,\frac{\partial f_c}{\partial x_3}(s,t,0,0)$. Hence, if the matrix $J$ was not invertible, there would be a relation

$$(\lambda_2 s+\mu_2 t)\frac{\partial f_c}{\partial x_2}(s,t,0,0) + (\lambda_3 s+\mu_3 t)\frac{\partial f_c}{\partial x_3}(s,t,0,0) = 0$$

identically in $s,t$, with $(\lambda_2,\mu_2,\lambda_3,\mu_3) \in \mathbb{C}^4\backslash\{0\}$. As homogeneous polynomials in two variables always decompose into linear factors, this means that $\frac{\partial f_c}{\partial x_2}(s,t,0,0)$ and $\frac{\partial f_c}{\partial x_3}(s,t,0,0)$ must have a common linear factor, i. e. that there is a point $p = (p_0,p_1,0,0) \in L$ with $\frac{\partial f_c}{\partial x_2}(p) = \frac{\partial f_c}{\partial x_3}(p) = 0$.

But as the line $L$ lies in the cubic $V_p(f_c)$, we also have $f_c(s,t,0,0) = 0$ for all $s,t$. Differentiating this with respect to $s$ and $t$ gives $\frac{\partial f_c}{\partial x_0}(p) = 0$ and $\frac{\partial f_c}{\partial x_1}(p) = 0$, respectively. Hence all partial derivatives of $f_c$ vanish at $p \in L \subset X$. By the Jacobi criterion of Exercise 10.12 (b) this means that $p$ would be a singular point of $X$, in contradiction to our assumption. Hence $J$ must be invertible, and part (b) of the lemma follows. □

**Corollary 11.5.** *Every smooth cubic surface contains exactly* 27 *lines.*

*Proof.* In this proof we will work with the classical topology throughout. Let $X \in U$ be a fixed smooth cubic, and let $L \subset \mathbb{P}^3$ be an arbitrary line. We distinguish two cases:

Case 1: If $L$ lies in $X$, Lemma 11.4 (b) shows that there is an open neighborhood $V_L \times W_L$ of $(X,L)$ in $U \times G(2,4)$ in which the incidence correspondence $M$ is the graph of a continuously differentiable function. In particular, every cubic in $V_L$ contains exactly one line in $W_L$.

Case 2: If $L$ does not lie in $X$ there is an open neighborhood $V_L \times W_L$ of $(X,L)$ such that no cubic in $V_L$ contains any line (since the incidence correspondence is closed by Lemma 11.4 (a)).

Now let $L$ vary. As the Grassmannian $G(2,4)$ is projective, and hence compact, there are finitely many $W_L$ that cover $G(2,4)$. Let $V$ be the intersection of the corresponding $V_L$, which is then again an open neighborhood of $X$. By construction, in this neighborhood $V$ all cubic surfaces have the same number of lines (namely the number of $W_L$ coming from case 1). As this argument holds for any cubic, we conclude that the number of lines contained in a cubic surface is a locally constant function on $U$.

To see that this number is also globally constant, it therefore suffices to show that $U$ is connected. But this follows from Exercise 10.22 (b): we know that $U$ is the complement of a proper Zariski-closed subset in $\mathbb{P}^{19}$. But as such a closed subset has complex codimension at least 1 and hence real codimension at least 2, taking this subset away from the smooth and connected space $\mathbb{P}^{19}$ leaves us again with a connected space. □

**Remark 11.6.**

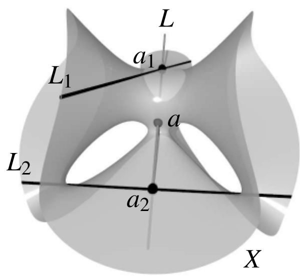
(a) In topological terms, the argument of the proof of Corollary 11.5 says that the map $\pi : M \to U$ of Construction 11.3 is a 27-sheeted covering map.

(b) Applying the methods of Lemma 11.4 and Corollary 11.5 to suitable incidence correspondences involving two resp. three lines in cubic surfaces, one can show similarly that the statements of Corollary 11.2 hold for an arbitrary smooth cubic surface $X$ as well: there are exactly 10 lines in $X$ meeting any given one, and exactly 5 lines in $X$ meeting any two disjoint given ones.

Note that a cubic surface $X$ is clearly not isomorphic to $\mathbb{P}^2$: by Remark 11.6 (b) there are two disjoint lines on $X$, whereas in $\mathbb{P}^2$ any two curves intersect by Exercise 6.32 (b). However, we will now see that $X$ is birational to $\mathbb{P}^2$, and that it is in fact isomorphic to a blow-up of $\mathbb{P}^2$ at six points.

**Proposition 11.7.** *Any smooth cubic surface is birational to $\mathbb{P}^2$.*

*Proof.* By Remark 11.6 (b) there are two disjoint lines $L_1, L_2 \subset X$. The following mutually inverse rational maps $X \dashrightarrow L_1 \times L_2$ and $L_1 \times L_2 \dashrightarrow X$ show that $X$ is birational to $L_1 \times L_2 \cong \mathbb{P}^1 \times \mathbb{P}^1$, and hence to $\mathbb{P}^2$:




"$X \dashrightarrow L_1 \times L_2$": By Exercise 6.30, for every point $a$ not on $L_1$ or $L_2$ there is a unique line $L$ in $\mathbb{P}^3$ through $L_1$, $L_2$, and $a$. Take the rational map from $X$ to $L_1 \times L_2$ sending $a$ to $(a_1, a_2) := (L_1 \cap L, L_2 \cap L)$, which is obviously well-defined away from $L_1 \cup L_2$.

"$L_1 \times L_2 \dashrightarrow X$": Map any pair of points $(a_1, a_2) \in L_1 \times L_2$ to the third intersection point of $X$ with the line $L$ through $a_1$ and $a_2$. This is well-defined whenever $L$ is not contained in $X$. □

**Proposition 11.8.** *Any smooth cubic surface is isomorphic to $\mathbb{P}^2$ blown up in 6 (suitably chosen) points.*

*Proof.* We will only sketch the proof. Let $X$ be a smooth cubic surface, and let $f : X \dashrightarrow L_1 \times L_2 \cong \mathbb{P}^1 \times \mathbb{P}^1$ be the rational map as in the proof of Proposition 11.7.

First of all we claim that $f$ is actually a morphism. To see this, note that there is a different description for $f$: if $a \in X \backslash L_1$, let $H$ be the unique plane in $\mathbb{P}^3$ that contains $L_1$ and $a$, and set $f_2(a) = H \cap L_2$. If one defines $f_1(a)$ similarly, then $f(a) = (f_1(a), f_2(a))$. Now if the point $a$ lies on $L_1$, let $H$ be the tangent plane to $X$ at $a$, and again set $f_2(a) = H \cap L_2$. Extending $f_1$ similarly, one can show that this extends $f = (f_1, f_2)$ to a well-defined morphism $X \to \mathbb{P}^1 \times \mathbb{P}^1$ on all of $X$.

Now let us investigate where the inverse map $\mathbb{P}^1 \times \mathbb{P}^1 \dashrightarrow X$ is not well-defined. As already mentioned in the proof of Proposition 11.7, this is the case if the point $(a_1, a_2) \in L_1 \times L_2$ is chosen so that $\overline{a_1 a_2} \subset X$. In this case, the whole line $\overline{a_1 a_2}$ will be mapped to $(a_1, a_2)$ by $f$, and it can be checked that $f$ is actually locally the blow-up of this point. By Remark 11.6 (b) there are exactly 5 such lines $\overline{a_1 a_2}$ on $X$. Hence $X$ is the blow-up of $\mathbb{P}^1 \times \mathbb{P}^1$ in 5 points, i. e. by Lemma 9.28 the blow-up of $\mathbb{P}^2$ in 6 suitably chosen points. □

**Remark 11.9.** It is interesting to see the lines on a cubic surface $X$ in the picture of Proposition 11.8 in which we think of $X$ as a blow-up of $\mathbb{P}^2$ in 6 points. It turns out that the 27 lines correspond to the following curves that we know already (and that are all isomorphic to $\mathbb{P}^1$):

- the 6 exceptional hypersurfaces,

- the strict transforms of the $\binom{6}{2} = 15$ lines through two of the blown-up points,
- the strict transforms of the $\binom{6}{5} = 6$ conics through five of the blown-up points (see Exercise 7.31 (c)).

In fact, it is easy to check by the above explicit description of the isomorphism of $X$ with the blow-up of $\mathbb{P}^2$ that these curves on the blow-up actually correspond to lines on the cubic surface.

It is also interesting to see again in this picture that every such “line” meets 10 of the other “lines”, as mentioned in Remark 11.6 (b):

- Every exceptional hypersurface intersects the 5 lines and the 5 conics that pass through this blown-up point.
- Every line through two of the blown-up points meets
  - the 2 exceptional hypersurfaces of the blown-up points,
  - the $\binom{4}{2} = 6$ lines through two of the four remaining points,
  - the 2 conics through the four remaining points and one of the blown-up points.
- Every conic through five of the blown-up points meets the 5 exceptional hypersurfaces at these points, as well as the 5 lines through one of these five points and the remaining point.

**Exercise 11.10.** As in Exercise 10.22 (b) let $U \subset \mathbb{P}^{\binom{4+5}{4}-1} = \mathbb{P}^{125}$ be the set of smooth (3-dimensional) hypersurfaces of degree 5 in $\mathbb{P}^4$. Prove:

(a) The incidence correspondence
$$\{(X,L) \in U \times G(2,5) : L \text{ is a line contained in } X\}$$
is smooth of dimension 125, i. e. of the same dimension as $U$.

(b) Although (a) suggests that a smooth hypersurface of degree 5 in $\mathbb{P}^4$ contains only finitely many lines, the Fermat hypersurface $V_p(x_0^5 + \cdots + x_4^5) \subset \mathbb{P}^4$ contains infinitely many lines.

# 12. Hilbert Polynomials and Bézout's Theorem

After our study of smooth cubic surfaces in the last chapter, let us now come back to the general theory of algebraic geometry. Our main goal of this chapter will be to determine the number of intersection points of given varieties (in case this number is finite). For example, let $X$ and $Y$ be two plane curves, with (principal) ideals generated by two polynomials $f$ and $g$, respectively. If they do not have a common irreducible component, their intersection will be zero-dimensional, and we can ask for the number of points in $X \cap Y$. We will see in Bézout's Theorem as in Corollaries 12.20 (b) and 12.26 (b) that this number of points is at most $\deg f \cdot \deg g$, and that we can even make this an equality if we count the points with suitable multiplicities. We have seen a special case of this already in Exercise 4.13, where one of the two curves was a line or a conic.

In particular, this statement means that the number of points in $X \cap Y$ (counted with multiplicities) depends only on the degrees of the defining polynomials, and not on the polynomials themselves. One can view this as a direct generalization of the statement that a degree-$d$ polynomial in one variable always has $d$ zeroes, again counted with multiplicities.

In order to set up a suitable framework for Bézout's Theorem, we have to take note of the following two technical observations:

- As mentioned above, we have to define suitable intersection multiplicities, e. g. for two plane curves $X$ and $Y$. We have motivated in Remark 1.27 already that such multiplicities are encoded in the (possibly non-radical) ideal $I(X) + I(Y)$. Most constructions in this chapter are therefore based on ideals rather than on varieties, and consequently commutative algebra will play a somewhat greater role than before.
- The simplest example of Bézout's Theorem is that two distinct lines in the plane always meet in one point. This would clearly be false in the affine setting, where two such lines might be parallel. We therefore have to work with projective varieties that can have intersection points at infinity in such cases. 19

Taking these two points into account, we see that our main objects of study will have to be homogeneous ideals in polynomial rings. The central concept that we will need is the Hilbert function of such an ideal.

**Definition 12.1** (Hilbert functions)**.**

(a) Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal. Then $K[x_0,\dots,x_n]/I$ is a finite-dimensional graded $K$-algebra by Lemma 6.10 (c). We can therefore define the function

$$h_I : \mathbb{N} \to \mathbb{N},\ d \mapsto \dim_K K[x_0,\dots,x_n]_d / I_d$$

encoding the dimensions of the graded parts of this quotient algebra. It is called the **Hilbert function** of $I$.

(b) For a projective variety $X \subset \mathbb{P}^n$ we set $h_X := h_{I(X)}$, so that

$$h_X : \mathbb{N} \to \mathbb{N},\ d \mapsto \dim_K S(X)_d,$$

where $S(X) = K[x_0,\dots,x_n]/I(X)$ is the homogeneous coordinate ring of $X$ as in Construction 6.18. We call $h_X$ the Hilbert function of $X$.

**Remark 12.2.** Note that the Hilbert function of an ideal is invariant under projective automorphisms as in Example 7.6 (a): an invertible matrix corresponding to an automorphism $\mathbb{P}^n \to \mathbb{P}^n$ also defines an isomorphism $\mathbb{A}^{n+1} \to \mathbb{A}^{n+1}$, and hence by Corollary 4.8 an isomorphism $K[x_0,\dots,x_n] \to K[x_0,\dots,x_n]$ of $K$-algebras. As this isomorphism respects the grading, any ideal has the same Hilbert function as its image under this isomorphism.

**Example 12.3.**

(a) The Hilbert function of $\mathbb{P}^n$ is given by $h_{\mathbb{P}^n}(d) = \dim_K K[x_0,\dots,x_n]_d = \binom{n+d}{n}$ for $d \in \mathbb{N}$.

(b) Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal with $V_p(I) = \emptyset$. Then $\sqrt{I} = (x_0,\dots,x_n)$ or $\sqrt{I} = (1)$ by the projective Nullstellensatz of Proposition 6.22. In both cases we have $x_i^{k_i} \in I$ for suitable $k_i \in \mathbb{N}$ for all $i$. This means that all monomials of degree at least $k := k_0 + \cdots + k_n$ are contained in $I$. Hence $I_d = K[x_0,\dots,x_n]_d$ for all $d \geq k$, or in other words

$$h_I(d) = 0 \quad \text{for almost all } d \in \mathbb{N},$$

where as usual we use the term "*almost all*" for "all but finitely many".

(c) Let $X = \{a\} \subset \mathbb{P}^n$ be a single point. To compute its Hilbert function we may assume by Remark 12.2 that this point is $a = (1:0:\cdots:0)$, so that its ideal is $I(a) = (x_1,\dots,x_n)$. Then $S(X) = K[x_0,\dots,x_n]/I(a) \cong K[x_0]$, and hence

$$h_X(d) = 1 \quad \text{for all } d \in \mathbb{N}.$$

**Exercise 12.4.** Compute the Hilbert function of...

(a) the ideal $(x_0^2x_1^2, x_0^3) \trianglelefteq K[x_0,x_1]$;

(b) two intersecting lines in $\mathbb{P}^3$;

(c) two non-intersecting lines in $\mathbb{P}^3$.

In order to work with Hilbert functions it is convenient to adopt the language of *exact sequences* from commutative algebra. The only statement that we will need about them is the following.

**Lemma and Definition 12.5** (Exact sequences)**.** *Let $f: U \to V$ and $g: V \to W$ be linear maps of vector spaces over $K$. Assume that $f$ is injective, $g$ is surjective, and that* $\operatorname{im} f = \ker g$*. These assumptions are usually summarized by saying that*

$$0 \longrightarrow U \xrightarrow{f} V \xrightarrow{g} W \longrightarrow 0$$

*is an **exact sequence*** [G5, Chapter 4].

*Then* $\dim_K V = \dim_K U + \dim_K W$.

*Proof.* This is just standard linear algebra: we have

$$\dim_K V = \dim_K \ker g + \dim_K \operatorname{im} g = \dim_K \operatorname{im} f + \dim_K \operatorname{im} g = \dim_K U + \dim_K W,$$

with the last equation following since $f$ is injective and $g$ is surjective. □

**Proposition 12.6.** *For any two homogeneous ideals $I, J \trianglelefteq K[x_0,\dots,x_n]$ we have*

$$h_{I\cap J} + h_{I+J} = h_I + h_J.$$

*Proof.* Set $R = K[x_0,\dots,x_n]$. It is easily checked that

$$\begin{array}{ccccccccc} 0 & \longrightarrow & R/(I\cap J) & \longrightarrow & R/I \times R/J & \longrightarrow & R/(I+J) & \longrightarrow & 0 \\ & & \overline{f} & \longmapsto & (\overline{f},\overline{f}) & & & & \\ & & & & (\overline{f},\overline{g}) & \longmapsto & \overline{f}-\overline{g} & & \end{array}$$

is an exact sequence. Taking its degree-$d$ part and applying Lemma 12.5 gives the desired result. □

**Example 12.7.**

(a) Let $X$ and $Y$ be disjoint projective varieties in $\mathbb{P}^n$. Then $I(X) \cap I(Y) = I(X \cup Y)$ by Remark 6.23. Moreover, by the same remark the ideal $I(X) + I(Y)$ has empty zero locus since $V(I(X)+I(Y)) = V(I(X)) \cap V(I(Y)) = X \cap Y = \emptyset$, and hence its Hilbert function is almost everywhere zero by Example 12.3 (b). Proposition 12.6 thus implies that

$$h_{X\cup Y}(d) = h_X(d) + h_Y(d) \quad \text{for almost all } d \in \mathbb{N}.$$

In particular, this means by Example 12.3 (c) that for a finite set $X = \{a_1, \dots, a_r\}$ of $r$ points we have

$$h_X(d) = r \quad \text{for almost all } d \in \mathbb{N}.$$

(b) Let $I = (x_1^2) \trianglelefteq K[x_0,x_1]$. It is a non-radical ideal whose projective zero locus consists of the single point $(1:0) \in \mathbb{P}^1$. In fact, it can be viewed as an ideal describing this point "with multiplicity 2" as in Remark 1.27.

The Hilbert function remembers this multiplicity: as $K[x_0,x_1]_d/I_d$ has basis $x_0^d$ and $x_0^{d-1}x_1$ for $d \geq 1$, we see that $h_I(d) = 2$ for almost all $d$, in the same way as for the Hilbert function of two distinct points as in (a).

(c) Let $X \subset \mathbb{P}^2$ be the union of three points lying on a line. Then there is a homogeneous linear polynomial in $K[x_0,x_1,x_2]$ vanishing on $X$, so that $\dim_K I(X)_1 = 1$. Hence $h_X(1) = \dim_K K[x_0,x_1,x_2]_1/I(X)_1 = 3-1 = 2$. On the other hand, if $X$ consists of three points not on a line, then no linear polynomial vanishes on $X$, and consequently $h_X(1) = \dim_K K[x_0,x_1,x_2]_1/I(X)_1 = 3-0 = 3$. So in particular, we see that in contrast to Remark 12.2 the Hilbert function is not invariant under arbitrary isomorphisms, since any set of three points is isomorphic to any other such set.

Together with the result of (a), for a finite set $X \subset \mathbb{P}^n$ we can say intuitively that $h_X(d)$ encodes the *number* of points in $X$ for large values of $d$, whereas it gives some information on the relative *position* of these points for small values of $d$.

Note that the intersection $X \cap Y$ of two varieties $X$ and $Y$ corresponds to the sum of their ideals. To obtain a formula for the number of points in $X \cap Y$ we therefore have to compute the Hilbert functions of sums of ideals. The following lemma will help us to do this in the case when one of the ideals is principal.

**Lemma 12.8.** *Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal, and let $f \in K[x_0,\dots,x_n]$ be a homogeneous polynomial of degree $e$. Assume that there is a number $d_0 \in \mathbb{N}$ with the following property:*

*for all homogeneous $g \in K[x_0,\dots,x_n]$ of degree at least $d_0$ with $fg \in I$ we have $g \in I$.*

*Then $h_{I+(f)}(d) = h_I(d) - h_I(d-e)$ for almost all $d \in \mathbb{N}$.*

*Proof.* Let $R = K[x_0,\dots,x_n]$. There is an exact sequence

$$0 \longrightarrow R_{d-e}/I_{d-e} \xrightarrow{\cdot f} R_d/I_d \longrightarrow R_d/(I+(f))_d \longrightarrow 0$$

for all $d$ with $d - e \geq d_0$, where the second map is just the quotient map. In fact, it is obvious that this quotient map is surjective, and that its kernel is exactly the image of the first map. The injectivity of the first map is precisely the assumption of the lemma.

The desired statement now follows immediately from Lemma 12.5. □

Before we can apply this lemma, we have to analyze the geometric meaning of the somewhat technical assumption that $fg \in I$ implies $g \in I$ for all polynomials $g$ (of sufficiently large degree).

**Remark 12.9.** Assume that $I = I(X)$ is the (radical) ideal of a projective variety $X$. Consider the irreducible decomposition $X = X_1 \cup \cdots \cup X_r$ of $X$, so that $I = I(X_1) \cap \cdots \cap I(X_r)$ by Remark 6.23 (c).

We claim that the assumption of Lemma 12.8 is then satisfied if $f$ does not vanish identically on any $X_i$. In fact, in this case $f$ is non-zero in the integral domain $S(X_i)$ for all $i$ (see Exercise 6.31 (b)). Hence $gf \in I$, i. e. $gf = 0 \in S(X_i)$, implies $g = 0 \in S(X_i)$ for all $i$, and thus $g \in I$.

If $I$ is not radical, a similar statement holds — but in order for this to work we need to be able to decompose $I$ as an intersection of ideals corresponding to irreducible varieties again. This so-called primary decomposition of $I$ is one of the main topics in the Commutative Algebra class [G5, Chapter 8]. We will therefore just quote the results that we are going to need.

**Remark 12.10** (*Primary decompositions*)**.** Let $I \trianglelefteq K[x_0,\dots,x_n]$ be an arbitrary ideal. Then there is a decomposition

$$I = I_1 \cap \cdots \cap I_r$$

into *primary ideals* $I_1,\dots,I_r$, which means by definition that $gf \in I_i$ implies $g \in I_i$ or $f \in \sqrt{I_i}$ for all $i$ and all polynomials $f,g \in K[x_0,\dots,x_n]$ [G5, Definition 8.9 and Proposition 8.16]. Moreover, this decomposition satisfies the following properties:

(a) The zero loci $V_a(I_i)$ are irreducible: by Proposition 2.9 and the Nullstellensatz, this is the same as saying that $K[x_0,\dots,x_n]/\sqrt{I_i}$ is an integral domain. So assume that $f$ and $g$ are two polynomials with $gf \in \sqrt{I_i}$. Then $g^k f^k \in I_i$ for some $k \in \mathbb{N}$. But this implies that $g^k \in I_i$ or $f^k \in \sqrt{I_i}$ since $I_i$ is primary, and hence that $g \in \sqrt{I_i}$ or $f \in \sqrt{\sqrt{I_i}} = \sqrt{I_i}$.

(b) Applying Remark 1.25 (c) to our decomposition, we see that

$$V_a(I) = V_a(I_1) \cup \cdots \cup V_a(I_r).$$

In particular, by (a) all irreducible components of $V_a(I)$ must be among the varieties $V_a(I_1),\dots,V_a(I_r)$. Moreover, we can assume that no two of these varieties coincide: if $V_a(I_i) = V_a(I_j)$ for some $i \neq j$, i. e. by the Nullstellensatz $\sqrt{I_i} = \sqrt{I_j}$, we can replace the two ideals $I_i$ and $I_j$ by the single ideal $I_i \cap I_j$ in the decomposition, which is easily seen to be primary again.

However, it may well happen that there are (irreducible) varieties among $V_a(I_1),\dots,V_a(I_r)$ that are strictly contained in an irreducible component of $V_a(I)$ [G5, Example 8.23]. These varieties are usually called the *embedded components* of $I$. In the primary decomposition, the ideals corresponding to the irreducible components are uniquely determined, whereas the ones corresponding to the embedded components are usually not [G5, Example 8.23 and Proposition 8.34].

Using these primary decompositions, we can now show for a homogeneous ideal $I \trianglelefteq K[x_0,\dots,x_n]$ that, by a suitable homogeneous linear change of coordinates, we can always achieve that $f = x_0$ satisfies the condition of Lemma 12.8. In fact, assume that $g$ is a homogeneous polynomial such that $gx_0 \in I_i$ for all $i$. We distinguish two cases:

- If $V_a(I_i) \subset \{0\}$ then $\sqrt{I_i} \supset (x_0,\dots,x_n)$ by the Nullstellensatz. Hence $K[x_0,\dots,x_n]_d \subset I_i$ for large $d$ in the same way as in Example 12.3 (b), which means that $g \in I_i$ if the degree of $g$ is big enough.
- If $V_a(I_i) \not\subset \{0\}$ a general homogeneous linear change of coordinates will assure that $V_a(I_i)$ is not contained in the hypersurface $V_a(x_0)$. Then $x_0$ is not identically zero on $V_a(I_i)$, so that $x_0 \notin I_a(V_a(I_i)) = \sqrt{I_i}$. Since $I_i$ is primary, we conclude that $g \in I_i$.

Let us now come back to our study of Hilbert functions. We have already seen that the important information in $h_I$ concerning the number of intersection points of varieties is contained in its values $h_I(d)$ for large $d$. We therefore have to study the behavior of $h_I(d)$ as $d \to \infty$. The central result in this direction is that the Hilbert function is eventually polynomial, with particularly the degree and the leading coefficient of this polynomial deserving special attention.

**Proposition and Definition 12.11** (Hilbert polynomials)**.** *Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal. Then there is a unique polynomial $\chi_I \in \mathbb{Q}[d]$ such that $\chi_I(d) = h_I(d)$ for almost all $d \in \mathbb{N}$. Moreover,*

(a) *The degree of $\chi_I$ is $m := \dim V_p(I)$.*

(b) *If $V_p(I) \neq \emptyset$, the leading coefficient of $\chi_I$ is $\frac{1}{m!}$ times a positive integer.*

*The polynomial $\chi_I$ is called the **Hilbert polynomial** of $I$. For a projective variety $X \subset \mathbb{P}^n$ we set $\chi_X := \chi_{I(X)}$.*

*Proof.* It is obvious that a polynomial with infinitely many fixed values is unique. So let us prove the existence of $\chi_I$ by induction on $m = \dim V_p(I)$. The start of the induction follows from Example 12.3 (b): for $V_p(I) = \emptyset$ we obtain the zero polynomial $\chi_I$.

Let us now assume that $V_p(I) \neq \emptyset$. By a homogeneous linear change of coordinates (which does not affect the Hilbert function by Remark 12.2) we can assume that the polynomial $x_0$ does not vanish identically on any irreducible component of $V_p(I)$. Hence $\dim V_p(I + (x_0)) \leq m-1$, and so by our induction on $m$ we know that $d \mapsto h_{I+(x_0)}(d)$ is a polynomial of degree at most $m-1$ for large $d$. For reasons that will be apparent later, let us choose $\binom{d}{0}, \ldots, \binom{d}{m-1}$ as a basis of the vector space of polynomials in $d$ of degree at most $m-1$, so that for suitable $c_0, \ldots, c_{m-1} \in \mathbb{Q}$ we can write

$$h_{I+(x_0)}(d) = \sum_{i=0}^{m-1} c_i \binom{d}{i} \quad \text{for almost all } d \in \mathbb{N}.$$

Moreover, by Remark 12.10 we can assume that Lemma 12.8 is applicable for $f = x_0$, so that

$$h_I(d) - h_I(d-1) = h_{I+(x_0)}(d) = \sum_{i=0}^{m-1} c_i \binom{d}{i} \quad \text{for almost all } d \in \mathbb{N}. \tag{1}$$

We will now show by induction on $d$ that there is a constant $c \in \mathbb{Q}$ such that

$$h_I(d) = c + \sum_{i=0}^{m-1} c_i \binom{d+1}{i+1} \quad \text{for almost all } d \in \mathbb{N}. \tag{2}$$

The start of the induction is trivial, since we can always adjust $c$ so that this equation holds at a single value of $d$ (chosen so that (1) holds for all larger values of $d$). But then for all larger $d$ we have

$$h_I(d+1) = h_I(d) + \sum_{i=0}^{m-1} c_i \binom{d+1}{i} = c + \sum_{i=0}^{m-1} c_i \left( \binom{d+1}{i+1} + \binom{d+1}{i} \right) = c + \sum_{i=0}^{m-1} c_i \binom{d+2}{i+1}$$

by (1) and the induction assumption. As the right hand side of (2) is a polynomial in $d$ (of degree at most $m$), this proves the existence of $\chi_I$.

Finally, if $V_p(I) \neq \emptyset$ let us show that the $d^m$-coefficient of $\chi_I$ is $\frac{1}{m!}$ times a positive integer, thus proving the additional statements (a) and (b). We will do this again by induction on $m$.

- $m = 0$: In this case $\chi_I$ is a constant, and it is clearly a non-negative integer, since it is by definition the dimension of $K[x_0, \ldots, x_n]_d / I_d$ for large $d$. Moreover, it cannot be zero, since otherwise $I_d = K[x_0, \ldots, x_n]_d$ for some $d$, which implies $x_i^d \in I_d$ for all $i$ and thus $V_p(I) = \emptyset$.
- $m > 0$: In this case $V_p(I)$ has an irreducible component of dimension $m$ (and none of bigger dimension). In our proof above, the zero locus of $x_0$ on this component is non-empty by Exercise 6.32 (b), and of dimension $m-1$ by Proposition 2.25 (c). Hence $\dim V_p(I + (x_0)) = m-1$, and so by induction $\chi_{I+(x_0)}$ is a polynomial of degree $m-1$, with $(m-1)!$ times the leading coefficient being a positive integer. But note that in the proof above this integer is just $c_{m-1}$, which is then also $m!$ times the leading coefficient of $\chi_I$ by (2). □

**Remark 12.12.** Of course, all our statements concerning the values of the Hilbert function $d \mapsto h_I(d)$ at large values of $d$ can be transferred immediately to the Hilbert polynomial. For example, Example 12.3 implies that $\chi_I = 0$ (as a polynomial) if $V_p(I) = \emptyset$, and $\chi_I = 1$ if $I$ is the ideal of a point. Similar statements hold for Proposition 12.6, Example 12.7, and Lemma 12.8.

**Definition 12.13** (Degree)**.** Let $I \trianglelefteq K[x_0, \ldots, x_n]$ be a homogeneous ideal with non-empty projective zero locus, and let $m = \dim V_p(I)$. Then $m!$ times the leading coefficient of $\chi_I$, which is a positive integer by Proposition 12.11, is called the **degree** $\deg I$ of $I$. The reason for this name will become clear in Example 12.17.

For a projective variety $X$, its degree is defined as $\deg X := \deg I(X)$.

**Example 12.14.**

(a) The degree of $\mathbb{P}^n$ is $n!$ times the $d^n$-coefficient of $\dim_K K[x_0, \ldots, x_n]_d = \binom{n+d}{n}$, i. e. $\deg \mathbb{P}^n = 1$. By Example 12.3 (c), the degree of a single point is 1 as well.

More generally, if $X \subset \mathbb{P}^n$ is any linear space then its homogeneous coordinate ring is isomorphic to a polynomial ring, and hence as above $\deg X = 1$ again.

(b) Let $X$ and $Y$ be projective varieties in $\mathbb{P}^n$ of the same dimension $m$, and assume that they do not have a common irreducible component. Then the zero locus $X \cap Y$ of $I(X) + I(Y)$ has dimension smaller than $m$, so that $\chi_{I(X)+I(Y)}$ has degree less than $m$. Moreover, we have $I(X) \cap I(Y) = I(X \cup Y)$, and hence considering $m!$ times the degree-$m$ coefficients in the Hilbert polynomials of $I(X) \cap I(Y)$, $I(X) + I(Y)$, $I(X)$, and $I(Y)$ yields

$$\deg(X \cup Y) = \deg X + \deg Y$$

by Proposition 12.6 (which of course holds for the Hilbert polynomials as well as for the Hilbert functions).

(c) Let $I \trianglelefteq K[x_0, \dots, x_n]$ be a homogeneous ideal with finite zero locus, consisting of $r$ points. Then $\chi_{\sqrt{I}} = r$ by Example 12.7 (a). But $\sqrt{I} \supset I$ also implies $\chi_{\sqrt{I}} \le \chi_I$, and so we conclude that

$$\deg I = \chi_I \ge \chi_{\sqrt{I}} = r.$$

In fact, in Corollary 12.26 we will refine this statement by interpreting $\deg I$ as a sum of multiplicities for each point in $V_p(I)$, with each of these multiplicities being a positive integer.

**Exercise 12.15.**

(a) Show that the degree of the Segre embedding of $\mathbb{P}^m \times \mathbb{P}^n$ is $\binom{n+m}{n}$.

(b) Show that the degree of the degree-$d$ Veronese embedding of $\mathbb{P}^n$ is $d^n$.

20

We are now ready to prove the main result of this chapter.

**Proposition 12.16 (Bézout's Theorem).** *Let $X \subset \mathbb{P}^n$ be a projective variety of dimension at least 1, and let $f \in K[x_0, \dots, x_n]$ be a homogeneous polynomial that does not vanish identically on any irreducible component of $X$. Then*

$$\deg(I(X) + (f)) = \deg X \cdot \deg f.$$

*Proof.* Let $m = \dim X$. By Definition 12.13, the Hilbert polynomial of $X$ is given by

$$\chi_X(d) = \frac{\deg X}{m!} d^m + a\, d^{m-1} + (\text{terms of degree less than } m-1)$$

for some $a \in \mathbb{Q}$. So by Remark 12.9 we can apply Lemma 12.8 and obtain with $e := \deg f$

$$\begin{aligned}
\chi_{I(X)+(f)}(d) &= \chi_X(d) - \chi_X(d-e) \\
&= \frac{\deg X}{m!} (d^m - (d-e)^m) + a\,(d^{m-1} - (d-e)^{m-1}) + (\text{terms of degree less than } m-1) \\
&= \frac{e \deg X}{(m-1)!} d^{m-1} + (\text{terms of degree less than } m-1).
\end{aligned}$$

By Definition 12.13 again, this means that $\deg(I(X) + (f)) = e \deg X = \deg X \cdot \deg f$. □

**Example 12.17.** Let $I = (f) \trianglelefteq K[x_0, \dots, x_n]$ be a principal ideal. Then Bézout's Theorem together with Example 12.14 (a) implies

$$\deg I = \deg((0) + (f)) = \deg \mathbb{P}^n \cdot \deg f = \deg f.$$

In particular, if $X \subset \mathbb{P}^n$ is a hypersurface, so that $I(X) = (f)$ for some homogeneous polynomial $f$ by Exercise 7.16 (a), then $\deg X = \deg f$. This justifies the name "degree" in Definition 12.13.

**Exercise 12.18.** Prove that every pure-dimensional projective variety of degree 1 is a linear space.

**Notation 12.19.** Let $X \subset \mathbb{P}^n$ be an irreducible projective variety. In certain cases there are commonly used names to describe the degree and / or the dimension of $X$ that we have probably used informally already several times:

(a) If $\deg X = 1$ then $X$ is called a **line** if $\dim X = 1$, a **plane** if $\dim X = 2$, and a **hyperplane** if $\dim X = n-1$.

(b) For any dimension, $X$ is called a **quadric** if $\deg X = 2$, a **cubic** if $\deg X = 3$, a **quartic** if $\deg X = 4$, and so on.

**Corollary 12.20** (Bézout's Theorem for curves)**.**

(a) *Let $X \subset \mathbb{P}^n$ be a curve, and let $f \in K[x_0,\dots,x_n]$ be a homogeneous polynomial that does not vanish identically on any irreducible component of $X$. Then*

$$|X \cap V(f)| \leq \deg X \cdot \deg f.$$

(b) *For any two curves $X$ and $Y$ in $\mathbb{P}^2$ without a common irreducible component we have*

$$|X \cap Y| \leq \deg X \cdot \deg Y.$$

*Proof.*

(a) As $I(X)+(f)$ is an ideal with zero locus $X \cap V(f)$, the statement follows from Bézout's Theorem together with Example 12.14 (c).

(b) Apply (a) to a polynomial $f$ generating $I(Y)$, and use Example 12.17. □

For the remaining part of this chapter we will focus on the case of curves as in Corollary 12.20. Our goal is to assign a natural multiplicity to each point in $X \cap V(f)$ (resp. $X \cap Y$) so that the inequality becomes an equality when all points are counted with their respective multiplicities. In order to achieve this we have to study the degree of a homogeneous ideal with zero-dimensional zero locus from a local point of view. It is convenient to do this in an affine chart of $\mathbb{P}^n$, and then finally in the local rings.

**Exercise 12.21.** Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal with $\dim V_p(I) = 0$. Assume that we have chosen coordinates so that all points in $V_p(I)$ have a non-vanishing $x_0$-coordinate. Prove that the degree of $I$ is then

$$\deg I = \chi_I = \dim_K K[x_1,\dots,x_n]/J,$$

where $J = \{f(1,x_1,\dots,x_n) : f \in I\} \trianglelefteq K[x_1,\dots,x_n]$.

The following lemma now expresses this dimension as a sum of local dimensions. In case you have attended the Commutative Algebra class already you will probably recognize this as precisely the Structure Theorem for Artinian rings, stating that an Artinian ring is always the product of its localizations [G5, Proposition 7.20].

**Lemma 12.22.** *Let $J \trianglelefteq K[x_1,\dots,x_n]$ be an ideal with finite affine zero locus $V_a(J) = \{a_1,\dots,a_r\}$. Then*

$$K[x_1,\dots,x_n]/J \cong \mathscr{O}_{\mathbb{A}^n,a_1}/J\,\mathscr{O}_{\mathbb{A}^n,a_1} \times \cdots \times \mathscr{O}_{\mathbb{A}^n,a_r}/J\,\mathscr{O}_{\mathbb{A}^n,a_r},$$

*where $J\,\mathscr{O}_{\mathbb{A}^n,a_i}$ denotes the ideal in $\mathscr{O}_{\mathbb{A}^n,a_i}$ generated by all elements $\frac{f}{1}$ for $f \in J$.*

*Proof.* Consider the primary decomposition of $J$ as in Remark 12.10. By part (b) of this remark it is of the form $J = J_1 \cap \cdots \cap J_r$ for some ideals $J_1,\dots,J_r$ with $V_a(J_i) = \{a_i\}$ for all $i$. Moreover, note that $J_i\,\mathscr{O}_{\mathbb{A}^n,a_j}$ is the unit ideal for $i \neq j$ since $a_j \notin V_a(J_i)$ implies that there is a polynomial in $J_i$ not vanishing at $a_j$, so that it is a unit in the local ring $\mathscr{O}_{\mathbb{A}^n,a_j}$. Hence it suffices to prove that

$$K[x_1,\dots,x_n]/J \cong \mathscr{O}_{\mathbb{A}^n,a_1}/J_1\,\mathscr{O}_{\mathbb{A}^n,a_1} \times \cdots \times \mathscr{O}_{\mathbb{A}^n,a_r}/J_r\,\mathscr{O}_{\mathbb{A}^n,a_r}.$$

We will do this by showing that the $K$-algebra homomorphism

$$\varphi : K[x_1,\dots,x_n]/J \to \mathscr{O}_{\mathbb{A}^n,a_1}/J_1\,\mathscr{O}_{\mathbb{A}^n,a_1} \times \cdots \times \mathscr{O}_{\mathbb{A}^n,a_r}/J_r\,\mathscr{O}_{\mathbb{A}^n,a_r}, \;\; \overline{f} \mapsto \left(\overline{f},\dots,\overline{f}\right)$$

is bijective.

- $\varphi$ is injective: Let $f$ be a polynomial with $\varphi(\overline{f}) = 0$. Then $f$ lies in $J_i \mathscr{O}_{\mathbb{A}^n,a_i}$ for all $i$, i. e. $\frac{f}{1} = \frac{g_i}{f_i}$ for some $g_i, f_i$ with $g_i \in J_i$ and $f_i \in K[x_0,\dots,x_n]$ such that $f(a_i) \neq 0$. This means that $h_i(f_i f - g_i) = 0$ for some $h_i$ with $h_i(a_i) \neq 0$, and hence $h_i f_i f \in J_i$. But $h_i f_i \notin I(a_i)$ means $h_i f_i \notin \sqrt{J_i}$, and thus $f \in J_i$ since $J_i$ is primary. As this holds for all $i$, we conclude that $f \in J$, i. e. $\overline{f} = 0$ in $K[x_1,\dots,x_n]/J$.
- $\varphi$ is surjective: By symmetry of the factors it suffices to prove that $(1,0,\dots,0) \in \operatorname{im}\varphi$. As $V(J_1 + J_i) = \{a_1\} \cap \{a_i\} = \emptyset$ for all $i > 1$ we see that $1 \in \sqrt{J_1 + J_i}$, and hence also $1 \in J_1 + J_i$. There are thus $a_i \in J_1$ and $b_i \in J_i$ with $a_i + b_i = 1$, so that $b_i \equiv 0 \mod J_i$ and $b_i \equiv 1 \mod J_1$. Hence the product $b_2 \cdot \dots \cdot b_r$ is an inverse image of $(1,0,\dots,0)$ under $\varphi$. □

It is now straightforward to translate Bézout's Theorem for curves into a local version.

**Definition 12.23** (Multiplicities)**.**

(a) Let $I \trianglelefteq K[x_0,\dots,x_n]$ be a homogeneous ideal with finite projective zero locus, and let $a \in \mathbb{P}^n$. Choose an affine patch of $\mathbb{P}^n$ containing $a$, and let $J$ be the corresponding affine ideal as in Exercise 12.21. Then

$$\operatorname{mult}_a(I) := \dim_K \mathscr{O}_{\mathbb{A}^n,a}/J\,\mathscr{O}_{\mathbb{A}^n,a}$$

is called the **multiplicity** of $I$ at $a$.

(b) Let $X \subset \mathbb{P}^n$ be a projective curve, and let $a \in X$ be a point. For a homogeneous polynomial $f \in K[x_0,\dots,x_n]$ that does not vanish identically on any irreducible component of $X$, the number

$$\operatorname{mult}_a(X,f) := \operatorname{mult}_a(I(X) + (f))$$

is called the **(vanishing) multiplicity** of $f$ at $a$. Note that $\operatorname{mult}_a(X,f)$ depends only on the class of $f$ modulo $I(X)$ and not on $f$ itself, so that we can also construct the multiplicity $\operatorname{mult}_a(X,f)$ for $f \in S(X)$. In this case, we will also often simplify its notation to $\operatorname{mult}_a(f)$.

If $n = 2$ and $Y \subset \mathbb{P}^2$ is another curve that does not share a common irreducible component with $X$, the **intersection multiplicity** of $X$ and $Y$ at $a$ is defined as

$$\operatorname{mult}_a(X,Y) := \operatorname{mult}_a(I(X) + I(Y)).$$

**Remark 12.24** (Positivity of multiplicities)**.** Continuing the notation of Definition 12.23 (a), note that $1 \notin J\,\mathscr{O}_{\mathbb{A}^n,a}$ if and only if $a \in V_p(I)$. It follows that $\operatorname{mult}_a(I) \geq 1$ if and only if $a \in V_p(I)$. Applying this to Definition 12.23 (b), we see that the vanishing multiplicity $\operatorname{mult}_a(X,f)$ is at least 1 if and only if $f(a) = 0$, and the intersection multiplicity $\operatorname{mult}_a(X,Y)$ is at least 1 if and only if $a \in X \cap Y$. In fact, we will show in Exercise 12.27 that there is also an easy geometric criterion for when $\operatorname{mult}_a(X,Y) = 1$.

**Remark 12.25** (Vanishing and intersection multiplicities in local rings)**.** It is often useful to express the multiplicities of Definition 12.23 (b) in terms of local rings as in Definition 12.23 (a). As above, we choose an affine patch $\{x \in \mathbb{P}^n : x_i \neq 0\} \cong \mathbb{A}^n$ of $\mathbb{P}^n$ containing $a$. By abuse of notation, if $f \in K[x_0,\dots,x_n]$ is a homogeneous polynomial, we will also denote by $f$ the (not necessarily homogeneous) polynomial obtained from it by setting $x_i$ equal to 1, and then also its quotient by 1 in the local ring $\mathscr{O}_{\mathbb{A}^n,a}$ (see Exercise 3.24). Then Definition 12.23 can be formulated as follows:

(a) Let $X \subset \mathbb{P}^n$ be a curve, and let $f \in K[x_0,\dots,x_n]$ be a homogeneous polynomial not vanishing identically on any irreducible component of $X$. Denote by $U = X \cap \mathbb{A}^n$ the affine part of $X$ in the chosen patch, and let $J = I(U)$ be its ideal. Then the vanishing multiplicity of $f$ at $a \in \mathbb{P}^n$ is equal to $\dim_K \mathscr{O}_{\mathbb{A}^n,a}/(J + (f))\,\mathscr{O}_{\mathbb{A}^n,a}$ by Definition 12.23. But $\mathscr{O}_{\mathbb{A}^n,a}/J\,\mathscr{O}_{\mathbb{A}^n,a} \cong \mathscr{O}_{X,a}$ by Exercise 3.23, and so we conclude that

$$\operatorname{mult}_a(X,f) = \dim_K \mathscr{O}_{X,a}/(f).$$

Note that we could use the same formula to define the vanishing multiplicity for any local function $f \in \mathscr{O}_{X,a}$ that does not vanish identically on any irreducible component of $X$ through $a$. In fact, for an irreducible variety we will even define such a multiplicity for rational functions in Construction 14.5, which then includes the case of local functions (see Exercise 9.8 (b) and Remark 14.7).

(b) For two curves $X, Y \subset \mathbb{P}^2$ without common irreducible component and ideals $I(X) = (f)$ and $I(X) = (g)$ their intersection multiplicity is
$$\text{mult}_a(X,Y) = \dim_K \mathscr{O}_{\mathbb{A}^2,a}/(f,g),$$
or alternatively with (a)
$$\text{mult}_a(X,Y) = \dim_K \mathscr{O}_{X,a}/(g) = \dim_K \mathscr{O}_{Y,a}/(f).$$

Of course, as we have defined the multiplicities above using affine charts, we could construct them equally well for affine instead of projective varieties. However, the projective case is needed for the local version of Bézout's Theorem, which we can now prove.

**Corollary 12.26** (Bézout's Theorem for curves, local version)**.**

(a) *Let $X \subset \mathbb{P}^n$ be a curve, and let $f \in K[x_0,\dots,x_n]$ be a homogeneous polynomial that does not vanish identically on any irreducible component of $X$. Then*
$$\sum_{a \in X \cap V(f)} \text{mult}_a(X,f) = \deg X \cdot \deg f.$$

(b) *For any two curves $X$ and $Y$ in $\mathbb{P}^2$ without a common irreducible component we have*
$$\sum_{a \in X \cap Y} \text{mult}_a(X,Y) = \deg X \cdot \deg Y.$$

*Proof.*

(a) By Exercise 12.21, Lemma 12.22, and the definition of multiplicities, all applied to the ideal $I(X)+(f)$, we have
$$\deg(I(X)+(f)) = \sum_{a \in X \cap V(f)} \text{mult}_a(X,f).$$
Hence the statement follows immediately from Proposition 12.16.

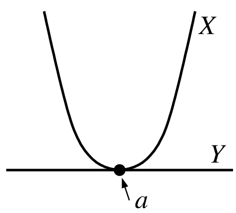
(b) This follows from (a) for $f$ a polynomial generating $I(Y)$, since $\deg Y = \deg f$ by Example 12.17. □

**Exercise 12.27** (Geometric interpretation of intersection multiplicities)**.** Let $X,Y \subset \mathbb{A}^2$ be two affine curves containing the origin. Moreover, let $I(X) = (f)$ and $I(Y) = (g)$ be their ideals. Show that the following statements are equivalent:

(a) $\dim_K \mathscr{O}_{\mathbb{A}^2,0}/(f,g) = 1$ (i. e. the intersection multiplicity of $X$ and $Y$ at the origin is 1).

(b) $X$ and $Y$ are smooth at 0 and have different tangent spaces there (i. e. "$X$ and $Y$ intersect transversely at the origin").

**Example 12.28.** Consider the two projective curves
$$X = V(x_0x_2 - x_1^2) \quad \text{and} \quad Y = V(x_2)$$
in $\mathbb{P}^2$, whose affine parts (which we have already considered in Remark 1.27) are shown in the picture on the right. Note that $\deg X = 2$ and $\deg Y = 1$ by Example 12.17, and that $a := (1:0:0)$ is the only point in the intersection $X \cap Y$. As $X$ and $Y$ have the same tangent space at $a$, we must have $\text{mult}_a(X,Y) \geq 2$ by Exercise 12.27.




In fact, it is easy to compute $\text{mult}_a(X,Y)$ explicitly: by definition we have
$$\text{mult}_a(X,Y) = \text{mult}_a(x_0x_2 - x_1^2, x_2) = \dim_K \mathscr{O}_{\mathbb{A}^2,0}/(x_2 - x_1^2, x_2) = \dim_K \mathscr{O}_{\mathbb{A}^2,0}/(x_1^2, x_2),$$
and since $a$ is the only intersection point of $X$ and $Y$ we can rewrite this by Lemma 12.22 as
$$\text{mult}_a(X,Y) = \dim_K K[x_1,x_2]/(x_1^2,x_2) = \dim_K K[x_1]/(x_1^2) = 2.$$

Note that this is in accordance with Bézout's Theorem as in Corollary 12.26 (b), since $\operatorname{mult}_a(X,Y) = 2 = \deg X \cdot \deg Y$.

# 13. Applications of Bézout's Theorem

Bézout's Theorem (as in Proposition 12.16 or Corollary 12.26) is clearly one of the most powerful results in algebraic geometry that we will discuss in this class. To illustrate this, let us now take some time to study several of its applications, which are in fact of quite different flavors.

Our first application is actually not much more than a simple remark. It concerns the question whether the ideal of a given variety of pure dimension $n-k$ in affine space $\mathbb{A}^n$ or projective space $\mathbb{P}^n$ can be generated by $k$ elements. We have seen in Exercise 7.16 (a) already that this is always the case if $k=1$. On the other hand, Example 0.11 and Exercise 2.32 show that for $k \geq 2$ one sometimes needs more than $k$ generators. Bézout's Theorem can often be used to detect when this happens, e. g. in the following setting.

**Proposition 13.1.** *Let $X \subset \mathbb{P}^3$ be a curve that is not contained in any proper linear subspace of $\mathbb{P}^3$. If $\deg X$ is a prime number, then $I(X)$ cannot be generated by two elements.*

*Proof.* Assume for a contradiction that we have $I(X) = (f,g)$ for two homogeneous polynomials $f,g \in K[x_0,x_1,x_2,x_3]$. Clearly, $g$ does not vanish identically on any irreducible component of $V(f)$, since otherwise the zero locus of $(f,g)$ would have codimension 1. By Proposition 12.16 and Example 12.17 we therefore have

$$\deg X = \deg((f)+(g)) = \deg(f) \cdot \deg g = \deg f \cdot \deg g.$$

As $\deg X$ is a prime number, this is only possible if $\deg f = 1$ or $\deg g = 1$, i. e. if one of these polynomials is linear. But then $X$ is contained in a proper linear subspace of $\mathbb{P}^3$, in contradiction to our assumption. □

**Example 13.2.** Let $X$ be the degree-3 Veronese embedding of $\mathbb{P}^1$ as in Construction 7.27, i. e.

$$X = \{(s^3 : s^2t : st^2 : t^3) : (s:t) \in \mathbb{P}^1\} \quad \subset \mathbb{P}^3.$$

By Exercise 12.15 (b) we know that $X$ is a cubic curve, i. e. $\deg X = 3$. Moreover, $X$ is not contained in any proper linear subspace of $\mathbb{P}^3$, since otherwise the monomials $s^3, s^2t, st^2, t^3$ would have to satisfy a $K$-linear relation. Hence Proposition 13.1 implies that the ideal $I(X)$ cannot be generated by two elements.

However, one can check directly that $I(X)$ can be generated by three elements. For example, we can write

$$I(X) = (x_0x_2 - x_1^2, x_1x_3 - x_2^2, x_0x_3 - x_1x_2).$$

In the spirit of Bézout's Theorem, we can also see geometrically why none of these three generators is superfluous: if we leave out e. g. the last generator and consider $I = (x_0x_2 - x_1^2, x_1x_3 - x_2^2)$ instead, we now have $\deg I = 2 \cdot 2 = 4$. Clearly, $V(I)$ still contains the cubic $X$, and hence by the additivity of degrees as in Example 12.14 (b) there must be another 1-dimensional component in $V(I)$ of degree 1. In fact, this component is easy to find: we have $V(I) = X \cup L$ for the line $L = V(x_1,x_2)$.

**Exercise 13.3.** Let $X \subset \mathbb{P}^n$ be an irreducible curve of degree $d$. Show that $X$ is contained in a linear subspace of $\mathbb{P}^n$ of dimension at most $d$.

21

As another application of Bézout's Theorem, we are now able to prove the result already announced in Example 7.6 (a) that any isomorphism of $\mathbb{P}^n$ is linear, i. e. a projective automorphism in the sense of this example. Note that the corresponding statement would be false in the affine case, as e. g. $f : \mathbb{A}^2 \to \mathbb{A}^2,\ (x_1,x_2) \mapsto (x_1, x_2 + x_1^2)$ is an isomorphism with inverse $f^{-1} : (x_1,x_2) \mapsto (x_1, x_2 - x_1^2)$.

**Proposition 13.4.** *Every isomorphism $f : \mathbb{P}^n \to \mathbb{P}^n$ is linear, i. e. of the form $f(x) = Ax$ for an invertible matrix $A \in \mathrm{GL}(n+1,K)$.*

*Proof.* Let $H \subset \mathbb{P}^n$ be a hyperplane (given as the zero locus of one homogeneous linear polynomial), and let $L \subset \mathbb{P}^n$ be a line not contained in $H$. Clearly, the intersection $L \cap H$ consists of one point with multiplicity 1 (i. e. $I(L)+I(H)$ has multiplicity 1 in the sense of Definition 12.23 (a)). As $f$ is an isomorphism, $f(L)$ and $f(H)$ must again be a curve resp. a hypersurface that intersect in one point with multiplicity 1. By the local version of Bézout's Theorem as in Corollary 12.26 (a), this means that $\deg f(L) \cdot \deg f(H) = 1$. This is only possible if $\deg f(H) = 1$. In other words, $f$ must map hyperplanes to hyperplanes.

By composing $f$ with a suitable projective automorphism (i. e. a linear map as in Example 7.6 (a)) we can therefore assume that $f$ maps the affine part $\mathbb{A}^n = \mathbb{P}^n \backslash V(x_0)$ isomorphically to itself. Passing to this affine part with coordinates $x_1, \dots, x_n$, the above argument shows that $f^{-1}(V(x_i)) = V(f^* x_i)$ is an affine linear space for all $i$, so that $f^* x_i$ must be a power of a linear polynomial. But $f^* : K[x_1, \dots, x_n] \to K[x_1, \dots, x_n]$ is an isomorphism by Corollary 4.8 and thus maps irreducible polynomials to irreducible polynomials again. Hence $f^* x_i$ is itself linear for all $i$, which means that $f$ is an affine linear map on $\mathbb{A}^n$, i. e. a linear map on $\mathbb{P}^n$. □

For the rest of this chapter we will now restrict to plane curves. One consequence of Bézout's Theorem in this case is that it gives an upper bound on the number of singular points that such a curve can have, in terms of its degree.

**Proposition 13.5.** *Let $X \subset \mathbb{P}^2$ be an irreducible curve of degree $d$. Then $X$ has at most $\binom{d-1}{2}$ singular points.*

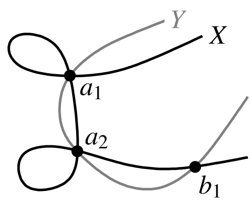
**Example 13.6.**

(a) A plane curve of degree 1 is a line, which is isomorphic to $\mathbb{P}^1$. A curve of degree 2, i. e. a conic, is always isomorphic to $\mathbb{P}^1$ as well, as we have seen in Example 7.6 (d). So in both these cases the curve does not have any singular points, in accordance with the statement of Proposition 13.5.

(b) By Proposition 13.5, a cubic curve in $\mathbb{P}^2$ can have at most one singular point. In fact, we have already seen both a cubic with no singular points (e. g. the Fermat cubic in Example 10.20) and a cubic with one singular point (e. g. $V_p(x_0x_2^2 - x_0x_1^2 - x_1^3)$ or $V_p(x_0x_2^2 - x_1^3)$, whose affine parts we have studied in Example 10.3).

(c) Without the assumption of $X$ being irreducible the statement of Proposition 13.5 is false: a union of two intersecting lines in $\mathbb{P}^2$ has degree 2, but a singular point (namely the intersection point). However, we will see in Exercise 13.7 that there is a similar statement with a slightly higher bound also for not necessarily irreducible curves.

*Proof of Proposition 13.5.* By Example 13.6 (a) it suffices to prove the theorem for $d \geq 3$. Assume for a contradiction that there are distinct singular points $a_1, \dots, a_{\binom{d-1}{2}+1}$ of $X$. Moreover, pick $d-3$ arbitrary further distinct points $b_1, \dots, b_{d-3}$ on $X$, so that the total number of points is




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

We claim that there is a curve $Y$ of degree at most $d-2$ that passes through all these points. The argument is essentially the same as in Exercise 7.31 (c): the space $K[x_0, x_1, x_2]_{d-2}$ of all homogeneous polynomials of degree $d-2$ in three variables is a vector space of dimension $\binom{d}{2}$, with the coefficients of the polynomials as coordinates. Moreover, the condition that such a polynomial vanishes at a given point is clearly a homogeneous linear equation in these coordinates. As $\binom{d}{2} - 1$ homogeneous linear equations in a vector space of dimension $\binom{d}{2}$ must have a non-trivial common solution, we conclude that there is a non-zero polynomial $f \in K[x_0, x_1, x_2]_{d-2}$ vanishing at all points $a_i$ and $b_j$. The corresponding curve $Y = V_p(f)$ then has degree at most $d-2$ (strictly less if $f$ contains repeated factors) and passes through all these points.

Note that $X$ and $Y$ cannot have a common irreducible component, since $X$ is irreducible and of bigger degree than $Y$. Hence Corollary 12.26 (b) shows that the curves $X$ and $Y$ can intersect in at most $\deg X \cdot \deg Y = d(d-2)$ points, counted with multiplicities. But the intersection multiplicity at all $a_i$ is at least 2 by Exercise 12.27 since $X$ 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 13.7.** Show that a (not necessarily irreducible) curve of degree $d$ in $\mathbb{P}^2$ has at most $\binom{d}{2}$ singular points. Can you find an example for each $d$ in which this maximal number of singular points is actually reached?

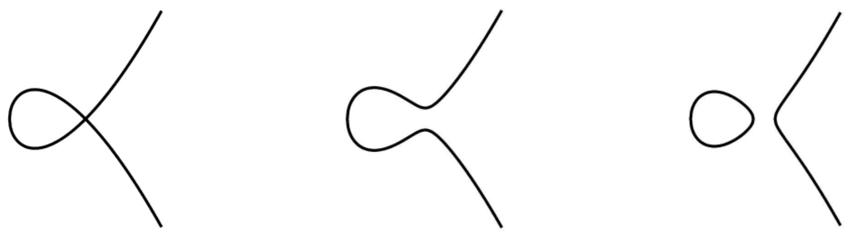
Let us now study smooth plane curves in more detail. An interesting topic that we have neglected entirely so far is the *classical* topology of such curves when we consider them over the real or complex numbers, e. g. their number of connected components in the standard topology. We will now see that 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 this class. 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 13.8** (Real curves)**.** Note that we have developed most of our theory only for algebraically closed ground fields, so that our constructions and results are not directly applicable to real curves. However, as we will not go very deep into the theory of real algebraic geometry it suffices to note that the definition of projective curves and their ideals works over $\mathbb{R}$ in the same way as over $\mathbb{C}$. To apply other concepts and theorems to a real curve $X$ with ideal $I(X) = (f)$ for a real homogeneous polynomial $f$, we will simply pass to the corresponding complex curve $X_\mathbb{C} = V_p(f) \subset \mathbb{P}^2_\mathbb{C}$ first. For example, we will say that $X$ is smooth or irreducible if $X_\mathbb{C}$ is.

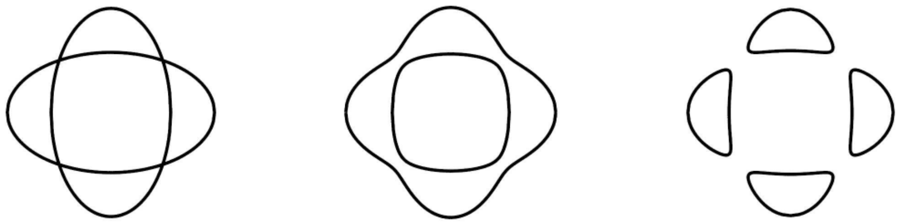
**Remark 13.9** (Loops of real projective curves)**.** Let $X \subset \mathbb{P}^2_\mathbb{R}$ be a smooth projective curve over $\mathbb{R}$. In the classical topology, $X$ is then a compact 1-dimensional manifold (see Remark 10.14). This means that $X$ is a disjoint union of finitely many connected components, each of which is homeomorphic to a circle. We will refer to these components as *loops* of $X$.

Note that $X$ can consist of several loops in the classical topology even if $f$ is irreducible (so that $X$ is irreducible in the Zariski topology). A convenient way to construct such curves is by deformations of singular curves. For example, consider the singular cubic curve $X$ in $\mathbb{P}^2_\mathbb{R}$ whose affine part in $\mathbb{A}^2_\mathbb{R}$ is the zero locus of $f_3 := x_2^2 - x_1^2 - x_1^3$ as in Example 10.3. It has a double point at the origin, as shown in the picture below on the left. In $\mathbb{P}^2_\mathbb{R}$, the curve contains one additional point at infinity that connects the two unbounded branches, so that $X$ is homeomorphic to two circles glued together at a point.



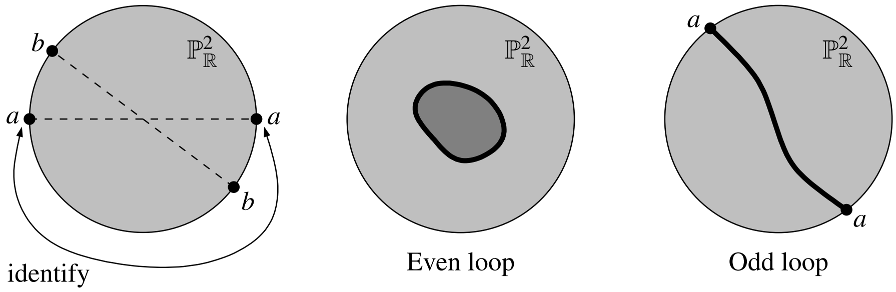
$f_3(x_1,x_2) = 0$ $\qquad$ $f_3(x_1,x_2) - \varepsilon = 0$ $\qquad$ $f_3(x_1,x_2) + \varepsilon = 0$

Let us now perturb $f_3$ and consider the curves $f_3(x_1,x_2) \pm \varepsilon = 0$ for a small number $\varepsilon \in \mathbb{R}_{>0}$ instead. This deforms $X$ into a smooth cubic with one or two loops depending on the sign of the perturbation, as shown in the picture above. The same technique applied to a singular quartic curve, e. g. the union of two ellipses given by $f_4 = (x_1^2 + 2x_2^2 - 1)(x_2^2 + 2x_1^2 - 1)$, yields two or four loops as in the following picture.



$f_4(x_1,x_2) = 0$ $\qquad$ $f_4(x_1,x_2) - \varepsilon = 0$ $\qquad$ $f_4(x_1,x_2) + \varepsilon = 0$

**Remark 13.10** (Even and odd loops)**.** Although all loops of smooth curves in $\mathbb{P}^2_{\mathbb{R}}$ 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 6.3 that $\mathbb{P}^2_{\mathbb{R}}$ is obtained from $\mathbb{A}^2_{\mathbb{R}}$ (which we will draw topologically as an open disc below) by adding a point at infinity for each direction in $\mathbb{A}^2_{\mathbb{R}}$. This has the effect of adding a boundary to the disc, with the boundary points corresponding to the points at infinity. But note that opposite points of the boundary of the disc belong to the same direction in $\mathbb{A}^2_{\mathbb{R}}$ and hence are the same point in $\mathbb{P}^2_{\mathbb{R}}$. In other words, $\mathbb{P}^2_{\mathbb{R}}$ is topologically equivalent to a closed disc with opposite boundary points identified, as in the picture below on the left.




identify $\qquad$ Even loop $\qquad$ Odd loop

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.

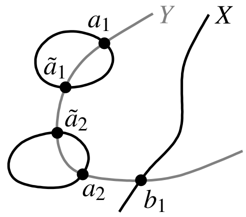
In principle, a real curve can have even as well as odd loops. There is one restriction however: as the complement of an odd loop is simply a disc, all other loops in this complement will have an interior and exterior, so that they are even. In other words, a smooth curve in $\mathbb{P}^2_{\mathbb{R}}$ can have at most one odd loop.

We are now ready to find a bound on the number of loops in a smooth curve in $\mathbb{P}^2_{\mathbb{R}}$ of a given degree. Interestingly, the idea in its proof is almost identical to that in Proposition 13.5, although the resulting statement is quite different.

**Proposition 13.11** (**Harnack's Theorem**)**.** *A smooth real projective curve of degree $d$ in $\mathbb{P}^2_{\mathbb{R}}$ has at most $\binom{d-1}{2} + 1$ loops.*

**Example 13.12.** A line ($d = 1$) has always exactly one loop. A conic ($d = 2$) is a hyperbola, parabola, or ellipse, 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 13.9. In fact, one can show that the bound given in Harnack's theorem is sharp, i. e. that for every $d$ one can find smooth real curves of degree $d$ with exactly $\binom{d-1}{2}+1$ loops.

*Proof sketch of Proposition 13.11.* By Example 13.12 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 13.10 that at least $\binom{d-1}{2}+1$ of these loops must be even. Hence we can pick points $a_1, \dots, a_{\binom{d-1}{2}+1}$ on distinct even loops of $X$, and $d-3$ more points $b_1, \dots, b_{d-3}$ on another loop (which might be even or odd). So as in the proof of Proposition 13.5, we have a total of $\binom{d}{2}-1$ points.




Again as in the proof of Proposition 13.5, it now follows that there is a real curve $Y$ of degree at most $d-2$ passing through all these points. Note that the corresponding complex curves $X_{\mathbb{C}}$ and $Y_{\mathbb{C}}$ as in Remark 13.8 cannot have a common irreducible component since $X_{\mathbb{C}}$ is assumed to be smooth, hence irreducible by Exercise 10.22 (a), and has bigger degree than $Y_{\mathbb{C}}$. So Bézout's Theorem as in Corollary 12.26 (b) implies that $X_{\mathbb{C}}$ and $Y_{\mathbb{C}}$ intersect in at most $d(d-2)$ points, counted with multiplicities. But recall that the even loops of $X$ containing the points $a_i$ divide the real projective plane into two regions, hence if $Y$ enters the interior of such a loop it has to exit it again at another point $\tilde{a}_i$ of the same loop as in the picture above (it may also happen that $Y$ is tangent to $X$ at $a_i$, in which case their intersection multiplicity is at least 2 there by Exercise 12.27). 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 in the classical topology. In fact, we have seen such an example already in Example 0.1 of the introduction.

**Remark 13.13** (Classical topology of complex curves)**.** Let $X \subset \mathbb{P}^2_{\mathbb{C}}$ be a smooth curve. Then $X$ is a compact 2-dimensional manifold in the classical topology (see Remark 10.14). Moreover, one can show:

(a) $X$ is always an *oriented manifold* in the classical topology, i. e. a "two-sided surface", as opposed to e. g. a Möbius strip. To see this, note that all tangent spaces $T_aX$ of $X$ for $a \in X$ are isomorphic to $\mathbb{C}$, and hence admit a well-defined multiplication with the imaginary unit $i$. Geometrically, this means that all tangent planes have a well-defined notion of a *clockwise* rotation by 90 degrees, varying continuously with $a$ — which defines an orientation of $X$. In fact, this statement holds for all smooth complex curves, not just for curves in $\mathbb{P}^2_{\mathbb{C}}$.

(b) In contrast to the real case that we have just studied, $X$ 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 $X$ had (at least) two connected components $X_1$ and $X_2$ in the classical topology, each of these components 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.

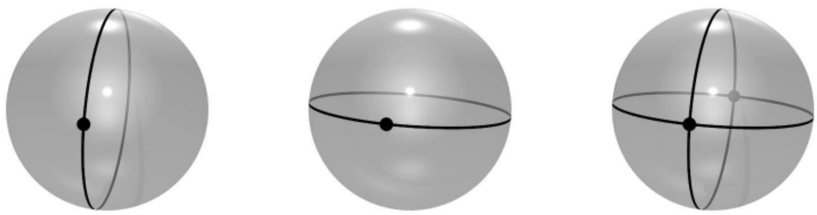
Of course, the methods needed to prove Bézout's Theorem in the topological setting are entirely different from ours in Chapter 12. If you know some algebraic topology, the statement here is that the 2-dimensional homology group $H_2(\mathbb{P}^2_{\mathbb{C}}, \mathbb{Z})$ is isomorphic to $\mathbb{Z}$. 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 **genus** of $X$. Hence every curve in $\mathbb{P}^2_{\mathbb{C}}$ can be assigned a genus that describes its topological type in the classical topology. The picture on the right shows a complex
22 curve of genus 2.

Our goal for the rest of this chapter will be to compute the genus of a smooth curve in $\mathbb{P}^2_{\mathbb{C}}$ in terms of its degree, as already announced in Example 0.3. To do this, we will need the following technique from topology.

**Construction 13.14** (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.



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:

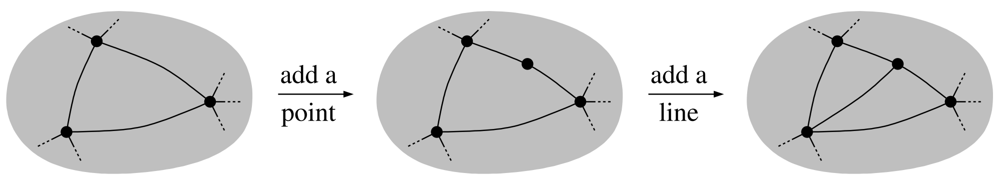
**Lemma and Definition 13.15** (Euler characteristic)**.** *Let $X$ be a compact 2-dimensional manifold. Consider a cell decomposition of $X$, consisting of $\sigma_0$ points, $\sigma_1$ lines, and $\sigma_2$ discs. Then the number*

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

*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 $\sigma_0$ and $\sigma_1$ by 1 as in the picture below, hence the alternating sum $\sigma_0 - \sigma_1 + \sigma_2$ does not change.
- Adding another line in a disc: in this case we raise $\sigma_1$ and $\sigma_2$ by 1, so again $\sigma_0 - \sigma_1 + \sigma_2$ 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 13.14 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 13.16** (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 $g = 1 - \frac{\chi}{2}$.

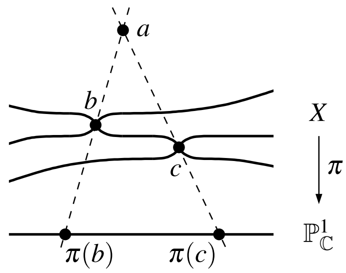
We are now ready to compute the genus of a smooth curve in $\mathbb{P}^2_{\mathbb{C}}$.

**Proposition 13.17** (**Degree-genus formula**)**.** *A smooth curve of degree $d$ in $\mathbb{P}^2_{\mathbb{C}}$ has genus $\binom{d-1}{2}$.*

*Proof sketch.* Let $I(X) = (f)$ for a homogeneous polynomial $f$ of degree $d$. By a linear change of coordinates we can assume that $a := (1:0:0) \notin X$. Then the projection

$$\pi : X \to \mathbb{P}^1_{\mathbb{C}}, \ \ (x_0 : x_1 : x_2) \mapsto (x_1 : x_2)$$

from $a$ as in the picture on the right is a well-defined morphism on $X$. Let us study its inverse images of a fixed point $(x_1 : x_2) \in \mathbb{P}^1_{\mathbb{C}}$. Of course, they are given by the values of $x_0$ such that $f(x_0, x_1, x_2) = 0$, so that there are exactly $d$ such points — unless the polynomial $f(\cdot, x_1, x_2)$ has a multiple zero in $x_0$ at a point in the inverse image, which happens if and only if $f$ and $\frac{\partial f}{\partial x_0}$ are simultaneously zero.




If we choose our original linear change of coordinates general enough, exactly two of the zeroes of $f(\cdot, x_1, x_2)$ will coincide at these points in the common zero locus of $f$ and $\frac{\partial f}{\partial x_0}$, so that $\pi^{-1}(x_1 : x_2)$ then consists of $d-1$ instead of $d$ points. These points, as e. g. $b$ and $c$ in the picture above, are usually called the *ramification points* of $\pi$. Note that the picture might be a bit misleading since it suggests that $X$ is singular at $b$ and $c$, which is not the case. The correct topological picture of the map is impossible to draw however since it would require a real 4-dimensional space, namely an affine chart of $\mathbb{P}^2_{\mathbb{C}}$.

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 13.16 since $\mathbb{P}^1_{\mathbb{C}}$ is topologically a sphere, i. e. of genus 0. Now lift this cell decomposition to a decomposition of $X$ 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 ramification points, which have one inverse image less. As the number of ramification points is $|V_p(f, \frac{\partial f}{\partial x_0})| = \deg f \cdot \deg \frac{\partial f}{\partial x_0} = d(d-1)$ by Bézout's Theorem, the resulting decomposition of $X$ has $d\sigma_0 - d(d-1)$ points, $d\sigma_1$ lines, and $d\sigma_2$ discs. Hence by Lemma 13.15 the Euler characteristic of $X$ 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 13.16 that its genus is

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

**Example 13.18.**

(a) A smooth curve of degree 1 or 2 in $\mathbb{P}^2_{\mathbb{C}}$ is isomorphic to $\mathbb{P}^1_{\mathbb{C}}$ (see Example 7.6 (d)). It is therefore topologically a sphere, i. e. of genus 0, in accordance with Proposition 13.17.

(b) By Proposition 13.17, a smooth curve of degree 3 in $\mathbb{P}^2_{\mathbb{C}}$ has genus 1, i. e. it is topologically a torus. We will study such plane cubic curves in detail in Chapter 15.

**Remark 13.19.** Note that every isomorphism of complex varieties is also a homeomorphism in the classical topology. In particular, two smooth connected projective curves over $\mathbb{C}$ of different genus cannot be isomorphic. Combining this with Proposition 13.17, we see that two smooth curves in $\mathbb{P}^2_{\mathbb{C}}$ of different degree are never isomorphic, unless these degrees are 1 and 2.

**Exercise 13.20** (Arithmetic genus)**.** For a projective variety $X$ the number $(-1)^{\dim X} \cdot (\chi_X(0) - 1)$ is called its *arithmetic genus*, where $\chi_X$ denotes as usual the Hilbert polynomial of $X$. Show that the arithmetic genus of a smooth curve in $\mathbb{P}^2_{\mathbb{C}}$ agrees with the (geometric) genus introduced above.

In fact, one can show that this is true for any smooth projective curve over $\mathbb{C}$. However, the proof of this statement is hard and goes well beyond the scope of these notes. Note that, as the definition of the arithmetic genus is completely algebraic, one can use it to extend the notion of genus to projective curves over arbitrary ground fields.

**Exercise 13.21.** Show that

$$\{((x_0 : x_1), (y_0 : y_1)) : (x_0^2 + x_1^2)(y_0^2 + y_1^2) = x_0 x_1 y_0 y_1\} \quad \subset \mathbb{P}^1_{\mathbb{C}} \times \mathbb{P}^1_{\mathbb{C}}$$

is a smooth curve of genus 1.

# 14. Divisors on Curves

In its version for curves, Bézout's Theorem determines the number of zeroes of a homogeneous polynomial on a projective curve (see Corollaries 12.20 and 12.26). For example, if $X \subset \mathbb{P}^2$ is a cubic curve then the zero locus of a homogeneous linear polynomial $f$ on $X$ consists of three points, counted with multiplicities. But of course not every collection of three points on $X$ can arise in this way, as three points will in general not lie on a line, and thus cannot be in the zero locus of $f$. So by reducing the question of the zeroes of polynomials to just their number we are losing information about their possible positions. To avoid this, we will now present a theory that is able to keep track of the actual configurations of points on curves.

It turns out that these configurations, called *divisors* below, are parametrized by a group that is naturally associated to the curve $X$. This will allow us to study and classify curves with methods from group theory, very much in the same way as in topology the fundamental group or the homology groups can be used to study and distinguish topological spaces. For example, using divisors we will be able to prove in Proposition 14.19 and Remark 14.20 that a smooth plane cubic curve as above is never isomorphic to $\mathbb{P}^1$. Note that for the ground field $\mathbb{C}$ we have already seen this topologically in Remark 13.19 since a smooth plane cubic is a torus whereas $\mathbb{P}^1_{\mathbb{C}}$ is a sphere — but of course this was using techniques from topology that would certainly require some work to make them rigorous. In contrast, our new proof here will be entirely self-contained and algebraic, so in particular applicable to any ground field.

The concept of divisors can be defined for arbitrary curves. For example, in the smooth affine case this leads to the notion of Dedekind domains studied in commutative algebra [G5, Chapter 13], and for singular curves one needs two different concepts of divisors, called *Weil divisors* and *Cartier divisors*. However, in our applications we will only need irreducible smooth projective curves. So for simplicity of notation we will restrict ourselves to this case from the very beginning, even if many of our constructions and results do not need all these assumptions.

Let us start by giving the definition of divisors. It should be noted that the name "divisor" in this context has historical reasons; it is completely unrelated to the notion of divisors in an integral domain.

**Definition 14.1** (Divisors)**.** Let $X$ be an irreducible smooth projective curve.

(a) A **divisor** on $X$ is a formal finite linear combination $k_1a_1 + \cdots + k_na_n$ of distinct points $a_1,\dots,a_n \in X$ with integer coefficients $k_1,\dots,k_n \in \mathbb{Z}$ for some $n \in \mathbb{N}$. Obviously, the divisors on $X$ form an Abelian group under pointwise addition of the coefficients. We will denote it by $\operatorname{Div} X$.

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

(b) A divisor $D = k_1a_1 + \cdots + k_na_n$ as above is called **effective**, written $D \geq 0$, if $k_i \geq 0$ for all $i = 1,\dots,n$. 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 $D_1$.

(c) The **degree** of a divisor $D = k_1a_1 + \cdots + k_na_n$ as above is the number $\deg D := k_1 + \cdots + k_n \in \mathbb{Z}$. Obviously, the degree is a group homomorphism $\deg : \operatorname{Div} X \to \mathbb{Z}$. Its kernel is denoted by

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

**Construction 14.2** (Divisors from polynomials and intersections)**.** Again, let $X \subset \mathbb{P}^n$ be an irreducible smooth curve. Our construction of multiplicities in Definition 12.23 (b) allows us to define divisors on $X$ as follows.

(a) For a non-zero homogeneous polynomial $f \in S(X)$ the *divisor of $f$* is defined to be
$$\operatorname{div} f := \sum_{a \in V_X(f)} \operatorname{mult}_a(f) \cdot a \quad \in \operatorname{Div} X,$$
where $V_X(f)$ denotes the zero locus of $f$ on $X$ as in Construction 6.18. In other words, the divisor $\operatorname{div} f$ contains the data of the zeroes of $f$ together with their multiplicities. By Bézout's Theorem as in Corollary 12.26 (a), its degree is $\deg(\operatorname{div} f) = \deg X \cdot \deg f$.

(b) If $n = 2$ and $Y \subset \mathbb{P}^2$ is another curve not containing $X$, the *intersection divisor* of $X$ and $Y$ is
$$X \cdot Y := \sum_{a \in X \cap Y} \operatorname{mult}_a(X,Y) \cdot a \quad \in \operatorname{Div} X.$$
By definition, this divisor is just the same as $\operatorname{div} f$ for a generator $f$ of $I(Y)$. Note that it is symmetric in $X$ and $Y$, so in particular the result can be considered as an element of $\operatorname{Div} Y$ as well if $Y$ is also smooth and irreducible. By Bézout's Theorem as in Corollary 12.26 (b), we have $\deg(X \cdot Y) = \deg X \cdot \deg Y$.

**Example 14.3.** Consider again the two projective curves $X = V(x_0x_2 - x_1^2)$ and $Y = V(x_2)$ in $\mathbb{P}^2$ of Example 12.28. We have seen in this example that $X$ and $Y$ intersect in a single point $a = (1\!:\!0\!:\!0)$ with multiplicity 2. Hence $X \cdot Y = 2a$ in $\operatorname{Div} X$ in the notation of Construction 14.2. Equivalently, we can write $\operatorname{div} x_2 = 2a$ on $X$, and $\operatorname{div}(x_0x_2 - x_1^2) = 2a$ on $Y$.

Note that so far all our multiplicities have been non-negative, and hence all the divisors in Construction 14.2 are effective. Let us now extend this construction to multiplicities and divisors of rational functions, which will lead to negative multiplicities at their poles, and thus to non-effective divisors. To do this, we need the following lemma first.

**Lemma 14.4.** *Let $X$ be an irreducible smooth projective curve, and let $f, g \in S(X)$ be two non-zero polynomials. Then*
$$\operatorname{mult}_a(fg) = \operatorname{mult}_a(f) + \operatorname{mult}_a(g)$$
*for all $a \in X$. In particular, we have $\operatorname{div}(fg) = \operatorname{div} f + \operatorname{div} g$ in $\operatorname{Div} X$.*

*Proof.* By Remark 12.25 (a) we have to show that
$$\dim_K \mathscr{O}_{X,a}/(fg) = \dim_K \mathscr{O}_{X,a}/(f) + \dim_K \mathscr{O}_{X,a}/(g).$$
for all $a \in X$. But this follows immediately by Lemma 12.5 from the exact sequence
$$0 \longrightarrow \mathscr{O}_{X,a}/(f) \xrightarrow{\cdot g} \mathscr{O}_{X,a}/(fg) \longrightarrow \mathscr{O}_{X,a}/(g) \longrightarrow 0$$
(for the injectivity of the first map note that $\mathscr{O}_{X,a}$ is an integral domain since $X$ is irreducible). Taking these results for all $a \in X$ together, we conclude that $\operatorname{div}(fg) = \operatorname{div} f + \operatorname{div} g$. □

**Construction 14.5** (Multiplicities and divisors of rational functions)**.** Let $X$ be an irreducible smooth projective curve, and let $\varphi \in K(X)^*$ be a non-zero rational function (see Construction 9.6). By Definition 7.1 we can write $\varphi = \frac{g}{f}$ for two homogeneous polynomials $f$ and $g$ of the same degree.

(a) We define the *multiplicity* of $\varphi$ at a point $a \in X$ to be
$$\operatorname{mult}_a(\varphi) := \operatorname{mult}_a(g) - \operatorname{mult}_a(f) \quad \in \mathbb{Z}.$$
Note that this is well-defined: if $\frac{g'}{f'} = \frac{g}{f}$ for two other homogeneous polynomials $f'$ and $g'$ of the same degree then $g'f - f'g = 0$ on a non-empty open subset, hence on all of $X$ since $X$ is irreducible, and consequently $g'f = f'g \in S(X)$. Lemma 14.4 thus implies that $\operatorname{mult}_a(g') + \operatorname{mult}_a(f) = \operatorname{mult}_a(f') + \operatorname{mult}_a(g)$, i. e. that $\operatorname{mult}_a(g') - \operatorname{mult}_a(f') = \operatorname{mult}_a(g) - \operatorname{mult}_a(f)$.

Geometrically, we can think of this multiplicity as the order of the zero (if $\operatorname{mult}_a(\varphi) > 0$) resp. pole (if $\operatorname{mult}_a(\varphi) < 0$) of $\varphi$ at $a$.

(b) Analogously to Construction 14.2, we define the *divisor* of $\varphi$ to be

$$\operatorname{div}\varphi := \sum_{a \in V_X(f) \cup V_X(g)} \operatorname{mult}_a(\varphi) \cdot a = \operatorname{div} g - \operatorname{div} f.$$

23

**Example 14.6.** The rational function $\varphi = \frac{x_0 x_1}{(x_0 - x_1)^2}$ on $\mathbb{P}^1$ has divisor

$$\operatorname{div}\varphi = (1:0) + (0:1) - 2\,(1:1).$$

**Remark 14.7** (Multiplicities of local functions)**.** By Exercise 9.8 (b) every local function $\varphi \in \mathscr{O}_{X,a}$ at a point $a$ of an irreducible smooth projective curve $X$ can be considered as a rational function on $X$. Hence Construction 14.5 also defines a multiplicity $\operatorname{mult}_a(\varphi)$ for any non-zero $\varphi \in \mathscr{O}_{X,a}$.

Moreover, note that $\varphi$ then has a representation of the form $\varphi = \frac{g}{f}$ with $f, g \in S(X)$ and $f(a) \neq 0$. By Remark 12.24, this means that $\operatorname{mult}_a(f) = 0$, and thus $\operatorname{mult}_a(\varphi) = \operatorname{mult}_a(g) \in \mathbb{N}$. By the same remark, we have $\operatorname{mult}_a(\varphi) = 0$ if and only if $g(a) \neq 0$ as well, i. e. if and only if $\varphi(a) \neq 0$.

**Remark 14.8.** As above, let $X$ be an irreducible smooth projective curve.

(a) Lemma 14.4 implies that $\operatorname{mult}_a(\varphi_1\varphi_2) = \operatorname{mult}_a \varphi_1 + \operatorname{mult}_a \varphi_2$ for any $a \in X$ and any two non-zero rational functions $\varphi_1, \varphi_2 \in K(X)^*$. We therefore also have $\operatorname{div}(\varphi_1\varphi_2) = \operatorname{div}\varphi_1 + \operatorname{div}\varphi_2$, i. e. the map $\operatorname{div} : K(X)^* \to \operatorname{Div} X$ is a homomorphism of groups.

(b) As any non-zero rational function on $X$ has the form $\varphi = \frac{g}{f}$ for two homogeneous polynomials of the same degree, we see by Construction 14.2 (a) that its divisor always has degree 0:

$$\deg \operatorname{div}\varphi = \deg(\operatorname{div} g - \operatorname{div} f) = \deg \operatorname{div} g - \deg \operatorname{div} f = \deg X \cdot \deg g - \deg X \cdot \deg f = 0.$$

Hence the homomorphism of (a) can also be viewed as a morphism $\operatorname{div} : K(X)^* \to \operatorname{Div}^0 X$.

This observation leads to the idea that we should give special attention to the divisors of rational functions, i. e. to the image subgroup of the above divisor homomorphism.

**Definition 14.9** (Principal divisors and the Picard group)**.** Let $X$ be an irreducible smooth projective curve.

(a) A divisor on $X$ is called **principal** if it is the divisor of a (non-zero) rational function. We denote the set of all principal divisors by $\operatorname{Prin} X$. Note that $\operatorname{Prin} X$ is just the image of the divisor homomorphism $\operatorname{div} : K(X)^* \to \operatorname{Div}^0 X$ of Remark 14.8 (b), and hence a subgroup of both $\operatorname{Div}^0 X$ and $\operatorname{Div} X$.

(b) The quotient

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

is called the **Picard group** or **group of divisor classes** on $X$. Restricting to degree zero, we also define $\operatorname{Pic}^0 X := \operatorname{Div}^0 X / \operatorname{Prin} X$. By abuse of notation, a divisor and its class in $\operatorname{Pic} X$ will usually be denoted by the same symbol.

**Remark 14.10.** The groups $\operatorname{Pic} X$ and $\operatorname{Pic}^0 X$ carry essentially the same information on $X$, since we always have

$$\operatorname{Pic} X / \operatorname{Pic}^0 X \cong \operatorname{Div} X / \operatorname{Div}^0 X \cong \mathbb{Z}.$$

It just depends on the specific application in mind whether it is more convenient to work with $\operatorname{Pic} X$ or with $\operatorname{Pic}^0 X$.

By construction, the group $\operatorname{Div} X$ of divisors on an irreducible smooth projective curve $X$ is a free Abelian group with an infinite number of generators, and hence not very interesting from a group-theoretic point of view. In contrast, the Picard group is rather "small" and has quite a rich structure that we want to study now in some examples.

**Example 14.11** (The Picard group of $\mathbb{P}^1$ is trivial)**.** On $\mathbb{P}^1$, every degree-0 divisor is principal: if $D = k_1\,(a_{1,0}:a_{1,1}) + \cdots + k_n\,(a_{n,0}:a_{n,1})$ for some points $(a_{i,0}:a_{i,1}) \in \mathbb{P}^1$ and integers $k_i$ for $i = 1,\dots,n$ with $k_1 + \cdots + k_n = 0$, the rational function given by

$$\varphi(x_0:x_1) = \prod_{i=1}^{n} (a_{i,1}x_0 - a_{i,0}x_1)^{k_i}$$

has divisor $\operatorname{div}\varphi = D$. Hence the divisor map $\operatorname{div}: K(\mathbb{P}^1)^* \to \operatorname{Div}^0\mathbb{P}^1$ is surjective, so that we have $\operatorname{Prin}\mathbb{P}^1 = \operatorname{Div}^0\mathbb{P}^1$, and consequently

$$\operatorname{Pic}^0\mathbb{P}^1 = \{0\} \quad \text{and} \quad \operatorname{Pic}\mathbb{P}^1 = \operatorname{Div}\mathbb{P}^1/\operatorname{Div}^0\mathbb{P}^1 \cong \mathbb{Z},$$

with the isomorphism $\deg: \operatorname{Pic}\mathbb{P}^1 \to \mathbb{Z}$.

Let us now move on to more complicated curves. We know already from Example 7.6 (d) that a smooth conic $X \subset \mathbb{P}^2$ (which is irreducible by Exercise 10.22 (a)) is isomorphic to $\mathbb{P}^1$. As the Picard group is clearly invariant under isomorphisms, this means that $\operatorname{Pic}^0 X$ will then be the trivial group again. So the next case to consider is a smooth cubic curve $X \subset \mathbb{P}^2$. Our main goal in this chapter is to prove that $\operatorname{Pic}^0 X$ is not trivial in this case, so that $X$ cannot be isomorphic to $\mathbb{P}^1$. In fact, in the next chapter in Proposition 15.2 we will even be able to compute $\operatorname{Pic}^0 X$ for a plane cubic explicitly.

However, even for the special case of plane cubics the computation of $\operatorname{Pic}^0 X$ is not easy, and so we will need some preliminaries first. The following lemma will be well-known to you if you have attended the Commutative Algebra class already, since it essentially states in algebraic terms that the local rings of $X$ are discrete valuation rings [G5, Lemma 12.1]. It is the first time in this chapter that the smoothness assumption on $X$ is essential.

**Lemma 14.12** (Local coordinates on a smooth plane curve)**.** *Let $X \subset \mathbb{P}^2$ be a smooth curve, and let $I_a \trianglelefteq \mathscr{O}_{X,a}$ be the maximal ideal in the local ring of a point $a \in X$ as in Definition 3.22.*

(a) *The ideal $I_a$ is principal, with $I_a = (\varphi_a)$ for a suitable $\varphi_a \in \mathscr{O}_{X,a}$ with $\operatorname{mult}_a(\varphi_a) = 1$.*

(b) *Any non-zero $\varphi \in \mathscr{O}_{X,a}$ can be written as $\varphi = c\cdot\varphi_a^m$, where $c \in \mathscr{O}_{X,a}\backslash I_a$ and $m = \operatorname{mult}_a(\varphi)$.*

*Proof.*

(a) Choose a linear function $\varphi_a$ vanishing at $a$ such that the line $V(\varphi_a)$ is not the tangent line to $X$ at $a$. Then $\varphi_a$ vanishes on $X$ with multiplicity 1 at $a$ by Exercise 12.27. Hence $\varphi_a \in I_a$, and

$$1 = \dim_K \mathscr{O}_{X,a}/(\varphi_a) \geq \dim_K \mathscr{O}_{X,a}/I_a > 0.$$

It follows that we must have equality, so in particular that $I_a = (\varphi_a)$.

(b) Note that $\varphi \notin I_a^{m+1}$, since by (a) the elements of $I_a^{m+1}$ are multiples of $\varphi_a^{m+1}$, and thus have multiplicity at least $m+1$ at $a$. Hence there is a maximal $n \in \mathbb{N}$ with $\varphi \in I_a^n$. By (a) this means $\varphi = c\cdot\varphi_a^n$ for some $c \in \mathscr{O}_{X,a}$. But we must have $c \notin I_a$ since $n$ is maximal. Hence $m = \operatorname{mult}_a(\varphi) = \operatorname{mult}_a(c\cdot\varphi_a^n) = n$, and the result follows. □

**Remark 14.13.** Thinking of a smooth curve $X \subset \mathbb{P}^2_{\mathbb{C}}$ as a 1-dimensional complex manifold, we can interpret the function $\varphi_a$ of Lemma 14.12 as a *local coordinate* on $X$ at $a$, i.e. as a function that gives an isomorphism of a neighborhood of $a$ with a neighborhood of the origin in $\mathbb{C}$ in the classical topology. By standard complex analysis, any local holomorphic function on $X$ at $a$ can then be written as a non-vanishing holomorphic function times a power of the local coordinate [G4, Lemma 10.4]. Lemma 14.12 (b) is just the corresponding algebraic statement. Note however that this is only a statement about the local ring — in contrast to the analytic setting it does not imply that $X$ has a Zariski-open neighborhood of $a$ isomorphic to an open subset of $\mathbb{A}^1_{\mathbb{C}}$!

**Remark 14.14** (Infinite multiplicity)**.** Let $X \subset \mathbb{P}^2$ be a smooth curve. For the following lemma, it is convenient to set formally $\operatorname{mult}_a(X,f) = \infty$ for all $a \in X$ if $f$ is a homogeneous polynomial that vanishes identically on $X$. Note that by Lemma 14.12 (b) we then have for an arbitrary homogeneous polynomial $f$ that $\operatorname{mult}_a(X,f) \geq m$ if and only if $f$ is a multiple of $\varphi_a^m$ in $\mathscr{O}_{X,a}$, where we interpret $f$ as an element in $\mathscr{O}_{X,a}$ as in Remark 12.25, and $\varphi_a$ is a local coordinate as in Lemma 14.12.

**Lemma 14.15.** *Let $X \subset \mathbb{P}^2$ be a smooth curve, and let $a \in X$ be a point.*

(a) *Let $f, g \in K[x_0, x_1, x_2]$ be homogeneous polynomials of the same degree with* $\operatorname{mult}_a(X, f) \geq m$ *and* $\operatorname{mult}_a(X, g) \geq m$ *for some $m \in \mathbb{N}$. Then:*

- $\operatorname{mult}_a(X, \lambda f + \mu g) \geq m$ *for all $\lambda, \mu \in K$;*
- *there are $\lambda, \mu \in K$, not both zero, such that* $\operatorname{mult}_a(X, \lambda f + \mu g) \geq m + 1$.

(b) *Let $Y \subset \mathbb{P}^2$ be another curve, and set $m = \operatorname{mult}_a(X, Y)$. If $f \in K[x_0, x_1, x_2]$ is a homogeneous polynomial with* $\operatorname{mult}_a(X, f) \geq m$, *then we also have* $\operatorname{mult}_a(Y, f) \geq m$.

*Proof.* As in Remark 12.25, we will consider $f$ and $g$ as elements in the local ring $\mathscr{O}_{X,a}$.

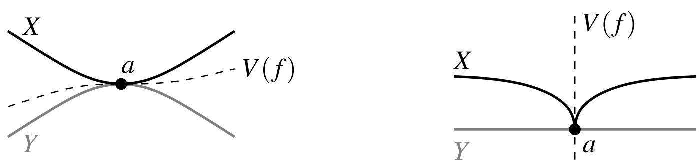
(a) We may assume that $f$ and $g$ do not vanish identically on $X$, since otherwise the statement is trivial. By Remark 14.14 there are then $u, v \in \mathscr{O}_{X,a}$ such that $f = u\,\varphi_a^m$ and $g = v\,\varphi_a^m$ in $\mathscr{O}_{X,a}$. So for any $\lambda, \mu \in K$, we have $\lambda f + \mu g = (\lambda u + \mu v)\,\varphi_a^m$, and thus $\operatorname{mult}_a(\lambda u + \mu v) \geq m$. Moreover, we can pick $\lambda$ and $\mu$ not both zero such that $\lambda u(a) + \mu v(a) = 0 \in K$. Then $\operatorname{mult}_a(\lambda u + \mu v) \geq 1$, and hence $\operatorname{mult}_a(\lambda f + \mu g) \geq m + 1$.

(b) As above we can assume that $f$ does not vanish identically on $X$ or $Y$. Let $g$ and $h$ be polynomials such that $I(X) = (g)$ and $I(Y) = (h)$. The assumption then means that $k := \operatorname{mult}_a(X, f) \geq m = \operatorname{mult}_a(X, h)$. Hence $f = u\,\varphi_a^k$ and $h = v\,\varphi_a^m$ for suitable units $u, v \in \mathscr{O}_{X,a}$ by Lemma 14.12 (b). This implies that $(f) \subset (h)$ in $\mathscr{O}_{X,a}$, so that $(f, g) \subset (g, h)$ in $\mathscr{O}_{\mathbb{A}^2,a}$. But then $(f, h) \subset (g, h)$ in $\mathscr{O}_{\mathbb{A}^2,a}$ as well, hence $\operatorname{mult}_a(f, h) \geq \operatorname{mult}_a(g, h)$, and thus $\operatorname{mult}_a(Y, f) \geq \operatorname{mult}_a(X, Y) = m$. $\square$

**Example 14.16.** The following examples show that the smoothness assumption in Lemma 14.15 (b) is crucial.

(a) Let $X$ and $Y$ be two smooth plane curves such that $\operatorname{mult}_a(X, Y) = 2$. By Exercise 12.27 this means that $X$ and $Y$ are tangent at $a$, as in the picture below on the left. Lemma 14.15 (b) then states that any polynomial vanishing on $X$ to order at least 2 at $a$ also vanishes on $Y$ to order at least 2 at $a$. In terms of the corresponding curve $V(f)$, this means using Exercise 12.27 again that any other curve that is (singular or) tangent to $X$ at $a$ is also (singular or) tangent to $Y$ at $a$ — which is obvious.

(b) In contrast to (a), in the picture on the right we have $X = V_p(x_2^3 - x_1^2 x_0)$, $Y = V_p(x_2)$, $f = x_1$, and thus at the origin $m = \operatorname{mult}_a(X, Y) = 2$, $\operatorname{mult}_a(X, f) = 3$, but $\operatorname{mult}_a(Y, f) = 1$. This fits well with the geometric interpretation of Remark 14.13: the curve $X$ is singular at the origin, so locally not a 1-dimensional complex manifold. Hence there is no local coordinate on $X$ around $a$, and the argument of Lemma 14.12 resp. Lemma 14.15 breaks down.



(a) (b)

**Lemma 14.17.** *Let $X \subset \mathbb{P}^2$ be a smooth curve, and let $g, h \in S(X)$ be two non-zero homogeneous polynomials.*

(a) *If $\operatorname{div} g = \operatorname{div} h$ then $g$ and $h$ are linearly dependent in $S(X)$.*

(b) *If $h$ is linear and $\operatorname{div} g \geq \operatorname{div} h$ then $h \mid g$ in $S(X)$.*

*Proof.* Let $f \in K[x_0, x_1, x_2]$ be a homogeneous polynomial with $I(X) = (f)$.

(a) By assumption we have $m_a := \operatorname{mult}_a(g) = \operatorname{mult}_a(h)$ for all $a \in X$. Moreover, Bézout's Theorem as in Corollary 12.26 (a) implies that $\sum_{a \in X} m_a = \deg X \cdot \deg g = \deg X \cdot \deg h$, so in particular we see already that $d := \deg g = \deg h$.

Now pick an arbitrary point $b \in X$. By Lemma 14.15 (a) there are $\lambda, \mu \in K$, not both zero, such that $\operatorname{mult}_a(\lambda g + \mu h) \geq m_a$ for all $a \in X$, and $\operatorname{mult}_b(\lambda g + \mu h) \geq m_b + 1$. Summing up, this means that $\sum_{a \in X} \operatorname{mult}_a(\lambda g + \mu h) \geq d \cdot \deg X + 1$. But $\lambda g + \mu h$ also has degree $d$, hence by Bézout again it follows that $\lambda f + \mu g$ must vanish identically on $X$, i. e. $\lambda g + \mu h = 0 \in S(X)$.

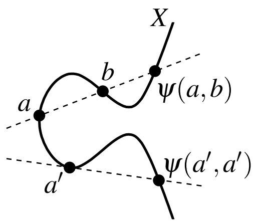
(b) Let $L = V(h)$, and choose a representative $\tilde{g} \in K[x_0, x_1, x_2]$ of $g \in S(X)$. We may assume that $\tilde{g}$ does not vanish identically on $L$, as otherwise $h \,|\, \tilde{g}$ in $K[x_0, x_1, x_2]$, and we are done.

By assumption, we have $\operatorname{mult}_a(X, \tilde{g}) \geq \operatorname{mult}_a(X, L)$ for all $a \in X \cap L$. As $X$ is smooth, Lemma 14.15 (b) shows that then $\operatorname{mult}_a(L, \tilde{g}) \geq \operatorname{mult}_a(X, L)$, and thus that $\operatorname{div} \tilde{g} \geq \operatorname{div} f$ on $L$. As $L \cong \mathbb{P}^1$, we can find a homogeneous polynomial $f' \in K[x_0, x_1, x_2]$ of degree $\deg g - \deg f$ with $\operatorname{div} f' = \operatorname{div} \tilde{g} - \operatorname{div} f$ on $L$ as in Example 14.11. Then $\operatorname{div}(ff') = \operatorname{div} \tilde{g}$ on $L$, which means by (a) that $ff'$ and $\tilde{g}$ are linearly dependent in $S(L)$. But then $\tilde{g} = ff' + ph$ for some homogeneous polynomial $p$ and after possibly multiplying $f'$ with a non-zero scalar, which means that $g = ph$ in $K[x_0, x_1, x_2]/(f) = S(X)$. □

24

We are now finally ready to prove that the Picard group of a smooth cubic curve in $\mathbb{P}^2$ is not trivial.

**Notation 14.18.** Let $a$ and $b$ be two points on a smooth cubic curve $X \subset \mathbb{P}^2$, not necessarily distinct. By Exercise 12.27 there is then a unique line $L \subset \mathbb{P}^2$ such that $a + b \leq L \cdot X$ as divisors on $X$ (in the sense of Definition 14.1 (b)), namely the line through $a$ and $b$ if these points are distinct, and the tangent line to $X$ at $a = b$ otherwise. But $L \cdot X$ is an effective divisor of degree 3 on $X$, and hence there is a unique point $c \in X$ (which need not be distinct from $a$ and $b$) with $L \cdot X = a + b + c$. In the following, we will denote this point $c$ by $\psi(a, b)$.




In geometric terms, for general $a, b \in X$ (i. e. such that none of the above points coincide) the point $\psi(a, b)$ is just the third point of intersection of $X$ with the line through $a$ and $b$. Hence the above definition is a generalization of our construction in Exercise 7.15. In fact, one can show that the map $\psi : X \times X \to X, (a, b) \mapsto \psi(a, b)$ is a morphism, but we will not need this result here.

**Proposition 14.19.** *Let $X \subset \mathbb{P}^2$ be a smooth cubic curve. Then for all distinct $a, b \in X$ we have $a - b \neq 0$ in $\operatorname{Pic}^0 X$, i. e. there is no non-zero rational function $\varphi$ on $X$ with $\operatorname{div} \varphi = a - b$.*

*Proof.* Assume for a contradiction that the statement of the proposition is false. Then there are a positive integer $d$ and homogeneous polynomials $f, g \in S(X)$ of degree $d$ such that the following conditions hold:

(a) There are points $a_1, \dots, a_{3d-1}$ and $a \neq b$ on $X$ such that
$$\operatorname{div} g = a_1 + \cdots + a_{3d-1} + a \quad \text{and} \quad \operatorname{div} f = a_1 + \cdots + a_{3d-1} + b$$
(hence $\operatorname{div} \varphi = a - b$ for $\varphi = \frac{g}{f}$).

(b) Among the $a_1, \dots, a_{3d-1}$ there are at least $2d$ distinct points. (If this is not the case in the first place, we can replace $f$ and $g$ by $f \cdot l$ and $g \cdot l$, respectively, for some homogeneous linear polynomial $l$ that vanishes on $X$ at three distinct points that are not among the $a_i$. This raises the degree of the polynomials by 1 and the number of distinct points by 3, so by doing this often enough we can get at least $2d$ distinct points.)

Pick $d$ minimal with these two properties.

If $d = 1$ then $\operatorname{div} g = a_1 + a_2 + a$ and $\operatorname{div} f = a_1 + a_2 + b$, so we must have $a = b = \psi(a_1, a_2)$ by Notation 14.18, in contradiction to our assumption. Hence we can assume that $d > 1$. Let us relabel the points $a_1,\dots,a_{3d-1}$ such that $a_2 \neq a_3$, and such that $a_1 = a_2$ if there are any equal points among the $a_i$.

Now consider linear combinations $\lambda f + \mu g$ for $\lambda, \mu \in K$, not both zero. As the polynomials $f$ and $g$ have different divisors they are linearly independent in $S(X)$, and hence $\lambda f + \mu g$ does not vanish identically on $X$. Moreover, by Lemma 14.15 (a) we have $a_1 + \cdots + a_{3d-1} \le \operatorname{div}(\lambda f + \mu g)$ for all $\lambda$ and $\mu$, and for any given $c \in X$ there are $\lambda$ and $\mu$ with $a_1 + \cdots + a_{3d-1} + c \le \operatorname{div}(\lambda f + \mu g)$. Of course, by Bézout's Theorem we must then have $\operatorname{div}(\lambda f + \mu g) = a_1 + \cdots + a_{3d-1} + c$.

In other words, by passing to linear combinations of $f$ and $g$ we can assume that the last points $a$ and $b$ in the divisors of $f$ and $g$ are any two points we like. Let us choose $a = \psi(a_1,a_2)$ and $b = \psi(a_1,a_3)$. Then

$$\begin{aligned} \operatorname{div} g &= (a_1 + a_2 + \psi(a_1,a_2)) + a_3 + a_4 + \cdots + a_{3d-1} \\ \text{and} \quad \operatorname{div} f &= (a_1 + a_3 + \psi(a_1,a_3)) + a_2 + a_4 + \cdots + a_{3d-1}. \end{aligned}$$

But $a_1 + a_2 + \psi(a_1,a_2)$ and $a_1 + a_3 + \psi(a_1,a_3)$ are divisors of homogeneous linear polynomials $k$ and $l$ in $S(X)$, respectively, and hence by Lemma 14.17 (b) there are homogeneous polynomials $f', g' \in S(X)$ of degree $d-1$ with $g = kg'$ and $f = lf'$, and thus with

$$\operatorname{div} g' = a_4 + \cdots + a_{3d-1} + a_3 \quad \text{and} \quad \operatorname{div} f' = a_4 + \cdots + a_{3d-1} + a_2.$$

Note that these new polynomials $f'$ and $g'$ satisfy (a) for $d$ replaced by $d-1$, as $a_3 \neq a_2$ by assumption. Moreover, $f'$ and $g'$ satisfy (b) because, if there are any equal points among the $a_i$ at all, then by our relabeling of these points there are only two distinct points among $a_1,a_2,a_3$, and so there must still be at least $2d-2$ distinct points among $a_4,\dots,a_{3d-1}$.

This contradicts the minimality of $d$, and therefore proves the proposition. □

**Remark 14.20.** In particular, Proposition 14.19 implies that $\operatorname{Pic}^0 X \neq \{0\}$ for any smooth cubic surface $X \subset \mathbb{P}^2$. So by Example 14.11 we can already see that $X$ is not isomorphic to $\mathbb{P}^1$. In fact, we will see in the next chapter that Proposition 14.19 suffices to compute $\operatorname{Pic}^0 X$ explicitly.

# 15. Elliptic Curves

At the end of the last chapter we have used Picard groups to show in Proposition 14.19 and Remark 14.20 that smooth cubic curves in $\mathbb{P}^2$ are not isomorphic to $\mathbb{P}^1$. In fact, if our ground field is not necessarily $\mathbb{C}$ (so that we cannot apply topological methods as in Remark 13.19), this is the first class of smooth projective curves for which we could prove rigorously that they are not isomorphic to $\mathbb{P}^1$. So let us now study these curves in more detail. We will see that they have a very rich structure, both from an algebraic and — over $\mathbb{C}$ — from an analytic point of view.

**Definition 15.1** (Elliptic curves)**.** In this chapter, by an **elliptic curve** we will simply mean a smooth cubic curve in $\mathbb{P}^2$.

Usually in the literature, an elliptic curve is defined to be a smooth complete curve of genus 1 (see Remark 13.13 and Exercise 13.20 for the definition of genus). Note that a smooth cubic curve in $\mathbb{P}^2$ is in fact complete by Example 7.22 (b) and of genus 1 by Example 13.18 (b) and Exercise 13.20. Conversely, one can show that every smooth complete curve of genus 1 can be embedded as a cubic curve in $\mathbb{P}^2$. Hence our somewhat non-standard definition of an elliptic curve is consistent with the literature.

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 13.9). 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 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 computing the degree-0 Picard group of an elliptic curve $X$, which (after the choice of a base point) turns out to be in natural bijection with $X$ itself.

**Proposition 15.2.** *Let $X \subset \mathbb{P}^2$ be an elliptic curve, and let $a_0 \in X$ be a point. Then the map*

$$\Phi : X \to \operatorname{Pic}^0 X, \quad a \mapsto a - a_0$$

*is a bijection.*

*Proof.* As $\deg(a - a_0) = 0$, the map $\Phi$ is clearly well-defined. It is also injective: if $\Phi(a) = \Phi(b)$ for $a, b \in X$ this means that $a - a_0 = b - a_0$, and hence $a - b = 0$, in $\operatorname{Pic}^0 X$. By Proposition 14.19 this is only possible if $a = b$.

To show that $\Phi$ is surjective, let $D$ be an arbitrary element of $\operatorname{Pic}^0 X$, which we can write as

$$D = a_1 + \cdots + a_m - b_1 - \cdots - b_m$$

for some $m \in \mathbb{N}_{>0}$ and not necessarily distinct $a_1, \dots, a_m, b_1, \dots, b_m \in X$. Assume first that $m \geq 2$. Then there are homogeneous linear polynomials $l, l'$ on $X$ such that $\operatorname{div} l = a_1 + a_2 + \psi(a_1, a_2)$ and $\operatorname{div} l' = b_1 + b_2 + \psi(b_1, b_2)$, where $\psi$ is as in Notation 14.18. The quotient of these polynomials is then a rational function on $X$, whose divisor $a_1 + a_2 + \psi(a_1, a_2) - b_1 - b_2 - \psi(b_1, b_2)$ is therefore zero in $\operatorname{Pic}^0 X$. It follows that we can also write

$$D = \psi(b_1, b_2) + a_3 + \cdots + a_m - \psi(a_1, a_2) - b_3 - \cdots - b_m \quad \in \operatorname{Pic}^0 X.$$

We have thus reduced the number $m$ of (positive and negative) points in $D$ by 1. Continuing this process, we can assume that $m = 1$, i. e. that $D = a_1 - b_1$ for some $a_1, b_1 \in X$.

In the same way, we then also have

$$a_0 + a_1 + \psi(a_0, a_1) - b_1 - \psi(a_0, a_1) - \psi(b_1, \psi(a_0, a_1)) = 0 \quad \in \operatorname{Pic}^0 X,$$

so that $D = a_1 - b_1 = \psi(b_1, \psi(a_0, a_1)) - a_0 \in \operatorname{Pic}^0 X$. Hence $D = \Phi(\psi(b_1, \psi(a_0, a_1)))$, i. e. $\Phi$ is surjective. $\square$

**Remark 15.3.** Let $X \subset \mathbb{P}^2$ be an elliptic curve. After choosing a base point $a_0 \in X$, Proposition 15.2 gives a canonical bijection between the variety $X$ and the Abelian group $\operatorname{Pic}^0 X$, i. e. between two totally different mathematical objects. So we can use this bijection to give $X$ the structure of an Abelian group, and $\operatorname{Pic}^0 X$ the structure of a smooth projective variety.

In fact, $\operatorname{Pic}^0 X$ can be made into a variety (the so-called *Picard variety*) for every smooth projective curve $X$. It is in general not isomorphic to $X$, however. One can only show that the map $\Phi : X \to \operatorname{Pic}^0 X$, $a \mapsto a - a_0$ of Proposition 15.2 is injective if $X$ is not $\mathbb{P}^1$, so that we can then think of $X$ as a subvariety of the Picard variety.

In contrast, the statement that $X$ can be made into an Abelian group is very special to elliptic curves. In the following, we want to explore this group structure in more detail.

**Construction 15.4** (The group structure on an elliptic curve)**.** Let $a_0$ be a fixed base point on an elliptic curve $X \subset \mathbb{P}^2$. As in Remark 15.3, we can use Proposition 15.2 to define a group structure on $X$. More precisely, if we denote this group operation by the symbol $\oplus$ (to distinguish it from the addition of points in $\operatorname{Div} X$ or $\operatorname{Pic} X$), then $a \oplus b$ for $a, b \in X$ is the unique point of $X$ satisfying

$$\Phi(a \oplus b) = \Phi(a) + \Phi(b).$$

To find an explicit description for $a \oplus b$, note that — as in the proof of Proposition 15.2 — both $a + b + \psi(a, b)$ and $a_0 + \psi(a, b) + \psi(a_0, \psi(a, b))$ are divisors of homogeneous linear polynomials, and thus

$$a + b + \psi(a,b) - a_0 - \psi(a,b) - \psi(a_0, \psi(a,b)) = 0 \quad \in \operatorname{Pic}^0 X.$$

Hence

$$\begin{aligned} a \oplus b &= \Phi^{-1}(\Phi(a) + \Phi(b)) \\ &= \Phi^{-1}(a - a_0 + b - a_0) \\ &= \Phi^{-1}(\psi(a_0, \psi(a,b)) - a_0) \\ &= \psi(a_0, \psi(a,b)). \end{aligned}$$

In other words, to construct the point $a \oplus b$ we draw a line through $a$ and $b$. Then we draw another line through the third intersection point $\psi(a, b)$ of this line with $X$ and the point $a_0$. The third intersection point of this second line with $X$ is then $a \oplus b$, as in the picture below on the left.

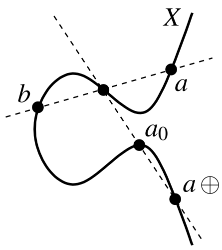
Similarly, to construct the inverse $\ominus a$ of $a$ in the above group structure we use the relation

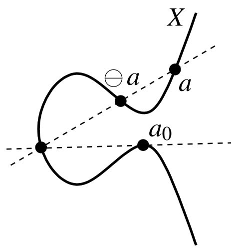
$$a_0 + a_0 + \psi(a_0, a_0) - a - \psi(a_0, a_0) - \psi(a, \psi(a_0, a_0)) = 0 \quad \in \operatorname{Pic}^0 X$$

to obtain

$$\begin{aligned} \ominus a &= \Phi^{-1}(-\Phi(a)) \\ &= \Phi^{-1}(a_0 - a) \\ &= \Phi^{-1}(\psi(a, \psi(a_0, a_0)) - a_0) \\ &= \psi(a, \psi(a_0, a_0)). \end{aligned}$$

So to construct the inverse $\ominus a$ we draw the tangent to $X$ through $a_0$. Then we draw another line through the other intersection point $\psi(a_0, a_0)$ of this tangent with $X$ and the point $a$. The third intersection point of this second line with $X$ is $\ominus a$, 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.

**Exercise 15.5.** Let $X$ and $Y$ be two distinct elliptic curves in $\mathbb{P}^2$, and assume that they intersect in 9 distinct points $a_1,\dots,a_9$. Prove that every elliptic curve passing through $a_1,\dots,a_8$ also has to pass through $a_9$.

Can you find a stronger version of this statement that applies in the case when the intersection multiplicities in $X \cap Y$ are not all equal to 1?

**Example 15.6** (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 $X$ and a base point $a_0 \in X$ for the group structure.

(a) Given $a \in X$ and $n \in \mathbb{N}$, the $n$-fold addition $n \odot a := a \oplus \cdots \oplus a$ can be computed very quickly, even for very large $n$ (think of numbers with hundreds of digits):

- By repeatedly applying the operation $a \mapsto a \oplus a$, we can compute all points $2^k \odot a$ for all $k$ such that $2^k \le n$.
- Now we just have to add these points $2^k \odot a$ for all $k$ such that the $k$-th digit in the binary representation of $n$ is 1.

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

(b) On the other hand, given two sufficiently general points $a, b \in X$ it is essentially impossible to compute an integer $n \in \mathbb{N}$ such that $n \odot a = b$ (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 $X$ over $K$, a base point $a_0 \in X$, and another point $a \in X$. Now Alice picks a secret (very large) integer $n$, computes $n \odot a$ as in (a), and sends (the coordinates of) this point to Bob. In the same way, Bob chooses a secret number $m$, computes $m \odot a$, and sends this point to Alice.

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

**Exercise 15.7.** Let $X$ be an elliptic curve of the form

$$X = \{(x_0 : x_1 : x_2) : x_2^2 x_0 = x_1^3 + \lambda x_1 x_0^2 + \mu x_0^3\} \quad \subset \mathbb{P}^2$$

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 $a_0 = (0:0:1)$ as the base point for the group structure on $X$. For given points $b=(b_0:b_1:b_2)$ and $c=(c_0:c_1:c_2)$ compute explicitly the coordinates of the inverse $\ominus b$ and of the sum $b \oplus c$. Conclude that the group structure on $X$ is well-defined even if the ground field $K$ is not necessarily algebraically closed. (This is important for practical computations, where one usually wants to work over finite fields in order to avoid rounding errors.)

25

Let us now restrict our attention to the ground field $\mathbb{C}$, so that an elliptic curve is topologically a torus by Example 13.18 (b). In the remaining part of this chapter we want to see how these tori arise in complex analysis in a totally different way. As we have not developed any analytic techniques in this class we will only sketch most arguments; more details can be found e. g. in [K, Section 5.1] (and many other books on complex analysis). Let us start by giving a quick review of what we will need from standard complex analysis.

**Remark 15.8** (Holomorphic and meromorphic functions)**.** Let $U \subset \mathbb{C}$ be an open set in the classical topology. Recall that a function $f: U \to \mathbb{C}$ 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 $f: U \to \mathbb{C} \cup \{\infty\}$ 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 $V$ of $z_0$ in $U$ such that

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

on $V$. If $f$ does not vanish identically in a neighborhood of $z_0$ we can moreover assume $\tilde{f}(z_0) \neq 0$ in this representation; the number $n$ is then uniquely determined. We will call it the *order* of $f$ at $z_0$ and denote it by $\operatorname{ord}_{z_0} f$. It is the analogue of the multiplicity of a rational function in Construction 14.5. If $n > 0$ we say that $f$ has a *zero* of order $n$ at $z_0$; if $n < 0$ then $f$ has a *pole* of order $-n$ there. A meromorphic function is holomorphic around a point $z_0$ if and only if its order at this point is non-negative.

Of course, every regular (resp. rational) function on a Zariski-open subset of $\mathbb{A}^1_{\mathbb{C}} = \mathbb{C}$ is holomorphic (resp. meromorphic). However, there are many holomorphic (resp. meromorphic) functions that are not regular (resp. rational), e. g. $f: \mathbb{C} \to \mathbb{C},\ z \mapsto e^z$.

**Remark 15.9** (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 is automatically infinitely differentiable: all higher derivatives $f^{(k)}$ for $k \in \mathbb{N}$ exist and are again holomorphic [G4, Corollary 8.1].

(b) Every holomorphic function $f$ is analytic, i. e. it can be represented locally around every point $z_0$ by its Taylor series. The radius of convergence is "as large as it can be", i. e. if $f$ is holomorphic in an open ball $U$ around $z_0$, then the Taylor series of $f$ at $z_0$ converges and represents $f$ at least on $U$. Consequently, a meromorphic function $f$ of order $n$ at $z_0$ can be expanded in a *Laurent series* as $f(z) = \sum_{k=n}^{\infty} c_k (z-z_0)^k$ [G4, Proposition 9.8]. The coefficient $c_{-1}$ of this series is called the *residue* of $f$ at $z_0$ and denoted by $\operatorname{res}_{z_0} f$.

   Residues are related to orders of meromorphic functions as follows: if $f(z) = (z-z_0)^n \tilde{f}(z)$ as in Remark 15.8 above, we obtain

$$\operatorname{res}_{z_0} \frac{f'(z)}{f(z)} = \operatorname{res}_{z_0} \left( \frac{n}{z-z_0} + \frac{\tilde{f}'(z)}{\tilde{f}(z)} \right) = n = \operatorname{ord}_{z_0} f.$$

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

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

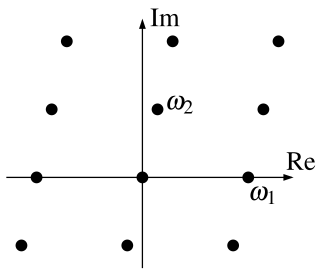
with the sum taken over all $z_0$ in the interior of $\gamma$ (at which $f$ has poles) [G4, Proposition 11.13]. In particular, if $f$ is holomorphic in the interior of $\gamma$ then this integral vanishes.

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

For our applications to elliptic curves we will need a particular meromorphic function. To describe its construction, 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. Note that $\Lambda$ is an additive subgroup of $\mathbb{C}$, and that the quotient $\mathbb{C}/\Lambda$ is topologically a torus. We want to see that it can be identified with an elliptic curve in a natural way, using a map that we are going to introduce now.




**Proposition and Definition 15.10** (The Weierstraß $\wp$-function)**.** *Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a lattice in $\mathbb{C}$. 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 $\omega$ 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 $\omega^3$. Let us add up these values according to the absolute value of $\omega$. 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 behaves like $\sum_{n=1}^\infty n \cdot \frac{1}{n^3} = \sum_{n=1}^\infty \frac{1}{n^2}$, 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 $\Lambda$ are clearly visible. □

**Remark 15.11** (Properties of the $\wp$-function)**.** It is a standard fact 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 contains only even exponents.

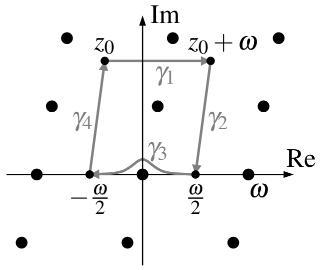
(b) Its derivative is $\wp'(z) = \sum_{\omega \in \Lambda} \frac{-2}{(z-\omega)^3}$. It is an odd function, i. e. $\wp'(z) = -\wp'(-z)$ for all $z$. In other words, its Laurent series at 0 contains only odd exponents. It has poles of order 3 exactly at the lattice points.

(c) The $\wp$-function is doubly periodic with respect to $\Lambda$, i. e. $\wp(z_0) = \wp(z_0 + \omega)$ for all $z_0 \in \mathbb{C}$ and $\omega \in \Lambda$. To show this note first that it is obvious from (b) that $\wp'(z_0) = \wp'(z_0 + \omega)$. Now integrate $\wp'(z)$ along the closed contour $\gamma = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$ 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 $\gamma_2$ cancels the integral along $\gamma_4$ as $\wp'(z)$ is periodic. The integral along $\gamma_3$ 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. $\wp$ is periodic with respect to $\Lambda$.




**Lemma 15.12.** *The $\wp$-function associated to a lattice $\Lambda$ 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 $\Lambda$).*

*Proof.* By Remark 15.11 (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

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

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 $\Lambda$-periodic by Remark 15.11 (c). Hence $f$ is holomorphic around all lattice points. But $f$ is also holomorphic around all other points, as $\wp$ and $\wp'$ are. In other words, $f$ is holomorphic on all of $\mathbb{C}$.

Moreover, the 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 15.9 (d) 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. $\square$

**Remark 15.13.** By an explicit computation one can show that the coefficients $c_3, c_2, c_1, c_0$ in Lemma 15.12 are given by

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

The proof of Lemma 15.12 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 15.12 is a (non-homogeneous) cubic equation in the two functions $\wp$ and $\wp'$, which are $\Lambda$-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 15.14.** *Let $\Lambda \subset \mathbb{C}$ be a fixed lattice, and let $X \subset \mathbb{P}^2_{\mathbb{C}}$ be the cubic curve*

$$X = \{(x_0 : x_1 : x_2) : x_2^2 x_0 = c_3 x_1^3 + c_2 x_1^2 x_0 + c_1 x_1 x_0^2 + c_0 x_0^3\}$$

*for the constants $c_3, c_2, c_1, c_0 \in \mathbb{C}$ of Lemma 15.12. Then there is a bijection*

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

*Proof.* As $\wp$ and $\wp'$ are periodic with respect to $\Lambda$ and satisfy the differential equation of Lemma 15.12, it is clear that $\Psi$ is well-defined as a map to $X$. (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 $\Psi(0)$ formally looks like $(1 : \infty : \infty)$. But by Remark 15.8 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 $\Psi$ as

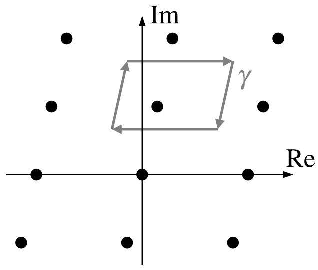
$$\Psi(0) = \lim_{z \to 0} (1 : \wp(z) : \wp'(z)) = \lim_{z \to 0} (z^3 : z f(z) : g(z)) = (0 : 0 : 1),$$

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

Now let $(x_0 : x_1 : x_2) \in X$ be a given point; we will show that it has exactly one inverse image point under $\Psi$. 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 will first look for a number $z \in \mathbb{C}$ such that $\wp(z) = x_1$. To do so, consider the integral

$$\int_\gamma \frac{\wp'(z)}{\wp(z) - x_1} \, dz$$




over the boundary of any "parallelogram of periodicity" (that does not meet the zeroes and poles of the function $z \mapsto \wp(z) - x_1$), as in the picture on the right. The integrals along opposite sides of the parallelogram vanish because of the periodicity of $\wp$ and $\wp'$, so that the total integral must be 0. Hence by Remark 15.9 (b) and (c) we get

$$0 = \sum_{z_0 \in \mathbb{C}/\Lambda} \mathrm{res}_{z_0} \frac{\wp'(z)}{\wp(z) - x_1} = \sum_{z_0 \in \mathbb{C}/\Lambda} \mathrm{ord}_{z_0}(\wp(z) - x_1).$$

In other words, the function $z \mapsto \wp(z) - x_1$ has as many zeroes as it has poles in $\mathbb{C}/\Lambda$, counted with multiplicities. (This is a statement in complex analysis analogous to the algebraic result of Remark 14.8 (b).) As $\wp$ has a pole of order 2 in the lattice points, it thus follows that there are exactly two points in $\wp^{-1}(x_1)$, counted with multiplicities.

For such a point $z$ with $\wp(z) = x_1$ we then have by Lemma 15.12

$$\wp'(z)^2 = c_3 \wp(z)^3 + c_2 \wp(z)^2 + c_1 \wp(z) + c_0 = c_3 x_1^2 + x_2 x_1^2 + c_1 x_1 + c_0 = x_2^2$$

since $(1 : x_1 : x_2) \in X$. 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 $z \mapsto \wp(z) - x_1$. 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 $z \mapsto \wp(z) - x_1$. As $\wp$ is even and $\wp'$ odd by Remark 15.11, 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 $\Psi$.

Altogether we conclude that $\Psi$ is bijective, as we have claimed. □

**Remark 15.15.** With Proposition 15.14 we are again in a similar situation as in Proposition 15.2: we have a bijection between a group $\mathbb{C}/\Lambda$ and a variety $X$, so that the map $\Psi$ of the above proposition can be used to construct a group structure on $X$. In fact, we will see in Exercise 15.17 that this group structure is precisely the same as that obtained by the map $\Phi$ of Proposition 15.2 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

$$\frac{1}{n}(i\omega_1 + j\omega_2) \quad \text{for } 0 \le i,j < n.$$

It should be said however that the analytic bijection of Proposition 15.14 differs from that of Proposition 15.2 in that both $\mathbb{C}/\Lambda$ and $X$ can independently be made into a 1-dimensional complex manifold, and the map $\Psi$ of the above proposition is then an isomorphism between these two manifolds.

**Exercise 15.16.** Using the identification of an elliptic curve $X$ with a torus $\mathbb{C}/\Lambda$ as in Proposition 15.14, reprove the statement of Proposition 14.19 that there is no rational function $\varphi$ on an elliptic curve $X$ with divisor $\operatorname{div}\varphi = a - b$ for distinct points $a, b \in X$.

**Exercise 15.17.** Let $X$ be an elliptic curve corresponding to a torus $\mathbb{C}/\Lambda$. Show that the group structure of $\operatorname{Pic}_X^0$ is isomorphic to the natural group structure of $\mathbb{C}/\Lambda$.

**Exercise 15.18.** Let $\Lambda \subset \mathbb{C}$ be a lattice. Given two points $z, w \in \mathbb{C}/\Lambda$, it is obviously 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 15.6? 26

# References


[AM] M. Atiyah, I. MacDonald, *Introduction to Commutative Algebra*, Addison Wesley (2004)
[E] D. Eisenbud, *Commutative Algebra with a View towards Algebraic Geometry*, Springer (2004)
[EGA] A. Grothendieck, J. Dieudonné, *Éléments de Géométrie Algébrique*, Publications Mathématiques IHES (various volumes)
[G1] A. Gathmann, *Algebraische Strukturen*, class notes TU Kaiserslautern (2017/18), www.mathematik.uni-kl.de/~gathmann/ags
[G2] A. Gathmann, *Grundlagen der Mathematik*, class notes TU Kaiserslautern (2015/16), www.mathematik.uni-kl.de/~gathmann/gdm
[G3] A. Gathmann, *Einführung in die Algebra*, class notes TU Kaiserslautern (2010/11), www.mathematik.uni-kl.de/~gathmann/algebra
[G4] A. Gathmann, *Einführung in die Funktionentheorie*, class notes TU Kaiserslautern (2016/17), www.mathematik.uni-kl.de/~gathmann/futheo
[G5] A. Gathmann, *Commutative Algebra*, class notes TU Kaiserslautern (2014), www.mathematik.uni-kl.de/~gathmann/commalg
[GP] G.-M. Greuel, G. Pfister, *A Singular Introduction to Commutative Algebra*, Springer (2002)
[H] R. Hartshorne, *Algebraic Geometry*, Springer Graduate Texts in Mathematics 52 (1977)
[Ha] J. Harris, *Algebraic Geometry*, Springer Graduate Texts in Mathematics 133 (1992)
[K] F. Kirwan, *Complex Algebraic Curves*, London Mathematical Society Student Texts 23, Cambridge University Press (1992)
[M] R. Miranda, *Algebraic Curves and Riemann Surfaces*, AMS Graduate Studies in Mathematics 5 (1995)
[M1] D. Mumford, *The Red Book of Varieties and Schemes*, Springer Lecture Notes in Mathematics 1358 (1988)
[M2] D. Mumford, *Algebraic Geometry I — Complex Projective Varieties*, Springer Classics in Mathematics (1976)
[R] M. Reid, *Undergraduate Algebraic Geometry*, London Mathematical Society Student Texts 12, Cambridge University Press (1988)
[S] W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, SINGULAR — a computer algebra system for polynomial computations, www.singular.uni-kl.de
[S1] I. Shafarevich, *Basic Algebraic Geometry I — Varieties in Projective Space*, Springer (1994)
[S2] I. Shafarevich, *Basic Algebraic Geometry II — Schemes and Complex Manifolds*, Springer (1994)


Note: This is a very extensive list of literature of varying usefulness. Here is a short recommendation which of the references you might want to use for what:

- For motivational aspects, examples, and a generally “fairy-tale” style introduction without much theoretical background, see [Ha], [R], or maybe [S1] and [S2].
- For questions concerning complex curves, in particular in relation to complex analysis, see [K] or [M].
- For a general reference on the commutative algebra background, see [AM] or [G5].
- For commutative algebra problems involving computational aspects, see [GP].
- For a good book that develops the theory thoroughly, but largely lacks motivations and examples, see [H]. You should not try to read the “hard-core” parts of this book without some motivational background.
- For the ultimate reference (“if it is not proven there, it must be wrong”), see [EGA]. Warning: this is unreadable if you do not have a decent background in algebraic geometry yet, and close to being unreadable even if you do.

# Index