GEIR ELLINGSRUD

JOHN CHRISTIAN OTTEM

MAT4210 – ALGEBRAIC GEOMETRY I

Contents

Acknowledgements

Thanks to Georges Elencwajg, Frank Gounelas, and Kristian Ranestad for comments and suggestions. Any further comments or corrections are welcome:

https://docs.google.com/document/d/17NlPG7tP_zrEWK1dYF93rNAi5sqY_mA_27QOJRG4IHM/edit

Introduction

Algebraic geometry has many ramifications, but roughly speaking there are two main branches. One could be called the "geometric" branch where the geometry is the main objective. One studies geometric objects like curves, surfaces, threefolds and varieties of higher dimensions, defined by polynomials (or more generally algebraic functions). The aim is to understand their geometry. Frequently techniques from several other fields are used like from algebraic topology, differential geometry or analysis, and the theory is tightly connected with these other branches of mathematics. This makes it natural to work over the complex field $\mathbb{C}$, even though other fields like function fields are also important. In fact, this interaction with other fields has been paramount in the development of algebraic geometry since the very beginning. The study of elliptic functions in the beginning of the 19th century, and subsequently of other algebraic functions, was the birth of modern algebraic geometry. The motivation and the origin was found in function theory, but the direction of research quickly took a geometric route. Later Riemann surfaces and algebraic curves appeared together with their function fields.

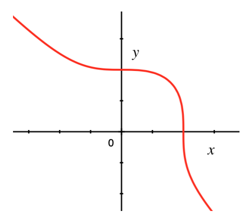


Figure 1: The affine Fermat curve $x^3 + y^3 = 1$.

The other branch one could call "arithmetic". Superficially presented, one studies numbers by geometric methods. An ultra famous example is Fermat's last theorem, now Andrew Wiles' theorem, that the equation $x^n + y^n = z^n$ has no integral solutions except the trivial ones for $n \geqslant 3$. The arithmetic branch also relies on techniques from other fields, like number theory, Galois theory and representation theory. One very commonly applied technique is reduction modulo a prime number p. Hence the importance of including fields of positive characteristic among the base fields. Of course another very natural base field for many of these "arithmetic" studies is the field $\overline{\mathbb{Q}}$ of algebraic numbers.

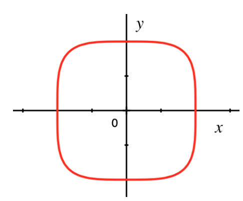


Figure 2: The affine Fermat curve $x^4 + y^4 = 1$.

Algebraic geometry is to the common benefit a triple marriage of geometry, algebra and arithmetic. All of the spouses claim influence on the development of the field which makes the field quite abstract; but also an immensely beautiful part of mathematics.

Chapter 1

Algebraic sets and the Nullstellensatz

Topics in Chapter 1: *The correspondence between ideals and algebraic sets – weak and strong versions of Hilbert's Nullstellensatz – the Rabinowitsch trick – two proofs of the Nullstellensatz, one elementary, and another totally different – radical ideals – intuition, drawings and figures.*

We shall almost exclusively work over an algebraic closed field which we shall denote by k. In general we do not impose further constraints on k, except for a few results that require the characteristic to be zero. A specific field to have in mind would be the field of complex numbers $\mathbb{C}$, but other important fields include the closures $\overline{\mathbb{Q}}$ of the rational numbers $\mathbb{Q}$ and $\overline{\mathbb{F}}_p$, the algebraic closure of the finite field $\mathbb{F}_p$ with p elements.

The *affine n-space* $\mathbb{A}^n$ is, as a set, simply k^n. The name change is here to underline that there is more to $\mathbb{A}^n$ than just its set of elements – we will soon equip it with the structure of a topological space, and we shall introduce the notion of polynomial maps $\mathbb{A}^n \to k$. Anyhow, in the beginning think about $\mathbb{A}^n$ as just k^n. Often the ground field will be tacitly understood, but when we want to be precise about it, we shall write $\mathbb{A}^2_k$. The ground field will always be algebraically closed unless the contrary is explicitly stated.

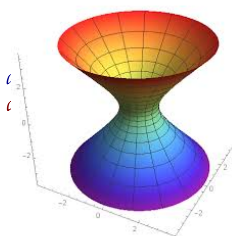


Figure 1.1: A one sheeted-hyperboloid.

1.1 *Closed algebraic sets*

Closed algebraic sets
lukkede algebraiske mengder

The first objects from algebraic geometry we shall meet, are the so-called *closed algebraic sets*, or simply *algebraic sets*. They are the subsets of the affine space $\mathbb{A}^n$ defined by a set of polynomial equations:

Definition *If S is a subset of the polynomial ring $k[x_1, \ldots, x_n]$, we define*

$$Z(S) = \{\, x \in \mathbb{A}^n \mid f(x) = 0 \text{ for all } f \in S \,\},$$

and the closed algebraic sets *are the subsets of $\mathbb{A}^n$ of this form.*

$\mathfrak{a}, \mathfrak{A}$	$\mathfrak{b}, \mathfrak{B}$	$\mathfrak{c}, \mathfrak{C}$	$\mathfrak{d}, \mathfrak{D}$	$\mathfrak{e}, \mathfrak{E}$
$\mathfrak{f}, \mathfrak{F}$	$\mathfrak{g}, \mathfrak{G}$	$\mathfrak{h}, \mathfrak{H}$	$\mathfrak{i}, \mathfrak{I}$	$\mathfrak{j}, \mathfrak{J}$
$\mathfrak{k}, \mathfrak{K}$	$\mathfrak{l}, \mathfrak{L}$	$\mathfrak{m}, \mathfrak{M}$	$\mathfrak{n}, \mathfrak{N}$	$\mathfrak{o}, \mathfrak{O}$
$\mathfrak{p}, \mathfrak{P}$	$\mathfrak{q}, \mathfrak{Q}$	$\mathfrak{r}, \mathfrak{R}$	$\mathfrak{s}, \mathfrak{S}$	$\mathfrak{t}, \mathfrak{T}$
$\mathfrak{u}, \mathfrak{U}$	$\mathfrak{v}, \mathfrak{V}$	$\mathfrak{w}, \mathfrak{W}$	$\mathfrak{x}, \mathfrak{X}$	$\mathfrak{y}, \mathfrak{Y}$
$\mathfrak{z}, \mathfrak{Z}$				

Mathematicians are always in shortage of symbols and use all kinds of alphabets. The germanic gothic letters are still in use in some context, like to denote ideals in some text.

Note that any expression of the form $\sum b_i f_i$, with b_i polynomials and $f_1, \ldots, f_r \in S$, also vanishes at points of $Z(S)$. The *ideal* $\mathfrak{a}$ generated by S therefore has the same zero set as S; that is, $Z(S) = Z(\mathfrak{a})$. We will therefore almost exclusively work with ideals and tacitly replace a set of polynomials by the ideal it generates.

Hilbert's basis theorem tells us that any ideal in $k[x_1, \ldots, x_n]$ is finitely generated, so that a closed algebraic subset is always described as the set of common zeros of *finitely* many polynomials.

Examples

1.2 The polynomial ring $k[x]$ in one variable is a PID[1], so if $\mathfrak{a}$ is an ideal in $k[x]$, then $\mathfrak{a} = (f(x))$. Because polynomials in one variable have only finitely many zeros, the closed algebraic subsets of $\mathbb{A}^1$ are just the finite subsets of $\mathbb{A}^1$.

[1] Recall that a ring is a PID, or a *principal ideal domain*, if it is an integral domain where every ideal is principal, *i.e.* generated by one element.

1.3 The traditional conic sections are closed algebraic sets in $\mathbb{A}^2$. For instance, the zeros of $y - x^2$ form a *parabola* and $xy - 1$ gives a *hyperbola*. Again, the usual curves are formed by the real solutions in $\mathbb{A}^2(\mathbb{R})$.

1.4 A more interesting example is the so-called *Clebsch diagonal cubic*; a surface in $\mathbb{A}^3(\mathbb{C})$ with equation

$$x^3 + y^3 + z^3 + 1 = (x + y + z + 1)^3.$$

An old plaster model of its real points; that is, the points in $\mathbb{A}^3(\mathbb{R})$ satisfying the equation, is depicted in the margin.

The Clebsch diagonal cubic

☆

Simple properties of $Z(\mathfrak{a})$

1.5 The more constraints one imposes, the smaller the solution set will be, so if $\mathfrak{b} \subseteq \mathfrak{a}$ are two ideals, one has $Z(\mathfrak{a}) \subseteq Z(\mathfrak{b})$. Sending $\mathfrak{a}$ to $Z(\mathfrak{a})$ is thus an inclusion-reversing map from the partially ordered set of ideals in $k[x_1, \ldots, x_n]$ to the partially ordered set of subsets of $\mathbb{A}^n$.

Recall that the *sum* $\mathfrak{a} + \mathfrak{b}$ of two ideals is given by

$$\mathfrak{a} + \mathfrak{b} = \{\, f + g \mid f \in \mathfrak{a} \text{ and } g \in \mathfrak{b} \,\},$$

This ideal has the intersection $Z(\mathfrak{a}) \cap Z(\mathfrak{b})$ as zero set. Indeed, both $\mathfrak{a}$ and $\mathfrak{b}$ are contained in $\mathfrak{a} + \mathfrak{b}$ so that $Z(\mathfrak{a} + \mathfrak{b}) \subseteq Z(\mathfrak{a}) \cap Z(\mathfrak{b})$, and the reverse inclusion holds since all polynomials in $\mathfrak{a}$ and all polynomials in $\mathfrak{b}$ vanish in $Z(\mathfrak{a}) \cap Z(\mathfrak{b})$.

In the same vein, the *product* $\mathfrak{a} \cdot \mathfrak{b}$ defines the union $Z(\mathfrak{a}) \cup Z(\mathfrak{b})$. This follows because $\mathfrak{a} \cdot \mathfrak{b}$ is generated by the products $f \cdot g$ of polynomials $f \in \mathfrak{a}$ and $g \in \mathfrak{b}$.

1.6 It might very well happen that two different ideals define the same algebraic set. The simplest example is (x) and (x^2); they both define the origin in the affine line $\mathbb{A}^1$. More generally, powers $\mathfrak{a}^n$ of an ideal $\mathfrak{a}$ have the same zeros as $\mathfrak{a}$:

Because $\mathfrak{a}^n \subseteq \mathfrak{a}$ it holds that $Z(\mathfrak{a}) \subseteq Z(\mathfrak{a}^n)$, and the other inclusion holds as well since $f^n \in \mathfrak{a}^n$ whenever $f \in \mathfrak{a}$. Recall that the *radical* $\sqrt{\mathfrak{a}}$ of an ideal is defined as the ideal

$$\sqrt{\mathfrak{a}} = \{\, f \mid f^r \in \mathfrak{a} \text{ for some } r > 0 \,\}.$$

The argument above yields that $Z(\mathfrak{a}) = Z(\sqrt{\mathfrak{a}})$ (in fact, since all ideals in the polynomial ring are finitely generated, a power of the radical is contained in $\mathfrak{a}$). Ideals with the same radical therefore have coinciding zero sets, and we shall soon see that the converse is true as well.

We sum up the present discussion in a proposition:

PROPOSITION 1.7 *Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in $k[x_1, \ldots, x_n]$.*

i) *If $\mathfrak{a} \subseteq \mathfrak{b}$, then $Z(\mathfrak{b}) \subseteq Z(\mathfrak{a})$;*
ii) $Z(\mathfrak{a} + \mathfrak{b}) = Z(\mathfrak{a}) \cap Z(\mathfrak{b})$;
iii) $Z(\mathfrak{a}\mathfrak{b}) = Z(\mathfrak{a} \cap \mathfrak{b}) = Z(\mathfrak{a}) \cup Z(\mathfrak{b})$;
iv) $Z(\mathfrak{a}) = Z(\sqrt{\mathfrak{a}})$.

David Hilbert (1862–1943) German mathematician.

By the way, this also shows that $Z(\mathfrak{a} \cap \mathfrak{b}) = Z(\mathfrak{a}) \cup Z(\mathfrak{b})$: Because of the inclusion $(\mathfrak{a} \cap \mathfrak{b})^2 \subseteq \mathfrak{a} \cdot \mathfrak{b}$ one has $Z(\mathfrak{a} \cap \mathfrak{b}) \subseteq Z(\mathfrak{a}) \cup Z(\mathfrak{b})$, and the other inclusion follows readily as $\mathfrak{a} \cap \mathfrak{b}$ is contained in both $\mathfrak{a}$ and $\mathfrak{b}$.

Notice also that the argument for the second assertion remains valid for any family of ideals $\{\mathfrak{a}_i\}_{i \in I}$. That is, one has

❏ $Z(\sum_{i \in I} \mathfrak{a}_i) = \bigcap_{i \in I} Z(\mathfrak{a}_i)$.

The ideal $I(X)$ of functions vanishing on a set

1.8 There is a partial converse to the construction of the zero locus $Z(\mathfrak{a})$ of an ideal:

DEFINITION *For a subset X of $\mathbb{A}^n$ we define the ideal $I(X)$ consisting of polynomials in $k[x_1, \ldots, x_n]$ that vanish along X; that is,*

$$I(X) = \{\, f \in k[x_1, \ldots, x_n] \mid f(x) = 0 \textit{ for all } x \in X \,\}.$$

This gives an inclusion-reversing map from the set of subsets of $\mathbb{A}^n$ to the set of ideals in the polynomial ring $k[x_1, \ldots, x_n]$. It holds tautologically true that $\mathfrak{a} \subseteq I(Z(\mathfrak{a}))$ (functions vanish where they vanish!), but in general equality cannot hold as for instance the equalities $Z(\mathfrak{a}^r) = Z(\mathfrak{a})$ show.

1.10 For a subset $X \subset \mathbb{A}^n$, we always have an inclusion $X \subseteq Z(I(X))$ (functions that vanish on X must vanish on X). However, the inclusion may well be strict: consider for instance, the case when $X \subset \mathbb{A}^1$ is an infinite subset, then clearly $I(X) = 0$ (non-zero polynomials have only finitely many zeros), so $Z(I(X)) = \mathbb{A}^1$. In fact, $Z(I(X))$ equals the *Zariski closure* of X:

❏ $Z(I(X)) = \overline{X}$

Indeed, since $X \subseteq Z(I(X))$ and $Z(I(X))$ is a closed subset, we get the $\supseteq$-inclusion. On the other hand, let $Z(\mathfrak{b})$ be a closed subset that contains X. Then $f(x) = 0$ for all $x \in X$ and $f \in \mathfrak{b}$. Hence $\mathfrak{b} \subseteq I(X)$ and so $Z(\mathfrak{b}) \supseteq Z(I(X))$, since $Z(-)$ reverses inclusions.

1.2 Hilbert's Nullstellensatz

1.11 For closed algebraic subset X, we just saw that $Z(I(X)) = X$. Hilbert's Nullstellensatz is about the composition of I and Z the other way around, namely about $I(Z(\mathfrak{a}))$. Polynomials in the radical $\sqrt{\mathfrak{a}}$ vanish along $Z(\mathfrak{a})$ and therefore $\sqrt{\mathfrak{a}} \subseteq I(Z(\mathfrak{a}))$, and the Nullstellensatz tells us that this inclusion is an equality:

Theorem 1.12 (Hilbert's Nullstellensatz) *Assume that k is an algebraically closed field, and that $\mathfrak{a}$ is an ideal in $k[x_1, \ldots, x_n]$. Then one has*

$$I(Z(\mathfrak{a})) = \sqrt{\mathfrak{a}}.$$

Note that the theorem fails immediately if the ground field is not algebraically closed: The simplest example of an ideal in a polynomial ring with empty zero locus is the ideal $(x^2 + 1)$ in $\mathbb{R}[x]$.

1.13 Note that $I(\emptyset)$ equals the entire polynomial ring (requesting that a polynomial f vanishes in all points in $\emptyset$ is an empty condition). Conversely, if $\mathfrak{a}$ is a proper ideal, then it is clear that $\sqrt{\mathfrak{a}}$ is not the entire polynomial ring, so in particular, the theorem asserts that $Z(\mathfrak{a}) = \emptyset$ if and only if $\mathfrak{a}$ equals the whole polynomial ring; that is, if and only if $1 \in \mathfrak{a}$. Hence we can conclude that $Z(\mathfrak{a})$ is not empty when $\mathfrak{a}$ is proper. This statement goes under the name of the Weak Nullstellensatz and is, despite the name, equivalent to the Nullstellensatz, as we shall see later on.

Theorem 1.14 (Weak Nullstellensatz) *Assume that k is an algebraically closed field. For every proper ideal $\mathfrak{a}$ in $k[x_1, \ldots, x_n]$, there is a point $x \in Z(\mathfrak{a})$.*

1.15 Consider now the ideals $(x_1 - a_1, \ldots, x_n - a_n)$ where the a_i's are elements from k. It is easy to see that this is a maximal ideal, since it is the kernel of the map $k[x_1, \ldots, x_n] \to k$ that evaluates a polynomial at the point $(a_1, \ldots, a_n)$.

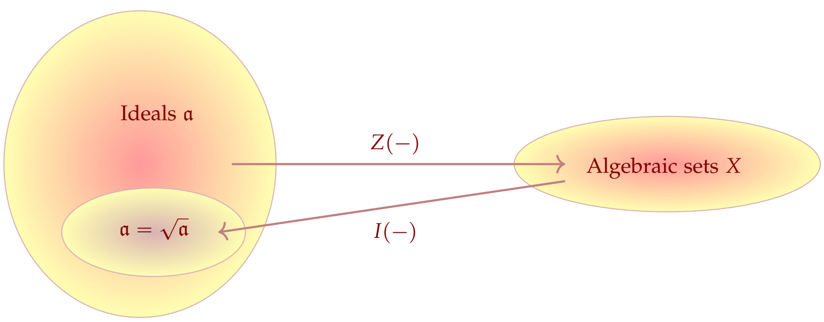
Amazingly, the converse also follows from the Nullstellensatz: Every maximal ideal in the polynomial ring is of this form. If $\mathfrak{m}$ is a maximal ideal, it is certainly a proper ideal, and by the (Weak) Nullstellensatz there is point $(a_1, \ldots, a_n)$ in $Z(\mathfrak{m})$. Consequently, it holds that $(x_1 - a_1, \ldots, x_n - a_n) \subseteq \mathfrak{m}$, but since $(x_1 - a_1, \ldots, x_n - a_n)$ is also maximal, the two ideals must coincide. Hence we have the following equivalent version of the Weak Nullstellensatz:

Theorem 1.16 (Weak Nullstellensatz II) *Assume the field k is algebraically closed. Then the maximal ideals in the polynomial ring $k[x_1, \dots, x_n]$ are exactly those of the form $(x_1 - a_1, \dots, x_n - a_n)$ with $(a_1, \dots, a_n) \in \mathbb{A}^n$.*

Proof: We already argued that the first version of the Weak Nullstellensatz implies the second. The first follows from second because any proper ideal $\mathfrak{a}$ is contained in a maximal ideal, say $\mathfrak{m}$, and the second two tells us that $\mathfrak{m} = (x_1 - a_1, \dots, x_n - a_n)$. As $\mathfrak{a} \subseteq \mathfrak{m}$, we may conclude that $(a_1, \dots, a_n) \in Z(\mathfrak{a})$. ❑

1.17 Recall that an ideal $\mathfrak{a}$ is a *radical ideal* if $\mathfrak{a} = \sqrt{\mathfrak{a}}$. In light of this, one may formulate the Nullstellensatz by saying that $X \mapsto I(X)$ and $I \mapsto Z(I)$ are mutually inverse mappings between the set of radical ideals and the set of closed algebraic sets. Both of these sets are partially ordered under inclusion, and both the mappings I and Z reverse the partial orders.

Radical ideals
radikale idealer



From the identity $\sqrt{\mathfrak{a} \cap \mathfrak{b}} = \sqrt{\mathfrak{a}} \cap \sqrt{\mathfrak{b}}$ it follows that if $\mathfrak{a}$ and $\mathfrak{b}$ are two radical ideals, their intersection $\mathfrak{a} \cap \mathfrak{b}$ is as well. On the other hand, the sum $\mathfrak{a} + \mathfrak{b}$ of two radical ideals is not in general radical. For instance, the ideals $(y - x^2)$ and (y) are both radical, but $(y - x^2) + (y) = (y - x^2, y) = (y, x^2)$ is not.

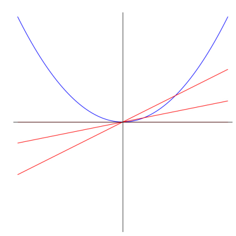


Figure 1.2: The parabola $y = x^2$ and some lines through the origin.

Exercise 1.1 Show that $(y - x^2)$ is prime, and hence radical. Let $\alpha \in k$ and let $\mathfrak{a} = (y - x^2, y - \alpha x)$. Show that $\mathfrak{a}$ is a radical ideal when $\alpha \neq 0$, but not when $\alpha = 0$. ★

The coordinate ring of a closed algebraic set

Definition *Assume that $X \subseteq \mathbb{A}^n$ is a closed algebraic subset. The quotient ring*

$$A(X) = k[x_1, \dots, x_n]/I(X)$$

is called the affine coordinate ring *of X.*

Two polynomials f and g in the variables $x_1, \dots, x_n$ restrict to the same function on X precisely when their difference $f - g$ belongs to the ideal $I(X)$.

Hence it is natural to interpret elements in $A(X)$ as being polynomial functions $X \to k$; that is, k-valued functions on X that are restrictions of a polynomials.

1.19 If Y is a closed algebraic set contained in X, it holds that $I(X) \subseteq I(Y)$, and conversely if $I(Y)$ contains $I(X)$, one has $Y \subseteq X$. Hence there is a one-to-one correspondence between radical ideals in the coordinate ring $A(X)$ and closed algebraic subsets contained in X. If $\mathfrak{a}$ is an ideal in $A(X)$, we denote by $Z(\mathfrak{a})$ the corresponding closed subset of X. And for a point $a = (a_1, \dots, a_n) \in X$ we let $\mathfrak{m}_a$ denote the image in $A(X)$ of the maximal ideal $(x_1 - a_1, \dots, x_n - a_n)$ of polynomials vanishing at x.

1.3 *Hilbert's Nullstellensatz – proofs*

In this section we discuss various proofs and various versions of the Nullstellensatz. The Nullstellensatz comes basically in two flavours, the strong Nullstellensatz and the weak one (of which we shall present three variations). Despite their names the different versions are equivalent. The strong version trivially implies the weak, but the reverse implication hinges on a trick frequently called the *Rabinowitsch trick*.

The Rabinowitsch trick

1.20 We proceed to present the trick proving that the weak version of the Nullstellensatz (Theorem 1.14 on page 12) implies the strong due to J.L. Rabinowitsch.

We need to show that $I(Z(\mathfrak{a})) \subseteq \sqrt{\mathfrak{a}}$ for any proper ideal $\mathfrak{a}$ in $k[x_1, \dots, x_n]$. The crux of the trick is to introduce a new auxiliary variable x_{n+1} and for each $g \in I(Z(\mathfrak{a}))$ to consider the ideal $\mathfrak{b}$ in the polynomial ring $k[x_1, \dots, x_{n+1}]$ given by

$$\mathfrak{b} = \mathfrak{a} \cdot k[x_1, \dots, x_{n+1}] + (1 - x_{n+1} \cdot g).$$

In geometric terms the zero-locus $Z(\mathfrak{b}) \subseteq \mathbb{A}^{n+1}$ is the intersection of the the subset $Z = Z(1 - x_{n+1} \cdot g)$ and the inverse image $\pi^{-1}Z(\mathfrak{a})$ of $Z(\mathfrak{a})$ under the projection $\pi \colon \mathbb{A}^{n+1} \to \mathbb{A}^n$ that forgets the auxiliary coordinate x_{n+1}. This intersection is empty since obviously g does not vanish along Z, but vanishes identically on $\pi^{-1}Z(\mathfrak{a})$.

The Weak Nullstellensatz therefore gives that $1 \in \mathfrak{b}$, and hence there are polynomials f_i in $\mathfrak{a}$ and h_i and h in $k[x_1, \dots, x_{n+1}]$ satisfying a relation of the form

$$1 = \sum f_i(x_1, \dots, x_n) h_i(x_1, \dots, x_{n+1}) + h \cdot (1 - x_{n+1} \cdot g).$$

We substitute $x_{n+1} = 1/g$ and multiply through by a sufficiently[2] high power g^N of g to obtain

Affine coordinate rings
affine koordinatringer

$$g^N = \sum f(x_1, \dots, x_n) H_i(x_1, \dots, x_n),$$

where $H_i(x_1,\ldots,x_n) = g^N \cdot h_i(x_1,\ldots,x_n,g^{-1})$. Hence $g \in \sqrt{\mathfrak{a}}$.

The third version of the Weak Nullstellensatz

1.21 As already mentioned there are several variants of the weak Nullstellensatz. We have already seen two, and here comes number three. This is the one we shall prove and from which we subsequently shall deduce the other versions. It has the virtue of being general, in that it is valid over any field k, not necessarily algebraically closed, and it the version well adapted to Grothendieck's marvellous world of schemes.

THEOREM 1.22 (WEAK NULLSTELLENSATZ III) *Let k a field and let $\mathfrak{m}$ be a maximal ideal in the polynomial ring $k[x_1,\ldots,x_n]$. Then $k[x_1,\ldots,x_n]/\mathfrak{m}$ is a finite field extension of k.*

Before proceeding to the proof of this we show how the Weak Nullstellensatz II (Theorem 1.14 on page 12) can be deduced from version III above.

PROOF OF II FROM III: Assume that k is algebraically closed and let $\mathfrak{m}$ be a maximal ideal in $k[x_1,\ldots,x_n]$. The salient point is that the field $k[x_1,\ldots,x_n]/\mathfrak{m}$ is a finite extension of k by version III above, and since k is algebraically closed by assumption, the two fields coincide. Thus there is an algebra homomorphism $k[x_1,\ldots,x_n] \to k$ having $\mathfrak{m}$ as kernel. Letting a_i be the image of x_i under this map, the ideal $(x_1 - a_1,\ldots,x_n - a_n)$ will be contained in $\mathfrak{m}$, and being maximal, it must be equal $\mathfrak{m}$. ❑

Proof of version III of the Nullstellensatz

1.23 The by far simplest proof of the Nullstellensatz we know, both technically and conceptually relies on no more sophisticated mathematics than the fact the polynomial ring $k[x]$ in one variable over a field is a PID. One establishes the following lemma, which obviously implies version III of the Weak Nullstellensatz:

LEMMA 1.24 *If $k \subseteq K$ is a finitely generated extension of fields which is not finite, and $a_1,\ldots,a_r$ are elements in K, then $k[a_1,\ldots,a_r]$ is not equal to K.*

PROOF: To begin with we treat the case that K is of transcendence degree one over k. Then there is a subfield $k(x) \subseteq K$ with x transcendental over which K is finite. Let $\{e_i\}$ be a basis for K over $k(x)$ with $e_0 = 1$, and let c_{ijk} be elements in $k(x)$ such that $e_i e_j = \sum_k c_{ijk} e_k$. Let s be a common denominator of the c_{ijk}. Then $A = \bigoplus_i k[x]_s e_i$ is a subalgebra of K which is free as a module over $k[x]_s$. Now, let $a_1,\ldots,a_r$ be elements in K, and express them in the basis $\{e_i\}$; that is, write $a_j = \sum d_{ij} e_i$ with $d_{ij} \in k(x)$. Let t be the common denominator of the d_{ij}'s.

[2] For instance the highest power of x_{n+1} that occurs in any of the h_i's.

Recall that $k[x]_s$ denotes the localization of $k[x]$ in the multiplicative set $\{1, s, s^2, \ldots\}$. Elements are of the form a/s^r with $a \in k[x]$.

Then $k[a_1, \ldots, a_r]$ is contained in A_t, and therefore can not be equal to K. Indeed, if $u \in k[x]$ is any irreducible element[3]that neither is a factor in s nor in t, then u^{-1} will not lie in A_t.

Finally, if the transcendence degree of K is more than one, we let $k' \subseteq K$ be a field containing k over which K is of transcendence degree 1. Then K is never equal to $k'[a_1, \ldots, a_r]$, hence *a fortiori* neither to $k[a_1, \ldots, a_r]$. ❑

1.4 Examples

Figures and intuition

To have some geometric intuition one frequently has real pictures of algebraic sets in mind. In this case, the ground field must be $\mathbb{C}$ and the algebraic set must be defined by real equations. The object depicted is the subset of the points in $Z(\mathfrak{a})$ whose coordinates are real numbers.

These real pictures can be very instructive (and beautiful) and sometimes they are unsurpassed to explain what happens. But they can as well be deceptive and must be taken with a rather large grain of salt – often they do not tell the whole story, and sometimes they do not say any thing at all. For instance, $x^2 + y^2 + 1$ has no real zeros, so $V(x^2 + y^2 + 1)$ has no real points, but of course there are infinitely many complex zeros.

[3] Even if k is a finite field, there are infinitely many irreducible polynomials in $k[x]$, see problem 1.12 on page 19.

1.25 Performing complex coordinate changes, which is perfectly legitimate when working over $\mathbb{C}$ and does not alter the complex geometry, can completely change the real picture. For instance, replacing y by iy in the above example, which is a simple scaling of one of the coordinates, gives the equation $x^2 - y^2 = -1$ whose real points constitute a hyperbola, and scaling both x and y by i gives the circle $x^2 + y^2 = 1$. So the real picture depends heavily on the complex coordinates one uses.

There is also a change in dimension. The affine plane $\mathbb{A}^2(\mathbb{C})$ is the 2-dimensional over $\mathbb{C}$, hence 4-dimensional when viewed over $\mathbb{R}$. Thus a line in $\mathbb{A}^2(\mathbb{C})$ is a linear subspace of real codimension two; that is, an $\mathbb{R}^2$ in $\mathbb{R}^4$. Complex algebraic sets will be of even (real) dimension where as the (real) dimension of their real counterparts will be half that dimension.

EXAMPLE 1.26 Consider the curve $y^2 = x(x+a)(x-b)$ in $\mathbb{A}^3(\mathbb{C})$; with a and b both positive real numbers. The real points is depicted in Figure 1.4. When this is regarded in the usual topology, there are two components: One compact, which is homeomorphic to a circle, and one which is unbounded. The complex points turn out to form a space homeomorphic to a torus $\mathbb{S}^1 \times \mathbb{S}^1$ minus one point.

Figure 1.3: The famous surface of degree six constructed by Wolf Barth. It has 65 double points. The picture is of a 3D-print of the surface from http://math-sculpture.com.

To underline the extent the real picture depends on the chosen coordinate system, in figure 1.5 we have depicted the cubic curve viewed in another coordinate system.

☆

EXERCISE 1.2 Contrary to what is true for quadratic curves, show that a cubic curve in $\mathbb{A}^2(\mathbb{C})$ defined by an equation with real coefficients always has real points. Generalize to curves with real equations of odd degree. HINT: Intersect with real lines. ★

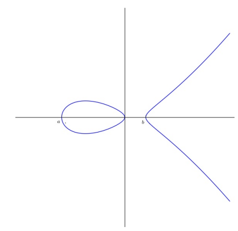


Figure 1.4: The real points of a cubic curve in the so-called Weierstrass normal form.

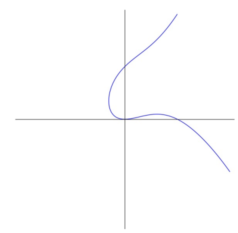


Figure 1.5: The real points of a cubic curve in the so-called Tate normal form.

EXAMPLE 1.27 *(Conics)* Curves in the affine plane $\mathbb{A}^2$ given by irreducible quadratic equations are called *conics* and they have a well-known classification. Up to an affine change of coordinates there are only two types:

- $y = x^2$ (parabola)
- $xy = 1$ (hyperbola)

To see why, note that any quadratic polynomial can be written as $Q(x, y) + L(x, y) + c$ where Q and L are homogeneous polynomials of degrees 2 and 1 respectively, and $c \in k$. Since k is algebraically closed, the quadratic form $Q = A_0x^2 + A_1xy + A_2y^2$ can be factored as the product of two linear forms. We change coordinates so that the two factors become x and y if they are different, or both become x if they coincide. In the former case the quadratic form Q turns into $Q(x, y) = xy$ and in the latter it becomes $Q(x, y) = x^2$. This brings the original quadratic polynomial on form

$$xy + ax + by + c = (x + a)(y + b) + c - ab$$

if $Q(x, y) = xy$, or

$$x^2 + ax + by + c$$

when $Q(x, y) = y^2$. The final necessary coordinate changes are then easy to find and left as an exercise. ☆

EXAMPLE 1.28 *(The affine twisted cubic)* In this example we take a closer look at a famous curve called the *twisted cubic*, or rather an *affine* version of it (there is also a *projective* version which we will come back to later). The word twisted in the name comes from the fact that the space curve is not contained in any plane.

The twisted cubic $C \subset \mathbb{A}^3$ is the image of the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^3$ given as $\phi(t) = (t, t^2, t^3)$. It is a closed algebraic set; indeed, we shall see that $C = Z(\mathfrak{a})$ where $\mathfrak{a}$ is the ideal

$$\mathfrak{a} = (z - x^3, y - x^2).$$

The inclusion $C \subseteq Z(\mathfrak{a})$ follows readily, and for the other inclusion we just observe that points in $Z(\mathfrak{a})$ are shaped like (x, x^2, x^3) so we can just take $t = x$. Moreover, it holds true that $I(X) = \mathfrak{a}$. To see this, we notice that every polynomial f can be represented as

$$f(x, y, z) = f(x, x^2, x^3) + h(x, y, z),$$

where $h \in \mathfrak{a}$. This is just a repeated application of the little Lemma 1.29 below; first with $y = x^2 - (x^2 - y)$ and then with $z = x^3 - (x^3 - z)$. That $f(x,y,z)$ vanishes on C means that $f(x, x^2, x^3)$ vanishes identically and hence $f \in \mathfrak{a}$.

As a by product of this reasoning, we obtain that the ideal $\mathfrak{a}$ is a prime ideal; indeed, it is the kernel of the restriction map

$$k[x,y,z] \to k[t]$$

that sends a polynomial to its restriction to C; in other words, x goes to t, y to t^2 and z to t^3. ☆

The following easy lemma is nothing but Taylor expansion to the first order, but is now and then useful:

Lemma 1.29 *Assume that R is any commutative ring. Let $P(z)$ be a polynomial in $R[z]$. Then $P(z+w) = P(z) + wQ(z,w)$ for some polynomial Q in $R[z,w]$.*

Proof: Observe that by the binomial theorem one has $(z+w)^i = z^i + wQ_i(z,w)$; the rest of the proof follows from this. ❑

Exercises

1.3 Let $f \in k[x_1, \ldots, x_n]$. Show that the ideal (f) is radical if and only if no factor of f is multiple.

Conics
kjeglesnitt

1.4 Assume that the characteristic of k is zero. Let $f(x)$ be a polynomial in $k[x]$. Show that the relation $\sqrt{(f)} = (f : f')$ holds (where f' is the derivative of f; see exercise 1.13). Give a counterexample if k is of positive characteristic.

1.5 Let $\mathfrak{p}$ be a prime ideal in $k[x_1, \ldots, x_n]$. Show that $\mathfrak{p}$ is the intersection of all the maximal ideals containing it; that is, $\mathfrak{p} = \bigcap_{\mathfrak{p} \subseteq \mathfrak{m}} \mathfrak{m}$. Hint: Show that $I(Z(\mathfrak{p})) = \bigcap_{\mathfrak{p} \subseteq \mathfrak{m}} \mathfrak{m}$, then use the Nullstellensatz.

1.6 Consider the closed algebraic set in $\mathbb{A}^2$ given by the vanishing of the polynomial $P(x) = y^2 - x(x+1)(x-1)$. Let $\alpha \in \mathbb{C}$ and let $\mathfrak{a} = (x - \alpha, P(x))$. Determine $Z(\mathfrak{a})$ for all α. For which α's is $\mathfrak{a}$ a radical ideal?

1.7 With the same notation as in the previous problem. Let $\mathfrak{b}$ be the ideal $\mathfrak{b} = (y - \alpha, P(x))$. Determine $Z(\mathfrak{b})$ for all α and decide for which α the ideal $\mathfrak{b}$ is radical. Hint: The answer depends on the characteristic of k, characteristic three being special.

1.8 Let $F_1, \ldots, F_r$ be homogenous polynomials in $k[x_1, \ldots, x_n]$ and let $X = Z(F_1, \ldots, F_r)$ be the closed algebraic subset they define. Show that X is a *cone* with apex at the origin; that is, show that if x is a point in X, the line joining x to the origin lies entirely in X. Hint: Show that $t \cdot x$ lies in X for all $t \in k$.

1.9 Assume that X is a cone in $\mathbb{A}^n$ with apex at the origin, and assume that f is a polynomial that vanishes on X. Show that also all the homogenous components of f vanish along X.

1.10 Let $\mathbb{M}_{n,m}$ be the space of $n \times m$-matrices with coefficients from k. It can be identified with the affine space $\mathbb{A}^{nm}$ with coordinates x_{ij} where $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$. Let r be a natural number less than both n and m, and let W_r be the set of $n \times m$-matrices of rank at most r. Show that W_r is a closed algebraic subset. Show that all the W_r's are cones over the origin. HINT: Determinants are polynomials.

For any two ideals $\mathfrak{a}$ and $\mathfrak{b}$ in a ring A recall that one denotes by $(\mathfrak{a} : \mathfrak{b})$ the ideal of those $a \in A$ such that $a \cdot \mathfrak{b} \subseteq \mathfrak{a}$; that is $(\mathfrak{a} : \mathfrak{b}) = \{ a \in A \mid a \cdot \mathfrak{b} \subseteq \mathfrak{a} \}$.

1.11 Let for every natural number $d \geqslant 2$ let $C_d \subseteq \mathbb{A}^d$ be the curve defined by the parametric representation $\phi(t) = (t, t^2, \dots, t^n)$, and let $\mathfrak{a}$ be the ideal $\mathfrak{a} = (x_i - x_1 x_{i-1} \mid 2 \leqslant i \leqslant d)$. Show that C_d is a closed algebraic set, that $I(C_d) = \mathfrak{a}$ and that $\mathfrak{a}$ is a prime ideal. The curves C_d are called *affine normal rational curves* and they are close relatives to the twisted cubic. For $d = 2$ we have a parabola in the plane and for $d = 3$ we get back the twisted cubic.

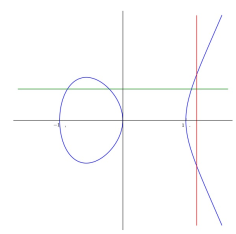


Figure 1.6: A cubic and two lines.

1.12 As usual $\mathbb{F}_q$ denotes the finite field with q elements. The aim of this exercise is to establish that there are infinitely many irreducible polynomials with coefficients in $\mathbb{F}_q$.

The simplest proof is a reuse of the good old argument of Euclid that there are infintely many prime numbers. Assume that $p_1, \dots, p_r$ are the only irreducible polynomials in $\mathbb{F}_q[t]$. Show that $1 + p_1 \cdot \ldots \cdot p_r$ is not a unit and is not divisible by any of the p_i's.

1.13 *(The formal derivative.)* Let $f(x) = \sum_i a_i x^i$ be a polynomial. Define the (formal) derivative of f to be $f'(x) = \sum_i i a_i x^{i-1}$. Show that the usual rules are still valid; *i.e.* derivation is a linear operation and Leibnitz's product rule holds true. Show that f' vanishes identically if and only if either f is constant or the characteristic of k is p and $f(x) = g(x^p)$ for some polynomial $g(x)$.

★

Chapter 2

The Zariski topology

TOPICS IN CHAPTER 2: *The Zariski topology on closed algebraic subsets; irreducible topological spaces; Noetherian topological spaces; primary decomposition and decomposition of noetherian spaces into irreducibles; hypersurfaces; polynomial maps; quadratic forms; determinantal varieties; Veronese surface.*

It is sometimes said that "Algebraic geometry is the geometry of zero sets of polynomials". This is a good first approximation, but the realm of algebraic geometry is much bigger than the corner occupied by the closed algebraic sets. There are many more geometric objects, several of which will be the principal objects of our interest. However, the closed algebraic sets are fundamental and serve as building blocks. Just as a smooth manifold locally looks like an open ball in Euclidean space, our spaces will locally look like a closed algebraic set, or in a more restrictive setting, like an affine variety. Before giving the general definition, we need to know what the term "locally" means, and of course, this will be encoded in a topology. The topologies that are used, are particularly well adapted to algebraic geometry, and they are called *Zariski topologies* after one of the great algebraic geometers Oscar Zariski.

Oscar Zariski (1899–1986) Russian–American mathematician

2.1 *The Zariski topology*

2.1 In Chapter 1, we established the intimate relationship between closed algebraic subsets of affine spaces and ideals in polynomial rings, and among those relations were the following two:

- $Z(\sum_{i\in I} \mathfrak{a}_i) = \bigcap_{i\in I} Z(\mathfrak{a}_i)$,
- $Z(\mathfrak{a} \cap \mathfrak{b}) = Z(\mathfrak{a}) \cup Z(\mathfrak{b})$,

where $\mathfrak{a}$ and $\mathfrak{b}$ are ideals and $\{\mathfrak{a}_i\}_{i\in I}$ any collection of ideals in a polynomial ring $k[x_1,\ldots,x_n]$. The first relation shows that the intersection of arbitrarily many closed algebraic sets is a closed algebraic set, and the second that the union of two is closed algebraic (hence the union of finitely many will be as well). And of course, both the empty set and the entire affine space are closed algebraic sets (zero loci of respectively the whole polynomial ring and the zero ideal). The closed algebraic sets in $\mathbb{A}^n$ therefore fulfill the axioms for being the closed sets of a topology. This is the topology we call the *Zariski topology.*

The Zariski topology
Zariski-topologien

2.2 Every closed algebraic set X in $\mathbb{A}^n$ carries a Zariski topology as well, namely the topology induced from the Zariski topology on $\mathbb{A}^n$. The closed sets are easily seen to be the closed subsets of $\mathbb{A}^n$ that are contained in X. These are the zero loci of ideals $\mathfrak{a}$ containing $I(X)$; that is, those loci shaped like $Z(\mathfrak{a})$ with $I(X) \subseteq \mathfrak{a}$. Such ideals are in a one-to-one correspondence with the ideals in the coordinate ring $A(X) = k[x_1, \ldots, x_n]/I(X)$. In other words, the Zariski closed sets in X are the zero loci of the ideals in $A(X)$. And if we request the $\mathfrak{a}$'s to be radical ideals, the correspondence is one-to-one (remember that $Z(\sqrt{\mathfrak{a}}) = Z(\mathfrak{a})$ by Proposition 1.7)

Proposition 2.3 *Let $X \subset \mathbb{A}^n$ be an algebraic set. There is a one-to-one correspondence between closed subsets $Y \subset X$ and radical ideals $\mathfrak{a} \subseteq A(X)$.*

Example 2.4 We showed earlier that the closed algebraic sets of $\mathbb{A}^1$ are, apart from the empty set and $\mathbb{A}^1$ itself, just the finite sets. Thus the Zariski topology on $\mathbb{A}^1$ coincides with the *cofinite topology*; the open sets are exactly those with a finite complement. ☆

Example 2.5 The closed sets in the affine plane $\mathbb{A}^2$ are more complicated. Later on we will show that they are finite unions of either points or subsets of the form[1] like $Z(f(x,y))$ where f is a polynomial in $k[x,y]$. While $\mathbb{A}^2 = \mathbb{A}^1 \times \mathbb{A}^1$, we note that the Zariski topology on $\mathbb{A}^2$ is not the same as the product topology. The product topology is generated by the inverse images of closed sets in the two factors which in our case are just points (apart from the empty set and the entire space), and the inverse images are thus sets of the form $Z(x-a)$ or $Z(y-a)$. The closed sets are thus finite union of intersections of these; that is, finite unions of points or lines parallel to one of the coordinate axes. However, a conic like the hyperbola $xy = 1$, for instance, is not among those, but it is Zariski closed. ☆

[1] Or see *e.g.* Proposition **??** in CA

2.6 Among the open sets there are some called *distinguished open sets* that play a special role. They are the sets where a single polynomial does not vanish. If f is any polynomial in $k[x_1, \ldots, x_n]$, we define

The distinguished open sets
prinsipale åpne mengder

$$D(f) = \{\, x \in X \mid f(x) \neq 0 \,\},$$

which clearly is open in X being the complement of $Z(f) \cap X$. Another common notation for $D(f)$ is X_f.

Proposition 2.7 *Let X be a closed algebraic set. The distinguished open sets form a basis for the Zariski topology on X.*

2

[2] Recall that a collection $\{U_i\}$ of open sets is a basis for the topology if every open set in X is the union of members of $\{U_i\}$.

PROOF: Fix an open set U. The complement U^c is closed and hence of the form $U^c = Z(\mathfrak{a})$ for some ideal $\mathfrak{a}$ in $A(X)$. If $\{f_i\}$ is a set of generators for $\mathfrak{a}$, it holds true that $Z(\mathfrak{a}) = \bigcap_i Z(f_i)$, and consequently $U = Z(\mathfrak{a})^c = \bigcup_i D(f_i)$. ❑

2.8 When the ground field is the field of complex numbers $\mathbb{C}$, the affine space $\mathbb{A}^n(\mathbb{C}) = \mathbb{C}^n$ has in addition to the Zariski topology the traditional Euclidean topology, and a closed algebraic set X in $\mathbb{A}^n$ inherits a topology from this. This induced topology on X is called *the Euclidean topology*.

the Euclidean topology
den Euclidske topologien

The Zariski topology is very different from the Euclidean topology. Polynomials are continuous in the Euclidean topology, so any Zariski open set is open in the usual topology, but the converse is far from being true. For example, in contrast to the usual topology on $\mathbb{C}$, the Zariski topology on the affine line $\mathbb{A}^1(\mathbb{C})$, as we saw in Example 2.4 above, is the cofinite topology; a non-empty set is open if and only if the complement is finite.

The Zariski topology has, however, the virtue of being defined whatever the ground field is (as long as it is algebraically closed, if one wants to stay with varieties[3]), and the field can very well be of positive characteristic.

[3] So-called *schemes* are very general geometric gadgets that also have a Zariski topology

2.2 *Irreducible topological spaces*

DEFINITION *A topological space X is called* irreducible *if it is not the union of two proper closed subsets. That is, if $X = X_1 \cup X_2$ with X_1 and X_2 both being closed, then either $X_1 = X$ or $X_2 = X$.*

Taking complements, we arrive at the following characterization of irreducible spaces:

LEMMA 2.10 *A topological space X is irreducible if and only if the intersection of any two non-empty open subsets is non-empty.*

PROOF: Assume first that X is irreducible and let U_1 and U_2 be two open subsets. If $U_1 \cap U_2 = \emptyset$, it would follow, when taking complements, that $X = U_1^c \cup U_2^c$, and X being irreducible, we could infer that $U_i^c = X$ for either $i = 1$ or $i = 2$; whence $U_i = \emptyset$ for one of the i's. To prove the other implication, assume that X is expressed as a union $X = X_1 \cup X_2$ with the X_i's being closed. Then $X_1^c \cap X_2^c = \emptyset$; hence either $X_1^c = \emptyset$ or $X_2^c = \emptyset$, and therefore either $X_1 = X$ or $X_2 = X$. ❑

2.11 There are a few properties irreducible spaces have that follow immediately from Lemma 2.10. Firstly, every open non-empty subset U of an irreducible space X is dense. Indeed, if $x \in X$ and V is any neighbourhood of x, the lemma tells us that $U \cap V \neq \emptyset$, and x therefore belongs to the closure of U.

Secondly, every non-empty open subset U of X is irreducible. This follows trivially since any two non-empty open sets of U are open in X, hence their intersection is *a fortiori* non-empty.

Thirdly, the closure $\overline{Y}$ of an irreducible subset Y of X is irreducible. For if U_1 and U_2 are two non-empty open subsets of $\overline{Y}$, it holds true that $U_i \cap Y \neq \emptyset$, and hence $U_1 \cap U_2 \cap Y \neq \emptyset$ since Y is irreducible, and *a fortiori* the intersection $U_1 \cap U_2$ is non-empty.

Fourthly, continuous images of irreducible spaces are irreducible. If $f\colon X \to Y$ is surjective and continuous and U_i for $i = 1, 2$ are open and non-empty subsets of Y, it follows that each $f^{-1}(U_i)$ is open and non-empty (the map f is surjective). When X is irreducible, it holds that $f^{-1}(U_1 \cap U_2) = f^{-1}(U_1) \cap f^{-1}(U_2) \neq \emptyset$, and so $U_1 \cap U_2$ is not empty.

We summarize this in the following lemma:

LEMMA 2.12 *Open non-empty sets of an irreducible set are irreducible and dense. Closures and continuous images of irreducible sets are irreducible.*

We should also mention that Zariski topologies are far from being Hausdorff; it is futile to search for disjoint neighbourhoods when all non-empty open subsets meet!

2.13 Closed algebraic sets in the affine space $\mathbb{A}^n$ are of special interest:

PROPOSITION 2.14 *An algebraic set $X \subseteq \mathbb{A}^n$ is irreducible if and only if the ideal $I(X)$ of polynomials vanishing on X is prime.*

In particular, we observe that the affine space $\mathbb{A}^n$ itself is irreducible.

PROOF: Assume that X is irreducible and let f and g be two polynomials such that $fg \in I(X)$, which implies that $X \subseteq Z(f) \cup Z(g)$. Since X is irreducible, it follows that either $Z(g) \cap X$ or $Z(f) \cap X$ equals X. Hence one has either $X \subseteq Z(f)$ or $X \subseteq Z(g)$, which for the ideal $I(X)$ means that either $f \in I(X)$ or $g \in I(X)$.

Conversely, assume that $I(X)$ is prime and that $X = Z(\mathfrak{a}) \cup Z(\mathfrak{b})$ with $\mathfrak{a}$ and $\mathfrak{b}$ being radical ideals. Then it holds that $I(X) = \mathfrak{a} \cap \mathfrak{b}$, and because $I(X)$ is prime, we deduce that $I(X) = \mathfrak{a}$ or $I(X) = \mathfrak{b}$. Hence $X = Z(\mathfrak{a})$ or $X = Z(\mathfrak{b})$. ❏

From the Nullstellensatz, we obtain the following dictionary between algebra and geometry:

PROPOSITION 2.15 *Let $A = k[x_1, \ldots, x_n]$ with k algebraically closed. Then the mappings $X \mapsto I(X)$ and $I \mapsto Z(I)$ give a one-to-one inclusion reversing correspondence between the objects in the left and right-hand columns in the following table:*

Algebra	*Geometry*
maximal ideals of A	*points of $\mathbb{A}^n$*
prime ideals of A	*irreducible closed subsets of $\mathbb{A}^n$*
radical ideals of A	*closed subsets of $\mathbb{A}^n$*
maximal ideals of $A/\mathfrak{a}$	*points of $Z(\mathfrak{a})$*

Hypersurfaces

2.16 Algebraic sets that are the zero locus of one single polynomial; that is, sets X such that $X = Z(f)$, are called *hypersurfaces*. They are quintessential players in our story. Curves in $\mathbb{A}^2$ and surfaces in $\mathbb{A}^3$ are well known examples of the sort.

Irreducible topological spaces
irreduktible topologiske rom

In general, hypersurfaces are somehow more manageable than general algebraic sets –even though the equation can be complicated, at least there is just one.

The simplest hypersurface is the one defined by a linear polynomial, in which case $Z(f)$ is called a *hyperplane*. This is just a linear subspace of dimension $n-1$ in $\mathbb{A}^n$.

Hypersurfaces
hyperflater

2.17 As we have seen, the ideal (f) and its radical $\sqrt{(f)}$ have the same zero locus, so every hypersurface has a polynomial f without multiple factors as defining polynomial. Moreover, we know that a polynomial f generates a prime ideal if and only it is irreducible –this is just the fact that polynomial rings are UFD's. Hence a hypersurface is irreducible if and only if it can be defined by an irreducible polynomial.

PROPOSITION 2.18 *Let f be a polynomial in $k[x_1, \ldots, x_n]$. If f is an irreducible polynomial, the hypersurface $Z(f)$ is irreducible. If the hypersurface $Z(f)$ is irreducible, then f is a power of some irreducible polynomial.*

The principal prime ideals in $k[x_1, \ldots, x_n]$ are height one prime ideals; that is, they are minimal among the non-zero primes, and conversely, since the polynomial ring is a UFD, all height one primes are principal. The Nullstellensatz then entails that the irreducible hypersurfaces are precisely the maximal proper closed subsets of $\mathbb{A}^n$ –and anticipating the notion of dimension, it is common usage to say they are of codimension one; *i.e.* one less than the ambient space. The hypersurfaces; that is, those of the form $Z(f)$ with f a non-constant polynomial, are precisely those closed subsets that are finite unions of maximal proper closed subsets.

Hyperplanes
hyperplan

Examples

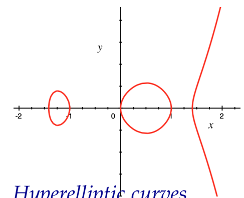
2.19 Polynomials shaped like $f(x) = y^2 - P(x)$ are irreducible unless $P(x)$ is a square; that is $P(x) = Q(x)^2$. Indeed; if $f = A \cdot B$ either both A and B are linear in y or one of them does not depend on y at all (just consider the terms of A and B of highest degree in y). In the former case $A(x,y) = y + a(x)$ and $B(x,y) = y + b(x)$ which gives

$$y^2 - P(x) = (y + a(x)) \cdot (y + b(x)) = y^2 + (a(x) + b(x)) \cdot y + a(x)b(x),$$

and it follows that $a(x) = -b(x)$. In the latter case one finds

$$y^2 - P(x) = (y^2 + a(x)) \cdot b(x),$$

which implies that $b(x) = 1$, and hence $f(x)$ is irreducible. When the polynomial $P(x)$ has merely simple zeros, the curve defined by f is called a *hyperelliptic curve*, and if $P(x)$ in addition is of the third degree, it is said to be an *elliptic curve*. In figures 2.1 and 2.2 in the margin we have depicted (the real points of) two, both with $P(x)$ of the eighth degree. Can you explain the qualitative difference between the two?



Hyperelliptic curves
Elliptic curves
hyperelliptiske kurver
elliptiske kurver
Figure 2.1: A hyperelliptic curve.

We already met some elliptic curves in Lecture 1. They are omnipresent in both geometry and number theory and we shall study them closely later on.

2.20 Our second example is a well known hypersurface, namely the determinant. It is one of many interesting algebraic sets that appear as subsets of the space of matrices $\mathbb{M}_{n,m} = \mathbb{A}^{nm}$ defined by rank conditions. The example is about the determinant $F = \det(x_{ij})$ of a "generic" $n \times n$-matrix $M = (x_{ij})$ with independent variables as entries. It is a homogenous polynomial of degree n. (The determinant is linear in each row, or column, so that $\det(tx_{ij}) = t^n \det(x_{ij})$.)

We shall see that it is irreducible by a so-called *specialization* technique. Consider matrices like

$$A = \begin{pmatrix} t & y_1 & 0 & 0 & \dots & 0 \\ 0 & t & y_2 & 0 & \dots & 0 \\ 0 & & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & t & y_{n-2} & 0 \\ 0 & \dots & 0 & 0 & t & y_{n-1} \\ y_n & \dots & 0 & 0 & 0 & t \end{pmatrix}$$

with a variable t's along the diagonal, and variables y_i's along the first "supra-diagonal" and y_n in the lower left corner; in other words, the specialization consists in putting $x_{ii} = t$, $x_{i,i+1} = y_i$ for $i \leqslant n-1$ and $x_{n1} = y_n$, and the rest of the x_{ij}'s are put to zero.

It is not difficult to show that $\det A = t^n - (-1)^n y_1 \cdot \ldots \cdot y_n$ and that this polynomial is irreducible (Problem 2.1 below). A potential factorization $\det(x_{ij}) = F \cdot G$ with F and G both of degree less than n must persist when we give the variables special values (but notice that one of the factors may become constant). But since $\det A$ is irreducible and of degree n, this cannot happen, and we conclude that there can be no such factorization.

☆

EXERCISE 2.1 With notation as in the example, show that the determinant $\det A$ is given as $\det A = t^n - (-1)^n y_1 \cdot \ldots \cdot y_n$ and that this is an irreducible polynomial. HINT: Consider the terms of highest and lowest degree in t of potential factors. ★

Decomposition into irreducible subsets

From commutative algebra we know that ideals in Noetherian rings have a *primary decomposition*. This is the Noether-Lasker theorem proven by Emanuel Lasker for ideals in polynomial rings in 1905. Some fifteen years later the general result, as we know it today, was established by Emmy Noether. It states that every ideal $\mathfrak{a}$ in a Noetherian ring can be expressed as an intersection

$$\mathfrak{a} = \mathfrak{q}_1 \cap \mathfrak{q}_2 \cap \ldots \cap \mathfrak{q}_r$$

where the $\mathfrak{q}_i$'s are primary ideals[4]. Recall that primary ideals have radicals that are a prime, so the ideals $\sqrt{\mathfrak{q}_i}$ are all prime. They are called the *primes associated* to $\mathfrak{a}$. Such a decomposition is not always unique, but there are partial uniqueness results. The associated prime ideals are unique as are the primary components $\mathfrak{q}_i$ whose associated prime ideals are minimal. However, the so-called *embedded components*[5] are not. For instance, one has $(x^2, xy) = (x) \cap (x^2, y)$ but also $(x^2, xy) = (x) \cap (x^2, xy, y^2)$ holds true.

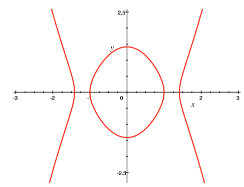


Figure 2.2: Another hyperelliptic curve.

[4] See *e.g.* Theorem **??** in CA

Associated primes
assosierte primidealer

2.21 Properties of ideals in the polynomial ring $k[x_1, \ldots, x_n]$ usually translate into properties of algebraic sets, and the same is true for the primary decomposition. In geometric terms, it means the following. Let $Y = Z(\mathfrak{a})$ for an ideal $\mathfrak{a} \subseteq k[x_1, \ldots, x_n]$, and write down the primary decomposition of $\mathfrak{a}$:

$$\mathfrak{a} = \mathfrak{q}_1 \cap \mathfrak{q}_2 \cap \ldots \cap \mathfrak{q}_r.$$

Putting $Y_i = Z(\sqrt{\mathfrak{q}_i})$, we find $Y = Y_1 \cup Y_2 \cup \ldots \cup Y_r$, where each Y_i is an irreducible closed algebraic set in $\mathbb{A}^n$. If the prime $\sqrt{\mathfrak{q}_i}$ is not minimal among the associated primes, say $\sqrt{\mathfrak{q}_j} \subseteq \sqrt{\mathfrak{q}_i}$, it holds that $Y_i \subseteq Y_j$, and the component Y_i contributes nothing to intersection and can be discarded.

[5] A primary component $\mathfrak{q}_i$ is *embedded* if $\sqrt{\mathfrak{q}_i}$ contains the radical $\sqrt{\mathfrak{q}_j}$ of another component $\mathfrak{q}_j$.

EXAMPLE 2.22 Consider the closed set $X = Z(I) \subset \mathbb{A}^3$ given by the ideal

$$I = (x^2 - y, xz - y^2, x^3 - xz)$$

Note first that if $x = 0$, then $y = 0$, so $Z(x, y) \subset X$. If $x \neq 0$, the third equation gives $z = x^2$, and so by the first and second equations we get $xz - y^2 = x^3 - x^4$, giving $x = 1$, $y = 1$ and $z = 1$. Hence

$$X = Z(x, y) \cap Z(x - 1, y - 1, z - 1)$$

That is, X is the union of the z-axis, and the point $(1, 1, 1)$. In fact, a primary decomposition of I is given by $I = \mathfrak{q}_1 \cap \mathfrak{q}_2 \cap \mathfrak{q}_3$, where

$$\mathfrak{q}_1 = (x, y), \quad \mathfrak{q}_2 = (x - 1, y - 1, z - 1), \quad \mathfrak{q}_3 = (x^2 - y, xy, y^2, z).$$

Taking radicals, we find that the primes associated to I are

$$\mathfrak{p}_1 = (x, y), \quad \mathfrak{p}_2 = (x - 1, y - 1, z - 1), \quad \mathfrak{p}_3 = (x, y, z).$$

Note that $\mathfrak{p}_1 \subset \mathfrak{p}_3$, so $\mathfrak{p}_3$ is an embedded component, so it does not show up in the decomposition above. ☆

Emmy Noether (1882–1935) German mathematician

2.23 In a more general context, a decomposition $Y = Y_1 \cup \ldots \cup Y_r$ of any topological space is said to be *redundant* if one can discard one or more of the Y_i's without changing the union. That a component Y_j can be omitted is equivalent to Y_j being contained in the union of rest; that is, $Y_j \subseteq \bigcup_{i \neq j} Y_i$. A decomposition that is not redundant, is said to be *irredundant*. Translating the Noether–Lasker theorem into geometry we arrive at the following:

Redundant decompositions *redundante dekomposisjoner*

Proposition 2.24 *Any closed algebraic set $Y \subseteq \mathbb{A}^n$ can be written as an irredundant union*

$$Y = Y_1 \cup \ldots \cup Y_r$$

where the Y_i's are irreducible closed algebraic subsets. The union is unique up to the order of the Y_i's.

Notice that since embedded components do not show up for radical ideals, we get a clear and clean uniqueness statement.

Example 2.25 Consider the algebraic set $X = Z(f, g)$ in $\mathbb{A}^3$ where

$$f = x^2 - yz \text{ and } g = xz - x.$$

Then X has three components: From $g = 0$, we find that $x = 0$, or $z = 1$. If $x = 0$, then $f = 0$ implies $y = 0$ or $z = 0$. Thus we find the two components $X_1 = Z(x, y)$ and $X_2 = Z(x, z)$. If $z = 1$, then f implies that $y = x^2$, so we obtain a component $X_3 = Z(y - x^2, z - 1)$. Thus using the Nullstellensatz, we find

$$\sqrt{(f, g)} = (x, z) \cap (x, z) \cap (y - x^2, z - 1)$$

☆

Example 2.26 Consider the algebraic set $W = Z(F, G)$ in $\mathbb{A}^4$ where $F = x^2 - yw$ and $G = xy - zw$. Then it is clear that W contains the plane $L = Z(x, w)$. Let

$$C = Z(x^2 - yw, xy - zw, y^2 - xz)$$

We claim that a decomposition into irreducibles is given by $W = C \cup L$.

The $\supseteq$-containment is clear, since $F, G \in I(C) \cap (x, w)$. Conversely, let $p = (x, y, z, w) \in W$. If $w = 0$, then $x = 0$ by F, so $p \in L$. If $w \neq 0$, we find that $y = x^2 w^{-1}$ and $z = w^{-1} xy = x^3 w^{-2}$. Thus $y^2 - xz = x^4 w^{-2} - x^4 w^{-2} = 0$, and hence $p \in C$.

In fact, a primary decomposition is given by

$$(F, G) = (x, w) \cap (x^2 - yw, zw - xy, xz - y^2).$$

☆

2.27 A decomposition result as in Proposition 2.24 above holds for a much broader class of topological spaces than the closed algebraic sets. The class in question is the class of the so-called *Noetherian topological spaces*; these satisfy the condition that every descending chain of closed subsets is eventually stable, i.e., if $\{X_i\}$ is a collection of closed subsets forming a chain

Irredundant decompositions
irreduntante dekomposisjoner

$$\ldots X_{i+1} \subseteq X_i \subseteq \ldots \subseteq X_2 \subseteq X_1,$$

it holds true that for some index r one has $X_i = X_r$ for $i \geqslant r$. It is easy to establish, and left to the zealous students, that any closed subset of a Noetherian space endowed with the induced topology is again Noetherian.

In algebraic geometry it is also customary to call a topological space *quasi-compact* if every open covering can be reduced to a finite covering (this notion usually goes under the name "compact" nowadays).

Quasi-compact
kvasikompakt

Noetherian topological spaces
noetherske topologiske rom

LEMMA 2.28 *Let X be a topological space. The following three conditions are equivalent:*

i) X is Noetherian;
ii) Every open subset of X is quasi-compact;
iii) Every non-empty family of closed subsets of X has a minimal member.

Common usage among several mathematicians is that a compact space is Hausdorff. The "Zariski-like" spaces are far from being Hausdorff, therefore the notion quasi-compact.

PROOF: Assume to begin with that X is Noetherian and let Σ be a family of closed sets without a minimal element. One then easily constructs a strictly descending chain that is not stationary by recursion. Assume a chain

$$X_r \subset X_{r-1} \subset \ldots \subset X_1$$

of length r has been found; to extend it just append any subset in Σ strictly contained in X_r, which does exist since Σ by assumption has no minimal member.

Next, assume that every non-empty family of closed subsets has a minimal member and let an open covering $\{U_i\}$ of an open subset U of X be given. Introduce the family Σ consisting of the closed sets that are finite intersections of complements of members of the covering; *i.e.* the sets of the shape $U_{i_1}^c \cap \ldots \cap U_{i_r}^c$. It has a minimal element Z. If U_j is any member of the covering, it follows that $Z \cap U_j^c = Z$, hence $U_j \subseteq Z^c$, and by consequence $U = Z^c$.

Finally, suppose that every open U in X is quasi-compact and let $\{X_i\}$ be a descending chain of closed subsets. The open set $U = X \backslash \bigcap_i X_i$ is quasi-compact by assumption, and it is covered by the ascending collection $\{X_i^c\}$, hence it is covered by finitely many of them. The collection $\{X_i^c\}$ being ascending, we can infer that $X_r^c = U$ for some r; that is, $\bigcap_i X_i = X_r$ and consequently it holds that $X_i = X_r$ for $i \geqslant r$. ❏

2.29 The Noether–Lasker decomposition of closed subsets in affine space as a union of irreducibles can be generalized to any Noetherian topological space:

Theorem 2.30 *Every closed subset Y of a Noetherian topological space X has an irredundant decomposition $Y = Y_1 \cup \ldots \cup Y_r$ where each is Y_i is a closed and irreducible subset of X. Furthermore, the decomposition is unique up to order.*

The Y_i's that appear in the theorem are called the *irreducible components* of Y. They are *maximal* among the closed irreducible subsets of Y.

The last statement in the lemma leads to the technique called *Noetherian induction* – proving a statement about closed subsets, one can work with a minimal "crook"; *i.e.* a minimal counterexample.

Proof: We shall work with the family Σ of those closed subsets of X that cannot be decomposed into a finite union of irreducible closed subsets; or phrased in a different way, the set of counter examples to the assertion – and of course, we shall prove that it is empty.

Assuming the contrary – that Σ is non-empty – we can find a minimal element Y in Σ because X by assumption is Noetherian. The set Y itself can not be irreducible, so $Y = Y_1 \cup Y_2$ where both the Y_i's are proper subsets of Y and therefore do not belong to Σ. Either is thus a finite union of closed irreducible subsets, and consequently the same is true for their union Y. We have a contradiction, and Σ must be empty.

As to uniqueness, assume that we have a counterexample; that is, two irredundant decomposition such that $Y_1 \cup \ldots \cup Y_r = Z_1 \cup \ldots \cup Z_s$ and such that one of the Y_i's, say Y_1, does not equal any of the Z_k's.

Since Y_1 is irreducible and $Y_1 = \bigcup_k (Z_k \cap Y_1)$, it follows that $Y_1 \subseteq Z_k$ for some k. A similarly argument gives $Z_k = \bigcup_i (Z_k \cap Y_i)$ and Z_k being irreducible, it holds that $Z_k \subseteq Y_i$ for some i, and therefore $Y_1 \subseteq Z_k \subseteq Y_i$. Since the union of the Y_i's is irredundant, we infer that $Y_1 = Y_i$, and hence $Y_1 = Z_k$. Contradiction. ❑

Exercise 2.2 Let X be a topological space and let $Z \subseteq X$ be an irreducible component of X. Let U be an open subset of X and assume that $U \cap Z$ is nonempty. Show that $Z \cap U$ is an irreducible component of U. ★

2.31 You should already have noticed the resemblance of the conditions of being Noetherian for topological spaces and for rings – both are chain conditions (from where the name Noetherian spaces comes). When X is a closed algebraic set in $\mathbb{A}^n$, the one-to-one correspondence between the prime ideals in the coordinate ring $A(X)$ and the closed irreducible sets in X, yields that X is a Noetherian space; indeed, Hilbert's Basis Theorem[6] implies that $A(X)$ is a Noetherian ring, so any ascending chain $\{I(X_i)\}$ of prime ideals corresponding to a descending chain of $\{X_i\}$ of closed irreducibles, is stationary. We have

Irreducible components
irreduktible komponenter

Proposition 2.32 *If X is a closed algebraic subset of $\mathbb{A}^n$, then X is a Noetherian space.*

[6] See *e.g.* Theorem **??** in CA

Example 2.33 *(Hypersurfaces again)* It is often difficult to prove that an algebraic set X is irreducible, or equivalently that the ideal $I(X)$ is a prime ideal. This can be challenging even when X is a hypersurface.

Generally, to find the primary decomposition of an ideal can be quite hard. In addition to the problems of finding the minimal primes and the corresponding primary components, which frequently can be attacked by geometric methods, one has the notorious problem posed by embedded components. They are annoyingly well hidden from geometry.

There are examples of non-noetherian rings with just one maximal ideal, so an ascending chain condition on prime ideals does not imply that the ring is Noetherian.

If $X = Z(f)$ is a hypersurface in $\mathbb{A}^n$, there will be no embedded components since the polynomial ring[7] is a UFD. Indeed, one easily sees that $(f) \cap (g) = (fg)$ for polynomial without common factors. Hence one infers by induction that

$$(f) = (f_1^{a_1}) \cap \ldots \cap (f_r^{a_r}),$$

where $f = f_1^{a_1} \cdot \ldots \cdot f_r^{a_r}$ is the factorization of f into irreducibles, and observes there are no inclusions among the prime ideals (f_i). ☆

EXAMPLE 2.34 *(Homogeneous polynomials)* Recall that a polynomial f is *homogenous* if all the monomials that appear (with a non-zero coefficient) in f are of the same total degree. Recollecting terms of the same total degree, one sees that any polynomial can be written as a sum $f = \sum_i f_i$ where the f_i's are homogeneous of degree i; and since homogeneous polynomials of different total degrees are linearly independent, such a decomposition is unique.

[7] For a proof of this see Theorem **??** in CA.

If a homogeneous polynomial f factors as a product $f = a \cdot b$, the polynomials a and b will also *homogeneous*. (Sometimes this can make life easier if you want *e.g.* to factor f or to show that f is irreducible.) Indeed, if $a = \sum_{0 \leqslant i \leqslant d} a_i$ and $b = \sum_{0 \leqslant j \leqslant e} b_j$ with a_i's and b_j's homogeneous of degree i and j respectively and with $a_d \neq 0$ and $b_e \neq 0$, one finds

$$f = ab = \sum_{i+j<d+e} a_i b_j + a_d b_e$$

Since the decomposition of f in homogeneous parts is unique, it follows that $f = a_e b_d$. ☆

EXAMPLE 2.35 *(Quadrics)* The polynomial $f(x) = x_1^2 + x_2^2 + \ldots + x_n^2$ is irreducible when $n \geqslant 3$ and the characteristic of k is not equal to two. To check this, we may clearly assume that $n = 3$ (*e.g.*, by specialization). Suppose there is a factorization like

$$x_1^2 + x_2^2 + x_3^2 = (a_1x_1 + a_2x_2 + a_3x_3)(b_1x_1 + b_2x_2 + b_3x_3).$$

Observing that $a_1 \cdot b_1 = 1$ and replacing a_i by a_i/a_1 and b_i by b_i/b_1, one may assume that $a_1 = b_1 = 1$ and the equation takes the shape

$$x_1^2 + x_2^2 + x_3^2 = (x_1 + a_2x_2 + a_3x_3)(x_1 + b_2x_2 + b_3x_3).$$

Putting $x_3 = 0$, and using the factorization $x_1^2 + x_2^2 = (x_1 + ix_2)((x_1 - ix_2)$, one easily brings the equation on the form

$$x_1^2 + x_2^2 + x_3^2 = (x_1 + ix_2 + a_3x_3)(x_1 - ix_2 + b_3x_3),$$

from which one obtains $a_3 = b_3$ and $a_3 = -b_3$ by equating cross terms. Since the characteristic is not equal to two, this is a contradiction. If k is of characteristic two the polynomial $x_1^2 + x_2^2 + x_3^2$ is not irreducible; it holds true that $x_1^2 + x_2^2 + x_3^2 = (x_1 + x_2 + x_3)^2$. ☆

Homogeneous polynomials
homogene polynomer

Example 2.36 *(Monomial ideals)* An ideal $\mathfrak{a}$ is said to be *monomial* if it is generated by monomials. Such an ideal $\mathfrak{a}$ has the property that if a polynomial f belongs to it, all the monomial terms appearing in f belong to it as well. To verify this, write f as a sum $f = \sum M_i$ of monomial terms[8] and choose a set $\{N_j\}$ of monomial generators for $\mathfrak{a}$. Then one infers that

Monomial ideals
monomiale idealer

$$f = \sum_i M_i = \sum_j P_j N_j = \sum_{j,k} A_{kj} N_j,$$

where the P_j's are polynomials whose expansions in monomial terms are $P_j = \sum_k A_{kj}$. Since different monomials are linearly independent (by definition of a polynomial), every term M_i is a linear combination of the monomial terms $A_{kj}N_j$ corresponding to the same monomial, and hence M_i lies in the ideal $\mathfrak{a}$. ☆

Example 2.37 Monomial ideals are much easier to work with than general ideals. As an easy example, consider the union of the three coordinate axes in $\mathbb{A}^3$. It is given as the zero locus of the ideal $\mathfrak{a} = (xy, xz, yz)$, and one has

$$(xy, xz, yz) = (x, y) \cap (x, z) \cap (y, z)$$

Indeed, one inclusion is trivial; for the other it suffices to show that a monomial in $(x, y) \cap (x, z) \cap (y, z)$ belongs to (xy, xz, yz). But $x^n y^m z^l$ lies in $(x, y) \cap (x, z) \cap (y, z)$ precisely when at least two of the three integers n, m and l are non-zero, which as well is the requirement to lie in $\mathfrak{a}$. ☆

2.3 Dimension

2.38 Let X be a topological space. Motivated by the Krull dimension of a ring, we try to define the dimension of X by considering strictly increasing chains of non-empty *closed* and *irreducible* subsets:

$$X_0 \subsetneq X_1 \subset \ldots \subsetneq X_r, \tag{2.1}$$

and call r (that is, the number of inclusions in the chain) *the length* of the chain. The *dimension* $\dim X$ of X is the supremum of the set of r's for which there is a chain as in (2.1).

The dimension of a topological space
dimensjonen til et topologisk rom

One says that the chain is *saturated* if there is no closed irreducible subset strictly in between any two of the X_i's; that is, if Z is a closed and irreducible subset such that $X_i \subseteq Z \subseteq X_{i+1}$, then either $Z = X_i$ or $Z = X_{i+1}$. Moreover, we shall call a saturated chain *maximal* if it neither can be lengthened upwards nor downwards; in other words, it is not part of any longer chain. Clearly, the

Saturated chains
mettede kjeder

[8] A *monomial term* is of the form $\alpha \cdot M$ where α is a scalar and M a monomial.

supremum over the lengths of saturated chains, or for that matter, of the maximal chains, will be equal to the dimension of the space. But be aware that a maximal chain is not necessarily of maximal length, there might be others that are longer.

2.39 Note that the dimension of X may be equal to ∞, and in fact, this can happen also for Noetherian spaces (see Exercise 2.5 below and Subection **??** on page **??** in *CA* for a Noetherian ring of infinite dimension). However, the spaces we will concern ourselves with, algebraic varieties, will not have these sorts of pathologies and they will always have finite dimension.

Exercises

2.3 The notion of dimension we introduced is only useful for "Zariski-like" topologies. Show that any Hausdorff space is of dimension zero. Hint: What are the irreducible subsets?

2.4 Assume that $Y = Y_1 \cup \ldots \cup Y_r$ is the decomposition of the Noetherian space Y into irreducible components. Show that $\dim Y = \max \dim Y_i$.

2.5 The following weird topology on the set $\mathbb{N}$ of natural numbers is one example of a Noetherian space with infinite dimension. The closed sets of this topology apart from the empty set and the entire space, are the sets defined by $Z_a = \{ x \in \mathbb{N} \mid x \leqslant a \}$ for $a \in \mathbb{N}$. They form a strictly ascending infinite chain and are irreducible, hence the dimension is infinite. On the other hand, any strictly descending chain is finite so the space is Noetherian. Prove these assertions.

★

A few basic properties of dimension

2.40 One immediately establishes the following basic properties

Lemma 2.41 *Assume that X is a topological space and that $Y \subseteq X$ is a closed subspace. Then $\dim Y \leqslant \dim X$. Assume furthermore that Y is irreducible and that X is of finite dimension. If $\dim Y = \dim X$, then Y is a component of X.*

Proof: Any chain as (2.1) in Y will be one in X as well; hence $\dim Y \leqslant \dim X$. For the second assertion, assume that $\dim Y = \dim X = r$, and let

$$Y_0 \subset Y_1 \subset \ldots \subset Y_r = Y$$

be a maximal chain in Y. In case Y is not a component of X, there is a closed and irreducible subset Z of X strictly bigger than Y, and we can extend the chain to

$$Y_0 \subset Y_1 \subset \ldots \subset Y_r \subset Z.$$

Hence $\dim X \geqslant r+1$, and we have a contradiction. ❏

2.42 It is slightly counterintuitive, but for general topological spaces, a dense open subset does not necessarily have the same dimension as the surrounding space even when the surrounding space is irreducible. But at least it cannot be bigger:

Proposition 2.43 *Assume that X is an irreducible topological space and that U is an open subset. Then $\dim U \leqslant \dim X$.*

Proof: We need to show that $\dim U \leqslant \dim X$, so let

$$U_0 \subset U_1 \subset \ldots \subset U_r$$

be a chain of closed irreducible subsets of U. By Lemma 2.12 on page 24 the closures $\overline{U_i}$ are irreducible closed subsets of X and they satisfy $\overline{U_i} \cap U = U_i$. Hence the chain $\{\overline{U_i}\}$ form a strictly ascending chain of closed irreducible sets in X. It follows that $r \leqslant \dim X$, and thence $\dim U \leqslant \dim X$ since the chain was arbitrary. ❏

2.44 In general strict inequality might hold in Lemma 2.43. The so-called Sierpiński space is a stupidly simple example. It has merely two points, η which is open, and x which is closed. Clearly $\{\eta\}$ is an open dense subset of dimension zero, whereas the Sierpiński space itself is of dimension one since it has the maximal chain $\{x\} \subseteq X$ of non-empty closed irreducible subsets.

The situation is, however, satisfactory for for varieties. As we shall establish later, their dimension coincides with the transcendence degree of their function field, and consequently equality prevails in Proposition 2.43, open dense subsets having the same function field as the surrounding variety (Corollary 7.24 on page 150).

2.45 Our concept of dimension coincides, when X is a closed irreducible subset of $\mathbb{A}^m$, with the Krull dimension of the coordinate ring $A(X)$. Indeed, the correspondence between closed irreducible subsets of X and prime ideals in $A(X)$, implied by the Nullstellensatz, yields a bijective correspondence between chains

$$X_0 \subset X_1 \subset \ldots \subset X_r$$

of closed irreducible subsets and chains

$$I(X_r) \subset \ldots \subset I(X_1) \subset I(X_0)$$

of prime ideals in $A(X)$. Hence the suprema of the lengths in the two cases are the same, and we have:

Proposition 2.46 *For each closed algebraic subset $X \subseteq \mathbb{A}^m$ it holds that*

$$\dim X = \dim A(X).$$

Maximal chains
maksimale kjeder

If X has a decomposition into irreducible subsets $X_1 \cup \ldots \cup X_s$, then the dimension of X equals the maximum of the dimensions of each X_i, i.e.,

$$\dim X = \sup_i \dim A(X_i).$$

PROPOSITION 2.47 $\dim \mathbb{A}^n = n$.

PROOF: This follows from the fact that the polynomial ring $k[x_1, \ldots, x_n]$ has Krull dimension n. See Chapter 11 in [CA]. ❏

The fact that the polynomial ring has Krull dimension n is not surprising, but is astonishingly subtle to establish. This reflects the fact that the polynomial ring $R[t]$ over a non-Noetherian ring R does not always behave decently when it comes to dimension; it can have a Krull dimension other than $\dim R + 1$. (See Subsection **??** on page **??** in CA.) In any case, we shall give a proof that $\dim \mathbb{A}^n = n$ using Noether's Normalization Lemma; see Theorem 7.22 on page 149 below.

From Lemma 2.41, we now deduce

COROLLARY 2.48 *Any closed algebraic sets $X \subset \mathbb{A}^n$ has finite dimension.*

EXAMPLE 2.49 Consider the closed algebraic set $X = Z(I)$ in $\mathbb{A}^2$, where

$$I = \left(x^2y + y^2 + y, x^3 + xy + x\right).$$

A primary decomposition of I is given by $I = \left(x^2 + y + 1\right) \cap (x, y)$. Thus X has two irreducible components $X_1 = Z(x^2 + y + 1)$, and $X_2 = Z(x, y)$. Note that X_1 has coordinate ring

$$A(X_1) = k[x, y]/(x^2 + y + 1) \simeq k[x]$$

which has Krull dimension 1. The second component, X_2, is a point (the origin in $\mathbb{A}^2$), so it has dimension 0. Hence X has dimension 1. ☆

EXAMPLE 2.50 In Example 2.26, we have $\dim C = \dim L = 2$, whence $\dim W = 2$ (as one would expect since W is defined by two equations.) ☆

Exercises

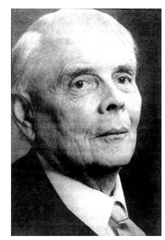
2.6 Let $X = Z(zx, zy) \subseteq \mathbb{A}^3$. Describe X and determine $\dim X$. Exhibit two maximal chains of irreducible subvarieties of different lengths. Exhibit a hypersurface Z so that $Z \cap X$ is of dimension zero.

2.7 Let $X = Z(xy, y(y-1)) \subseteq \mathbb{A}^2$. Describe all chains of closed algebraic sets in X.

2.8 Let $A = k[x_1, x_2, x_3]$ and let $\mathfrak{p}$ and $\mathfrak{q}$ be the two primeideals $\mathfrak{p} = (x_1)$ and $\mathfrak{q} = (x_2, x_3)$. Let S multiplicative system $S = A \backslash (\mathfrak{p} \cup \mathfrak{q})$. Show that $B = A_S$ is a Noetherian semi-local domain with the two maximal ideals $\mathfrak{m} = \mathfrak{p}A_s$ and $\mathfrak{n} = \mathfrak{q}A_S$. Show that $\dim A_{\mathfrak{m}} = 1$ and $\dim B_{\mathfrak{n}} = 2$, and conclude that A is a Noetherian domain with two maximal chains whose lengths differ.

2.4 *Polynomial maps between algebraic sets*

The final topic we approach in this chapter are the so-called *polynomial maps* between closed algebraic sets. This is a precursor of the notion of "morphisms" between general varieties in the next section, but we find the present premature mentioning worthwhile. Polynomial maps between algebraic sets are conceptually much simpler than morphisms, and having seen it should make it easier for students to absorb the general definition. And in fact, in many concrete cases one works with polynomials. In the end of course, the two concepts of polynomial maps and morphisms between closed algebraic sets turn out to be the same.



Wolfgang Krull
(1899–1971)
German mathematician

The coordinate ring

Polynomial functions
polynomfunksjoner

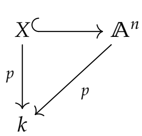
2.51 Let $X \subseteq \mathbb{A}^n$ be a closed algebraic set. A *polynomial function* (later on they will also be called *regular functions*) on X is just the restriction to X of a polynomial on $\mathbb{A}^n$; in other words, it is a polynomial $p \in k[x_1, \ldots, x_n]$ regarded as a function on X. Two polynomials p and q restrict to the same function precisely when the difference $p - q$ vanishes on X; that is to say, the difference $p - q$ belongs to the ideal $I(X)$. We infer that the polynomial functions on X correspond exactly to the elements in the coordinate ring $A(X) = k[x_1, \ldots, x_n]/I(X)$.

Regular functions
regulære funksjoner

Polynomial maps

2.52 Now, given another closed algebraic subset $Y \subseteq \mathbb{A}^m$ and a map $\phi\colon X \to Y$. Composing ϕ with the inclusion of Y in $\mathbb{A}^m$, we may consider ϕ as a map from X to $\mathbb{A}^m$ that takes values in Y; and as such, it has m component functions $q_1, \ldots, q_m$, and we write

$$\begin{aligned} \phi : X &\rightarrow Y \\ x &\mapsto (q_1(x), \ldots, q_m(x)) \end{aligned}$$



We say that ϕ is a *polynomial map* if the component q_i are polynomial functions on X. The set of polynomial maps from X to Y will be denoted by hom[AS]XY.

EXAMPLE 2.53 We have already seen several examples. For instance, the parametrization of a rational normal curve[9] C_n is a polynomial map from $\mathbb{A}^1$ to C_n whose component functions are the powers t^i. ☆

Polynomial maps
polynomavbildniger, polynomiale avbildninger

The condition that ϕ takes values in Y, imposes constraints on ϕ, namely that $f \circ \phi = 0$ must hold for all polynomials f on $\mathbb{A}^m$ which vanish on Y; that is to say, for all $f \in I(Y)$. Expressed in terms of the component functions q_i of ϕ, this means that $f(q_1(x), \dots, q_n(x)) = 0$ for all $x \in X$ and all $f \in I(Y)$. Thus the composed map $f \circ \phi$ only depends on the residue class of f mod $I(Y)$, and compositing with ϕ induces an algebra homomorphism ϕ^*

$$\phi^* \colon A(Y) \to A(X) \qquad f \mapsto f \circ \phi.$$

It turns out that the association $\phi \mapsto \phi^*$ is a bijection:

THEOREM 2.54 *Let X and Y be two closed algebraic sets. Then the map*

$$\mathrm{Hom}_{\mathsf{AS}}(X, Y) \longrightarrow \mathrm{Hom}_{\mathsf{Alg}}(A(Y), A(X))$$

that sends ϕ to ϕ^* is a bijection from the set of polynomial maps to the set of algebra homomorphisms.

PROOF: Note that the effect of the canonical surjection $\pi \colon k[y_1, \dots, y_m] \to A(Y)$ on a polynomial is to restrict it to Y. So if $\phi \colon X \to Y$ is a polynomial map, the functions $\phi^*(\pi(y_i))$ on X are just the component functions of ϕ. Hence, an equality $\phi^* = \psi^*$ between polynomial maps implies that ϕ and ψ have same components, and consequently they are equal.

On the other hand, given an algebra homomorphism $\alpha \colon A(Y) \to A(X)$, we let ϕ be the map on X having components $q_i = \alpha(\pi(y_i))$. Since α kills $I(Y)$, our function ϕ takes values in Y, and obviously $\phi^* = \alpha$ since they coincide on the generators $\pi(y_i)$. ❑

PROPOSITION 2.55 *Polynomial maps are continuous in the Zariski topology.*

PROOF: Assume that $\phi \colon X \to Y$ is the polynomial map. Given a polynomial function q on Y, clearly $\phi^* q = q \circ \phi$ is a polynomial function on X, and hence $\phi^{-1}(Z(q)) = Z(\phi^* q)$ is closed in X. Consequently, for any closed set $Z(q_1, \dots, q_m)$, we find that the inverse image $\phi^{-1} Z(q_1, \dots, q_r) = \bigcap_i \phi^{-1} Z(q_i)$ is closed. ❑

EXAMPLE 2.56 Consider the curve C given by the image of $\phi : \mathbb{A}^1 \to \mathbb{A}^3$ given by $\phi(t) = (t^3, t^4, t^5)$. Looking at monomials, it is easy to find some elements in $I(C)$:

$$q_0 = y^2 - xz, \quad q_1 = yz - x^3, \quad q_2 = z^2 - x^2 y$$

Let $J = (q_0, q_1, q_2) \subset I(C)$. We claim that the reverse containment also holds. For this, introduce a non-standard grading on $k[x, y, z]$ by setting $\deg x = 3, \deg y = 4, \deg z = 5$. Then the f_0, f_1, f_2 are homogeneous with respect to this grading. Let $f \in I(C)$. Then modulo J, we may write

$$f(x, y, z) = A(x) + B(x)y + C(x)z$$

If we evaluate both sides on (t^3, t^4, t^5), we see get

$$f(t^3, t^4, t^5) = A(t^3) + B(t^3)t^4 + C(t^3)t^5$$

Since $f \in I(C)$, this should be the zero-polynomial in t. Considering the t-degrees of the polynomials appearing on the right hand side, we see that $A = B = C = 0$ (as elements of $k[x]$). Hence $f = 0$ modulo J, and thus $I(C) = J$. Note that $I(C)$ is not a complete intersection: this follows because q_0, q_1, q_2 have degrees $8, 9, 10$ and there are no relations in degree $\leqslant 7$. ☆

This correspondence is in fact part of a more general phenomenom in commutative algebra.

If $\phi : A \to B$ is any ring homomorphism, then the association $\mathfrak{q} \mapsto \phi^{-1}(\mathfrak{q})$ induces a map between the set of prime ideals in B to those in A, i.e., a map

$$\operatorname{Spec}(B) \to \operatorname{Spec}(A)$$

If we apply this to $A = A(Y)$ and $B = A(X)$, and restrict this map to the set of maximal ideals, the resulting map is exactly the map $\phi : X \to Y$ which corresponds to $A(Y) \to A(X)$ in the correspondence above.

Images and fibres

2.57 Let $\phi\colon X \to Y$ be a polynomial map between closed algebraic sets. To understand a map it is of course important to understand the fibres, i.e., the preimages of points in Y. The following lemma gives a simple criterion for a point to lie in a given fibre.

LEMMA 2.58 *Let $\phi\colon X \to Y$ be a polynomial map between the two closed algebraic sets X and Y, induced by a ring map $h : A(Y) \to A(X)$. Then for closed subsets $Z \subseteq X$ and $W \subseteq Y$, we have*

i) $\overline{\phi(Z)} \subseteq W$ *if and only if* $I(W) \subseteq h^{-1}I(Z)$;

ii) $I(\overline{\phi(Z)}) = h^{-1}I(Z)$;

iii) *If $x \in X$ and $y \in Y$ are two points. Then $\phi(x) = y$ if and only if $\phi^*\mathfrak{m}_y \subseteq \mathfrak{m}_x$.*

PROOF: The situation is depicted below: Applying our correspondence between morphisms and ring maps, we see that any diagram on the left gives rise to one on the right, and vice versa.

$$\begin{array}{ccc} X & \xrightarrow{\phi} & Y \\ \uparrow & & \uparrow \\ Z & \xrightarrow{\phi|_Z} & W \end{array} \qquad\qquad \begin{array}{ccc} A(X) & \xleftarrow{h} & A(Y) \\ \downarrow & & \downarrow \\ A(Z) & \longleftarrow & A(W) \end{array}$$

Since $A(Z) = A(X)/I(Z)$, and $A(W) = A(Y)/I(W)$, we see that the diagram on the right commutes if and only if $h(I(W)) \subseteq I(Z)$, which implies $I(W) \subseteq h^{-1}I(Z)$. This gives (i).

For (ii), we need to prove that $\overline{\phi(Z)} = Z(h^{-1}I(Z))$. The '$\supseteq$' inclusion follows from (i). For the other inclusion, let $y = \phi(x) \in \phi(Z)$. If $f \in h^{-1}I(Z)$, then $h(f) \in I(Z)$, so $f(y) = f(\phi(x)) = (f \circ \phi)(x) = h(f)(x) = 0$, and so $y \in Z(h^{-1}I(Z))$. It follows that $\phi(Z) \subseteq Z(h^{-1}I(Z))$. Taking closures we also get $\overline{\phi(Z)} \subseteq Z(h^{-1}I(Z))$, and so $I(\overline{\phi(Z)}) \supset I(Z(h^{-1}I(Z)) = h^{-1}I(Z)$, by the Nullstellensatz. This gives (ii).

(iii): One has $\phi(x) = y$, if and only if $f(\phi(x)) = 0$ for all $f \in \mathfrak{m}_y$; that is, if and only if $\phi^*(f) = f \circ \phi \in \mathfrak{m}_x$ for all $f \in \mathfrak{m}_y$. ❑

In particular, the fibre $\phi^{-1}(y)$ of ϕ over the a point $y \in Y$ is the closed algebraic set in X given by the ideal $\phi^*\mathfrak{m}_y$. The fibre can of course be empty, in which case $\phi^*\mathfrak{m}_y = A(X)$, and the ideal $\phi^*\mathfrak{m}_y$ need in general not be radical.

Exercises

2.9 Let $\phi\colon \mathbb{A}^1 \to \mathbb{A}^1$ be the map $\phi(t) = t^n$. For each point $a \in \mathbb{A}^1$ determine the ideal $\phi^*\mathfrak{m}_a$ and the fibre $\phi^{-1}(a)$.

2.10 Let $\psi\colon \mathbb{A}^2 \to \mathbb{A}^2$ be the map $\psi(x, y) = (x, xy)$. Determine the ideals $\psi^*\mathfrak{m}_{(a,b)}$ and the fibres $\psi^{-1}(a, b)$ for all points $(a, b) \in \mathbb{A}^2$.

2.11 Let $\psi\colon \mathbb{A}^3 \to \mathbb{A}^3$ be given as $(x, y, z) \mapsto (yz, xz, xy)$. Find all fibres of ψ and their ideals.

★

2.5 *Examples*

We conclude this lecture with a few examples to illustrate different phenomena surrounding polynomial maps. One example to warn that images of polynomial maps can be very complicated, followed by two examples to illustrate that important and interesting sets naturally originating in linear algebra are irreducible algebraic sets.

In the first case, we consider general matrices of rank at most r are considered, but in the second the matrices are confined to be symmetric; *i.e.* we consider quadratic forms of rank at most r. In both cases the technique is to exhibit parameterizations of the sets whose parameter spaces easily are seen to be irreducible, and leaning on Lemma 2.12 on page 24 we then infer that the sets themselves are irreducible.

Images of polynomial maps

EXAMPLE 2.59 Images of polynomial maps can be complicated. In general they are neither closed nor open, but they will as we are to see later on, always be constructible[10].

For example, consider the map $\phi\colon \mathbb{A}^2 \to \mathbb{A}^2$ given as $\phi(u, v) = (u, uv)$. Pick a point (x, y) in the target $\mathbb{A}^2$. If $x \neq 0$, it holds that $\phi(x, x^{-1}y) = (x, y)$ so points lying off the y-axis are in the image. Among points with $x = 0$ however, only the origin belongs to the image since $uv = 0$ when $u = 0$. Hence the image is equal to the union $\mathbb{A}^2 \backslash Z(x) \cup \{(0, 0)\}$. This set is neither closed (it contains an open set and therefore it's closure is all of $\mathbb{A}^2$) nor open (the complement equals $Z(x)\backslash\{(0, 0)\}$ which has closure $Z(x)$, hence it is not closed).

The map ϕ collapses the v-axis to the origin, and, consequently, lines parallel to the u-axis map to lines through the origin; the intersection point with the v-axis is mapped to the origin. Pushing these lines out towards infinity, their images approach the v-axis. So in some sense, the "missing line" that should have covered the v axis, is the "line at infinity". ☆

[9] See Problem 1.11 on page 19

[10] See Pargraph 10.13 and Theorem 10.16 respectively on page 195 and 195.

EXERCISE 2.12 Let ϕ be the map in Example 2.59. Show that ϕ maps lines parallel to the u-axis (that is, those with equation $v = c$) to lines through the origin. Show that lines through the origin (those having equation $v = cu$) are mapped to parabolas. ★

EXERCISE 2.13 Describe the image of the map $\phi\colon \mathbb{A}^3 \to \mathbb{A}^3$ given as $\phi(u, v, w) = (u, uv, uvw)$. ★

Matrix multiplication and the general linear group

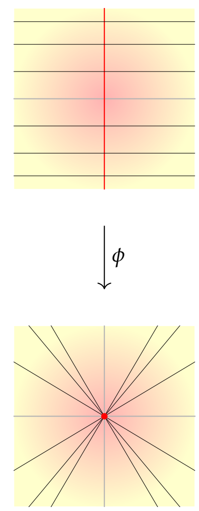
EXAMPLE 2.60 Let n be a natural number and consider the space $\mathbb{M}_{n,n}$ of $n \times n$-matrices with coefficients in k. We may identify $\mathbb{M}_{n,n}$ with the affine space $\mathbb{A}^{n^2}$, and using coordinates $\{x_{ij}\}$ with double indices, both running from 1 to n, we may express a typical element M as the matrix $M = (x_{ij})_{ij}$.

Later on we shall give a very general construction of products of varieties, but for the moment we observe that the Cartesian product $\mathbb{M}_{n,n} \times \mathbb{A}^n$ is naturally identified with $\mathbb{A}^{n^2+n}$ when the latter is equipped with coordinates $\{x_{ij}\}$ and $\{x_i\}$ with $1 \leqslant i, j \leqslant n$. The point of this example is the simple observation that matrix multiplication $\mathbb{M}_{n,n} \times \mathbb{A}^n \to \mathbb{A}^n$ is a polynomial map; indeed, the product is given by the familiar formula

$$(x_{ij})_{ij}(x_i)_i = \left(\sum_r x_{ir}x_r\right)_i$$

and the coordinate functions are quadratic polynomials (in fact, linear in each set of variables).

Inside $\mathbb{M}_{n,n}$ resides the *general linear group* $\mathrm{GL}(n)$ – that is, the set of invertible $n \times n$-matrices – as the complement of the locus where the determinant vanishes. Now, the determinant is a polynomial (of degree n) in the x_{ij}'s, and consequently $\mathrm{GL}(n)$ is an open subset of $\mathbb{M}_{n,n}$ (in fact, it is a distinguished open subset). From Lemma 2.12 on page 24 we see that $\mathrm{GL}(n)$ is irreducible.



There is coordinate free variant of this. Consider a k-vector space E of dimension n. The space $\mathrm{End}_k(E, E) = \mathrm{Hom}_k(E, E)$ of linear endomorphisms of E is a vector space of dimension n^2. Endomorphisms have a determinant, and the endomorphisms for which the determinant does not vanish, constitute the *general linear group* $\mathrm{GL}(E)$.

Of course, depending on a choice of basis for E, the space $\mathrm{End}_k(E, E)$ is isomorphic to matrix space $\mathbb{M}_{n\times n}$ with $\mathrm{GL}(E)$ corresponding to $\mathrm{GL}(n)$. And the map $\mathrm{GL}(E) \times E \to E$ sending (g, α) to $g(\alpha)$ is a polynomial map. ☆

EXERCISE 2.14 Let E and F be two vector spaces. Identify the Cartesian product $\mathrm{GL}(F) \times \mathrm{GL}(E)$ with an open subset of $\mathrm{End}(F, F) \times \mathrm{End}(E, E)$ considered as an affine space isomorphic to $\mathbb{A}^{n^2+m^2}$, and conclude that it is irreducible. Convince yourself that the composition map $((g, h), A) \mapsto g \circ A \circ h$ is polynomial. ★

EXERCISE 2.15 Show that the multiplication map $\mathbb{M}_{n,r} \times \mathbb{M}_{r,m} \to \mathbb{M}_{n,m}$ is a polynomial map. ★

Determinantal varieties

EXAMPLE 2.61 Determinantal varieties are as the name indicates, closed algebraic sets defined by the vanishing of certain determinants. They are much studied and play a prominent role in mathematics.

Let E and F be two vector spaces whose dimensions are n and m respectively and assume that $m \leqslant n$. Inside the space $\mathbb{M} = \mathrm{hom}[k]FE$ of k-linear maps from F to E, we consider the subset of maps having rank at most r:

$$W_r = \{\phi : F \to E \mid \mathrm{rank}\ \phi \leqslant r\}.$$

The subsets W_r form a chain: one obviously has $W_{r-1} \subseteq W_r$, and of course W_m is the entire space $\mathbb{M}$. We shall prove:

PROPOSITION 2.62 *The sets W_r are irreducible closed subsets of $\mathbb{M}$.*

PROOF: We may certainly assume that $r < m$ since$W_m = \mathbb{M}$. The proof proceeds in two steps. We shall show that the open subset $V = W_r \backslash W_{r-1}$ of W_r consisting of maps of rank exactly equal to r, is irreducible and that W_{r-1} lies in the closure of V in $\mathbb{M}$. Then we conclude that W_r is irreducible using Lemma 2.12 on page 24.

Choosing bases $\{f_i\}$ and $\{e_j\}$ for F and E we may identify $\mathbb{M}$ with the space $\mathbb{A}^{mn}$ of matrices with coordinates x_{ij}. That the rank of a matrix M is at most r can be expressed by the condition that the $(r+1) \times (r+1)$-minors of M vanish. These minors are polynomials in the x_{ij}'s and hence W_r is a closed algebraic subset.

We begin with showing that the open subset $V = W_r \backslash W_{r-1}$ of W_r consisting of maps of rank exactly equal to r is irreducible. The point is that the natural action of the group $\mathrm{GL}(E) \times \mathrm{GL}(F)$ on $\mathbb{M}$ leaves V invariant (this is obvious) and is transitive.

Let A_0 be the map that for $1 \leqslant i \leqslant r$ sends the basis vector f_i to e_i and maps the rest of the f_i's to zero; that is,

$$A_0(f_i) = \begin{cases} e_i & \text{when } 1 \leqslant i \leqslant r; \\ 0 & \text{else.} \end{cases}$$

Clearly A_0 has rank r. If A is any other map of rank r, we choose a basis $\{f_i'\}$ for F such that the last elements $f_{r+1}', \ldots, f_m'$ form a basis for the kernel $\ker A$. Then the first vectors $e_i' = A(f_i')$ for $1 \leqslant i \leqslant r$ will be linearly independent in E and can be extended to a basis for E by adjoining elements e_i' for $r+1 \leqslant i \leqslant n$. So displayed it looks like

$$A(f_i') = \begin{cases} e_i' & \text{when } 1 \leqslant i \leqslant r; \\ 0 & \text{else.} \end{cases}$$

Define linear maps $h\colon F \to F$ by $h(f_i') = f_i$ and $g\colon E \to E$ by $g(e_i) = e_i'$. They are both invertible, and one easily verifies that $A = g \circ A_0 \circ h$. Hence the map

$$\mathrm{GL}(E) \times \mathrm{GL}(F) \to V$$

that sends (g, h) to $g \circ A_0 \circ h$ is surjective. It is a polynomial map and according to Lemma 2.12 continuous image of irreducible sets are irreducible, and consequently the subset V is irreducible.

Our next observation will be that W_r equals the closure of $W_r \backslash W_{r-1}$ in $\mathbb{M}$. To see this, let A be any map in W_r and let its rank be s, which we may assume is less than r. It will suffice to exhibit a polynomial map[11] $\alpha_t\colon \mathbb{A}_1 \to W_r$ with $\alpha_0 = A$ and so that α_t is of rank r for $t \neq 0$. Indeed, if U is open neighbourhood of A in W_r the inverse image $\alpha^{-1}(U)$ will be non-empty and open in $\mathbb{A}_1$ and hence it meets $\mathbb{A}^1 \backslash \{0\}$ (because $\mathbb{A}^1$ is irreducible); consequently U meets $W_r \backslash W_{r-1}$ and A lies in the closure.

The general linear group
den generelle lineære gruppen

To construct the desired one-parameter family α we choose bases $\{e_i\}$ and $\{f_j\}$ in such a way that A satisfies

$$A(f_i) = \begin{cases} e_i & \text{when } i \leqslant s\ ; \\ 0 & \text{else,} \end{cases}$$

which is possible because A has rank s, and we obtain our one-parameter family simply by putting

$$\alpha_t(f_i) = \begin{cases} e_i & \text{when } i \leqslant s, \\ te_i & \text{for } s < i \leqslant r, \\ 0 & \text{else.} \end{cases}$$

❏

☆

EXERCISE 2.16 The subset W_1 of $\mathrm{Hom}_k(F, E)$ of maps of rank one has another parametrization which will be described in this exercise.

a) Show that any rank one map $F \to E$ factors as

$$F \xrightarrow{\phi} k \xrightarrow{\iota} E$$

where ϕ and ι are linear maps.

b) Conclude that W_1 is the image of $F^* \times E \to \mathrm{Hom}_k(F, E)$ that sends a pair (ϕ, ι) to the composition $\iota \circ \phi$.

c) Prove that W_1 is equal to the subset of $F^* \otimes_k E$ consisting of decomposable tensors.

★

Quadratic forms

EXAMPLE 2.63 Recall that a *quadratic form* is a homogeneous polynomial of degree two. It is shaped like $P(x) = \sum_{i,j} a_{ij} x_i x_j$ where both indices i and j run from 1 to n. In that sum $a_{ij} x_i x_j$ and $a_{ji} x_j x_i$ appear as separate terms, but as a matter of notation, one organizes the sum so that $a_{ij} = a_{ji}$. Coalescing the terms $a_{ij} x_i x_j$ and $a_{ji} x_j x_i$, the coefficient in front of $x_i x_j$ becomes $2a_{ij}$. For instance, when $n = 2$, a quadratic form is shaped like

$$P(x) = a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2.$$

The *coefficient matrix* of P is the symmetric matrix $A = (a_{ij})$ with $1 \leqslant i, j \leqslant j$, and in terms of A one may express $P(x)$ as the matrix product

$$P(x) = xAx^t$$

[11] Such a map is frequently called a *one-parameter family*

where x denotes the row vector $x = (x_1, \dots, x_n)$ and x^t the transposed vector (which is a column vector).

The set of such quadratic forms – or of such symmetric matrices – constitute a linear space which we shall denote by S_n. It has a basis formed by the monomials x_i^2 and $2x_ix_j$ for $1 \leqslant i, j \leqslant j$, and in our language S_n is isomorphic to an affine space $\mathbb{A}^N$ whose dimension N equals the number of distinct monomials x_ix_j; that is, $N = n(n+1)/2$. The coordinates with respect to this basis are denoted by a_{ij}.

We are interested in the subspaces $Q_r \subseteq S_n$ where the rank of A is at most r. They form a descending chain; that is $Q_{r-1} \subseteq Q_r$; and clearly $Q_n = S_n$ and $Q_0 = \{0\}$.

The Q_r's are all closed algebraic subsets, and the aim of this example is to show they are irreducible:

PROPOSITION 2.64 *The sets Q_r of quadric forms in n variables and of rank at most r are irreducible closed algebraic subset of the space S_n of all quadric forms in n variables.*

PROOF: That the Q_r's are closed, hinges on the fact that a matrix is of rank at most r if and only if all its $(r+1) \times (r+1)$-minors vanish – and of course, these minors are polynomial expressions in the entries of the matrix.

To see that the Q_r's are irreducible, we shall resort to the same technique as in Example 2.61 above and exhibit Q_r as the image of a polynomial map with an irreducible source (*i.e.* $\mathbb{A}^{n^2}$).

By the classical Gram-Schmidt process, any symmetric matrix A can be diagonalized. It may be expressed as a product

$$BAB^t = D$$

where B is an invertible matrix and D a diagonal matrix of the following special form: with r denoting the rank of A the r first diagonal elements of D are 1's and the rest are 0's. Introducing $C = B^{-1}$, we obtain the relation

$$A = CDC^t.$$

Allowing C to be any $n \times n$-matrix, not merely an invertible one, we obtain in this way all symmetric matrices A of rank at most r. Rendering the above considerations into geometry, we introduce a parametrization of the locus Q_r of quadrics of rank at most r, namely the map

$$\Phi\colon \mathbb{M}_{n,n} \to \mathbb{A}^N$$

that sends an $n \times n$-matrix C to CDC^t. This is not a one-to-one map, several parameter values correspond to the same point, but it is a polynomial map, and the Gram-Schmidt process just described, shows that the image equals Q_r. So Φ serves our purpose, and we can conclude that Q_r is irreducible. ❑

☆

Rank one quadrics

EXAMPLE 2.65 To get a better understanding of how a quadratic form of rank r is shaped, one introduces new coordinates $\{y_i\}$, adapted to a specific form with matrix $A = BDB^t$, by the relations $yB = x$; which is legitimate since B and therefore B^t is invertible. Then $xAx^t = yB^tAB^ty^t = yDy^t$. So, in view of the particular shape of D, the quadratic form $P(x)$ expressed in the new coordinates takes the form:

$$P(y) = y_1^2 + \ldots + y_r^2.$$

Of special interest are the sets Q_1 of rank one quadrics, *i.e.* those with $r = 1$. By what we just saw these quadrics are all squares of a linear form in the variables x_i's (remember that y_1 is a linear form in the original coordinates, the x_i's); that is, one has an expression

Quadratic forms
kvadratiske former

$$P(x) = \left(\sum_i u_i x_i\right)^2 = \sum_i u_i^2 x_i^2 + \sum_{i<j} 2u_iu_jx_ix_j.$$

This gives us another parametrization of Q_1, namely the one sending a linear form to its square. The linear forms constitute a vector space of dimension n (one coefficient for each variable), so the "squaring map" is the map

$$\nu \colon \mathbb{A}^n \to \mathbb{A}^N \tag{2.2}$$

sending $(u_1, \ldots, u_n)$ to the point whose coordinates are all different products u_iu_j with $i \leqslant j$ (remember we use the basis for the space of quadrics made up of the squares x_i^2 and the cross terms $2x_ix_j$, in some order). When $n = 3$ we get a mapping of $\mathbb{A}^3$ into $\mathbb{A}^6$ whose image is called the *cone over the Veronese surface*. We will meet the Veronese surface again in Chapter 5.

☆

EXERCISE 2.17 Show that the map v above in (2.2) is not injective, but satisfies $\nu(-u) = \nu(u)$. Show that if $\nu(u) = \nu(u')$, then either $u = u'$ or $u = -u'$. ★

Exercises

2.18 Show that an irreducible space is Hausdorff if and only if it is reduced to a single point.

2.19 Endow the natural numbers $\mathbb{N}$ with the topology whose closed sets apart from $\mathbb{N}$ itself are the finite sets. Show that $\mathbb{N}$ with this topology is irreducible. What is the dimension?

2.20 Show that any countably infinite subset of $\mathbb{A}^1$ is Zariski dense.

2.21 Let X be an infinite set and $Z_1, \ldots Z_r \subseteq X$ be proper infinite subsets of X such any two of them intersect in at most a finite set. Let $\mathcal{T}$ be the set of subsets of X that are either finite, the union of some of the Z_i's and a finite set, the empty set or the entire set X. Show that $\mathcal{T}$ is the set of closed sets for a topology on X. When is it irreducible?

2.22 Let $X \subseteq \mathbb{A}^4$ be the union of the four coordinate axes. Determine the ideal $I(X)$ by giving generators. Describe the Zariski topology on X.

Jacob Steiner (1996–1863)
Swiss mathematician

2.23 Show that any reduced[12] algebra of finite type over k is the coordinate ring of a closed algebraic set.

2.24 Show that any integral domain finitely generated over k is the coordinate ring of an irreducible closed algebraic set.

2.25 Let $\mathfrak{a}$ be the ideal $\mathfrak{a} = (xz, xw, zy, wy)$ in the polynomial ring $k[x, y, z, w]$. Describe the algebraic set $W = Z(\mathfrak{a})$ in $\mathbb{A}^4$ geometrically, and show that the primary decomposition of $\mathfrak{a}$ is

$$\mathfrak{a} = (x, y) \cap (z, w).$$

2.26 Continuing the previous exercise, let $\mathfrak{b}$ be the ideal $\mathfrak{b} = (w - \alpha y)$ with α a non-zero element in k, and let $X = Z(\mathfrak{b})$. Describe geometrically the intersection $W \cap X$. Show that the image $\mathfrak{c}$ of the ideal $\mathfrak{a} + \mathfrak{b}$ in $k[x, y, z]$ under the map that sends w to αy is given as

$$\mathfrak{c} = (xz, xy, zy, y^2),$$

and determine a primary decomposition of $\mathfrak{c}$. What happens if $\alpha = 0$?

2.27 Let two quadratic polynomials f and g in $k[x, y, z, w]$ be given as $f = xz - wy$ and $g = xw - zy$. Describe geometrically the algebraic subset $Z(f, g)$ and find a primary decomposition of the ideal (f, g).

2.28 Let $f = y^2 - x(x-1)(x-2)$ and $g = y^2 + (x-1)^2 - 1$. Show that $Z(f, g) = \{(0, 0), (2, 0)\}$. Determine the primary decomposition of (f, g).

2.29 Let $\mathfrak{a}$ be the ideal $(wy - x^2, wz - xy)$ in $k[x, y, z, w]$. Show that the primary decomposition of $\mathfrak{a}$ is

$$\mathfrak{a} = (w, x) \cap (wz - xy, wy - x^2, y^2 - zx).$$

2.30 Let $\mathfrak{a} = (wz - xy, wy - x^2, y^2 - zx)$. Show that $Z(\mathfrak{a})$ is irreducible and determine $I(X)$.

★

[12] Reduced means that there are no non-zero nilpotent elements.

Chapter 3

Varieties

TOPICS IN CHAPTER 3: *Sheaves of rings; regular function; rational functions; affine varieties; general varieties; morphisms; morphisms into affine varieties; the Hausdorff axiom; products of varieties.*

Jean Leray (1906–1998) French mathematician

Figure 3.1: A sheaf

A central feature of modern geometry is that a space of some geometric type comes equipped with a distinguished set of functions. For instance, topological spaces carry *continuous functions*, smooth manifolds carry C^∞-*functions* and complex manifolds carry *holomorphic functions*. There is a common way of *axiomatically* introducing the different types of 'functions' on a topological space, namely by the so-called *sheaves of rings*. A topological space equipped with a sheaf of rings is called a *ringed space*.

There are many variants of sheaves involving all kinds of structures other than ring structures. However, we confine ourselves to 'sheaves of functions' in this introductory course. Our sole reason to introduce sheaves is that we need them to give a uniform and clear definition of varieties, which are after all our main objects of study. So we cut the theory to a bare minimum (those pursuing studies of algebraic geometry will certainly have the opportunity to be well acquainted with sheaves of all sorts, and we hope they will find it helpful already having seen some sheaves when meeting the full crowd).

Sheaves were invented by the french mathematician Jean Leray during his imprisonment as a prisoner of war during WWII. The (original) french name[1] is *faisceau*.

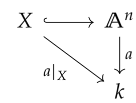
Varieties and the maps between them will be our main objects of study, and in general in geometry they occupy a dominating place. We already met the affine varieties, although disguised as irreducible closed algebraic sets, and in this chapter the club will be substantially enlarged. The newcomers are obtain by 'gluing' affine varieties together; a process that requires some knowledge of sheaves. This 'sheafy way' of eastblishing the basics of algebraic geometry was laid down in the highly influential paper 'Faisceaux algébriques cohérents' ([2]) from 1955 by Jean-Pierre Serre.

[1] In Norwegian one says "knippe" which is close to the meaning of the French word faisceau. When sheaves were introduced in Norway, a discussion arose among the mathematicians about the terminology, some proposed "feså"!!

[2]

3.1 Functions on irreducible algebraic sets

3.1 Consider an irreducible closed algebraic set $X \subseteq \mathbb{A}^n$. Its coordinate ring $A(X) = k[x_1, \ldots, x_n]/I(X)$ is an integral domain, identified with the ring of *polynomial functions* on X (as we saw in Paragraph 2.51 on page 36); they are the functions on X which are restrictions of polynomials in $k[x_1, \ldots, x_n]$.



Definition *For an irreducible closed algebraic set $X \subset \mathbb{A}^n$, we define the* rational function field, *or* the function field *of X, to be the fraction field of the coordinate ring $A(X)$. The elements of $k(X)$ are called* rational functions.

The name comes from the case of the affine line $\mathbb{A}^1$ whose function field equals $k(x)$; the field of rational functions in one variable. Similarly, the function field of $\mathbb{A}^n$ is the field $k(x_1, \ldots, x_n)$ of rational functions in n variables; *i.e.* functions expressible as quotients of two polynomials in the x_i's.

Example 3.3 Consider the algebraic set $X = Z(xw - yz)$ in $\mathbb{A}^4$ with x, y, z, w as coordinates. One readily verifies that $xw - yz$ is an irreducible polynomial, and hence X is irreducible with ideal $I(X) = (xw - yz)$. The coordinate ring

$$A(X) = k[x, y, z, w]/(xw - yz)$$

is an integral domain (but not an UFD). In the fraction field, the element x is invertible, and we may eliminate w using $w = x^{-1}yz$. Hence we have $k(X) \simeq k(x, y, z)$. ☆

3.4 Strictly speaking, rational functions are not functions on X; they are only defined on open subsets of X. Let us make this more precise: if $p \in X$ is a point, one says that a rational function $f \in k(X)$ is *defined* at p, or is *regular* at p, if f can be represented as a fraction $f = a/b$ of two elements in $A(X)$ with the denominator b non-zero at p; that is, $b \notin \mathfrak{m}_p$, where $\mathfrak{m}_p$, where $\mathfrak{m}_p \subset \mathbb{A}(X)$ is the maximal ideal corresponding to p.

The field of rational functions
kroppen av rasjonale funksjoner

An element $f = a/b$ with $a, b \in A(X)$ is certainly defined at all points in the distinguished open subset $D(b)$ of X where the denominator b does not vanish. Be aware, however, that it might very well be defined on a larger set; the stupid example being the constant 1 since $1 = a/a$ for all a. For a less stupid example, consider $X = Z(xw - yz)$ in $\mathbb{A}^3$ and the function $f = x/y = z/w$. It is defined on $D(y) \cup D(w)$ which is not a distinguished open set (see example Example 3.8 below for details). The subtlety lies in the fact that f has different representations as a fraction whose denominators vanish in different sets.

3.5 To a rational function $f \in k(X)$ one associates the ideal $\mathfrak{a}_f$ of *denominators*. It is defined as $\mathfrak{a}_f = \{\, b \in A(X) \mid bf \in A(X) \,\}$, so that $\mathfrak{a}_f$ consists of the different denominators that appear when f is expressed as a fraction in different ways. The role of $\mathfrak{a}_f$ is made clear by the following lemma:

Rational functions
rasjonale funksjoner

Proposition 3.6 *Let $X \subseteq \mathbb{A}^n$ be an irreducible closed algebraic set. The maximal open set where a rational function f is defined, is given by $X \backslash Z(\mathfrak{a}_f)$. A rational function f on X is regular if and only if it belongs to $A(X)$.*

The last statement says that the rational functions which are regular on the entire X, are precisely the restrictions of the polynomial functions.

Proof: Let $p \in X$ be a point. If $\mathfrak{a}_f \nsubseteq \mathfrak{m}_p$, there is an element $b \in \mathfrak{a}_f$ which does not vanish at p and such that $f = a/b$ for some a. Hence f is regular at p. If f is regular at p, one can write $f = a/b$ with $b \notin \mathfrak{m}_p$, hence $\mathfrak{a}_f \nsubseteq \mathfrak{m}_p$. For the second statement, the Nullstellensatz tells us that $Z(\mathfrak{a}_f)$ is empty if and only if $1 \in \mathfrak{a}_f$, which is equivalent to f lying in $A(X)$. ❑

Proposition 3.7 *Let $X \subseteq \mathbb{A}^n$ be an irreducible closed algebraic set and let $b \in A(X)$. The regular functions on the distinguished open subset $D(b)$ of X equals $A(X)_b$.*

Regular functions
regulære funksjoner

Proof: Clearly functions of the form a/b^r are regular on the distinguished open set $D(b)$ since b does not vanish there. For the inclusion the other way round: assume that f is regular on $D(b)$. Because $D(b)^c = Z(b)$, this means that $Z(\mathfrak{a}_f) \subseteq Z(b)$, and we infer by the Nullstellensatz that $b \in \sqrt{\mathfrak{a}_f}$, so $f = a/b^r$ for some r and some $a \in A(X)$, which is precisely to say that $f \in A(X)_b$. ❑

Example 3.8 Consider again the quadric $X = Z(xw - yz)$ in $\mathbb{A}^4$. In the function field $k(X)$ the equalities $f = x/y = z/w$ hold, so that the rational function f is defined on the open set $D(y) \cup D(w)$. We contend that this is the maximal set of definition of f. It will suffice to show that $\mathfrak{a}_f = (y, w)$; the inclusion $(y, w) \subseteq \mathfrak{a}_f$ is clear. So assume that $t \cdot x/y = a \in A(X)$. This means that an equality $tx - ay = cxw - cyz$ holds in $k[x, y, z, w]$ for some polynomial c. Rearranging gives $(t - cw)x = (a - cz)y$, and since the polynomial ring is a UFD and y does not divide x, it follows that $t - cw = dy$ for some polynomial d. But this means that $t \in (y, w)A(X)$. ☆

3.9 The coordinate ring $A(X)$ from the previous example is not a UFD – in fact, it is in some sense the prototype of a k-algebra that is not a UFD – and this is the reason behind f not being defined on a distinguished open subset. One has

Proposition 3.10 *Let $X \subseteq \mathbb{A}^n$ be an irreducible closed algebraic set, and assume that the coordinate ring is a UFD. Then the maximal open subset where a rational function f is defined, is of the form $D(b)$.*

Proof: Let $f \in k(X)$ be a rational function and let $b', b \in \mathfrak{a}_f$ be two elements; that is, it holds true that $f = a/b = a'/b'$ for appropriate members $a, a' \in A(X)$. It follows that $ab' = a'b$, and we may well cancel common factors and assume that a and b (respectively a' and b') are without common factors (remember, the coordinate ring $A(X)$ is a UFD). Now, we can write $b = cg$ and $b' = c'g$ with c and c' without common factors. It follows that $ac' = a'c$, and hence c is a factor in a and c' one in a'. We infer that c and c' are units; indeed, c (respectively c') divides both a and b (respectively both a' and b') which are without common factors. So b and b' are both equal to g up to a unit. Hence $\mathfrak{a}_f = (g)$. ❑

EXAMPLE 3.11 When $n \geqslant 2$, any regular function on $\mathbb{A}^n \backslash \{0\}$ extends to the entire $\mathbb{A}^n$ and is thus a polynomial function. Indeed, the coordinate ring of $\mathbb{A}^n$ is the polynomial algebra $k[x_1, \ldots, x_n]$ which is a UFD, hence the maximal set where a regular function is defined, is of the form $D(f) \subset \mathbb{A}^n$, but when $n \geqslant 2$, $\mathbb{A}^n \backslash \{0\}$ is not of this form (the ideal $(x_1, \ldots, x_n)$ is not a principal ideal). ☆

EXERCISE 3.1 Show that the union $D(y) \cup D(w)$ that appear in Example 3.8 above is not contained in any proper distinguished open subset. ★

EXERCISE 3.2 Let R be a UFD. Show that any prime ideal of height one (that is, a prime ideal properly containing no other prime ideals but the zero ideal) is principal. ★

The local ring at a point

There is a very important subring of $k(X)$ given by all functions that are regular at a point p; the *the local ring* at p :

The ideal of denominators
nevneridealer

$$\mathcal{O}_{X,p} = \{\, f \in k(X) \mid f \text{ is regular at } \ p \,\}.$$

This is indeed a local ring since it coincides with the localization of $A(X)$ at the maximal ideal $\mathfrak{m}_p$. Moreover, the maximal ideal is the ideal $\mathfrak{m}_p \mathcal{O}_{X,p}$; the ideal of functions regular near p which vanish at p.

EXAMPLE 3.12 The local ring of $\mathbb{A}^n$ at the origin is given by

$$\mathcal{O}_{\mathbb{A}^n,0} = k[x_1, \ldots, x_n]_{(x_1,\ldots,x_n)}.$$

The maximal ideal $(x_1, \ldots, x_n)$ consists of all rational functions, regular in a neighbourhood of 0, that vanish at the origin. ☆

EXAMPLE 3.13 The local ring of $X = Z(y^2 - x^3)$ at the origin is given by

$$\mathcal{O}_{X,0} = k[x, y]_{(x,y)}/(y^2 - x^3).$$

The maximal ideal $\mathfrak{m} = (x, y)$ is minimally generated by two elements. Note however that there are $\mathfrak{m}$-primary ideals generated by one element, for instance $\sqrt{(x)} = \mathfrak{m}$. ☆

3.2 Ringed spaces and sheaves

Before introducing the concept of a sheaf of rings, let us remind you of a three very familiar aspects of functions—say continuous functions on open subsets of some space X taking values in a field k equipped with a topology compatible with the field operations. We will write $\mathcal{C}(U)$ for the set of continuous maps $f\colon U \to k$ on an open set U in X.

The first aspects is that $\mathcal{C}(U)$ is *a ring*—one can add and multiply functions—it is even a k-algebra—one may scale by constants. The second aspect is that one may *restrict* continuous functions on U to smaller open sets and the restriction persists being continuous, and of course restrictions of sums are sums and of products are products. Finally, the third aspect is that if two functions say f_1 and f_2 defined on two opens U_1 and U_2 agree on the intersection $U_1 \cap U_2$, they define a function on the union $U_1 \cup U_2$—the slang is that they can be *glued together*. And of course, the same applies to a any larger bunch $\{f_i\}_{i\in I}$ of functions on opens $\{U_i\}_{i\in I}$ as long as they match on the overlaps; that is, $f_i|_{U_i\cap U_j} = f_j|_{U_i\cap U_j}$ for all i and j.

3.14 A sheaf of rings is just an axiomatization of these three aspects. However, we shall[3] refrain from giving the most general definition and just define *a sheaf of (continuous) k-valued functions*, which will be sufficient for our needs.

The ring $A(X)_b$ is the localization of $A(X)$ in the element b; *i.e.* the ring $S^{-1}A(X)$ where S is the multiplicative system $S = \{\, b^i \mid i \in \mathbb{N}\,\}$.

We have assumed that k has been given a field topology, including the fields $\mathbb{R}$ or $\mathbb{C}$ with the Euclidean topology. In our main example we will be identifying k with $\mathbb{A}^1$ with the Zariski topology, so that $f\colon U \to k$ is continuous if and only if $f^{-1}(a) \subset X$ is closed for each $a \in k$. Note that each $\mathcal{C}(U)$ is a k-algebra, and for each open $V \subset U$ there is a natural restriction map $\mathcal{C}(U) \to \mathcal{C}(V)$.

Definition *A* sheaf $\mathcal{O}_X$ of k-valued functions *is a collection of k-subalgebras $\mathcal{O}_X(U) \subset \mathcal{C}(U)$, one for each open subset $U \subset X$ satisfying the following properties:*

i) (Restriction) If $V \subset U$ are open, and $f \in \mathcal{O}_X(U)$, the restriction $f|_V \in \mathcal{C}(V)$ is an element of $\mathcal{O}_X(V)$.

ii) (Gluing axiom) Given an open set $U \subset X$, an open covering $\{U_i\}_{i\in I}$ of U, and elements $f_i \in \mathcal{O}_X(U_i)$ that agree on the overlaps $U_i \cap U_j$; that is

$$f_i|_{U_i\cap U_j} = f_j|_{U_i\cap U_j} \quad \forall i, j \in I.$$

then there is a (unique) $f \in \mathcal{O}_X(U)$ such that $f_i = f|_{U_i}$.

In the context of the last condition, the f_i's may be glued together as continuous functions f, and the request is that this f belongs to $\mathcal{O}_X(U)$.

Definition *A* ringed space $(X, \mathcal{O}_X)$ *is a topological space X equipped with a sheaf of k-valued functions $\mathcal{O}_X$.*

Thus a ringed space is a topological space together with a distinguished class of continuous functions; a collection of rings that behave well with respect to the various restriction maps.

3.17 The concepts 'ringed space' and 'sheaf' are in fact much more general than in the above definition. In general one defines a sheaf to be a collection $\mathcal{F}(U)$ of abelian groups, sets, modules, etc, and restriction maps $\mathcal{F}(U) \to \mathcal{F}(V)$, homomorphisms which are part of the data. In addition to the 'Gluing axiom', there is also an additional axiom, the 'Locality axiom', which essentially says that if two elements $f, g \in \mathcal{F}(U)$ agree on each open set of a covering $\{U_i\}$, they must be equal (this condition is automatically satisfied for our sheaves of k-valued functions). 'Ringed spaces' also allow $\mathcal{O}_X$ to be any sheaf of rings, not just k-algebras of continuous functions.

The general theory of sheaves is rich and powerful, and the extra level of abstraction is necessary in order to fully classify and study algebraic varieties. In this introductory book however, we take a more economical approach and introduce only the sheaves we need, *i.e.* all sheaves will be sheaves of k-valued continuous functions.

Example 3.18 Taking $\mathcal{O}_X(U) = \mathcal{C}(U)$ gives a sheaf of k-valued functions. The only thing to check is that the gluing axiom is satisfied: Given continuous functions $f_i : U_i \to k$ defined on a covering U_i of U, we define $f : U \to k$ by $f(x) = f_i(x)$ for any i such that $x \in U_i$ (this is well defined since the f_i agree on the overlaps). ☆

Example 3.19 Let X be a manifold and let $\mathcal{O}_X$ denote the *sheaf of differentiable functions*, i.e., $\mathcal{O}_X(U)$ consists of differentiable functions $f : U \to \mathbb{R}$. Then $(X, \mathcal{O}_X)$ is a ringed space; one checks that $\mathcal{O}_X$ satisfies the Gluing axiom as in the previous example. ☆

Example 3.20 Here is a non-example. Let $X = \mathbb{R}$ and define $\mathcal{O}_X(U) = \mathbb{R}$ for each $U \subset X$ (so that $\mathcal{O}_X(U)$ consists of the *constant functions* $f : U \to \mathbb{R}$). Then the gluing axiom is not satisfied: If $U = U_1 \cup U_2$ is a *disjoint union*, we can define f_1, f_2 by $f_1(x) = 0$ and $f_2(x) = 1$. These two functions trivially agree on the overlap $U_1 \cap U_2$ (which is empty), but there is no *constant* function $f : U \to \mathbb{R}$ restricting to different values on each U_i. ☆

3.21 If $\phi : X \to Y$ is a continuous map between two topological spaces, we get for each $U \subset Y$ an induced map of k-algebras

The local ring at a point
den lokale ringen i et punkt

$$\begin{aligned} \phi_U^* : \mathcal{C}(U) &\to \mathcal{C}(\phi^{-1}U) \\ f &\mapsto f \circ \phi. \end{aligned}$$

Note that these are compatible with the restriction maps, i.e., if $V \subset U$ are opens in X, we have a diagram

$$\begin{array}{ccc}
\mathcal{C}(U) & \xrightarrow{\phi_U^*} & \mathcal{C}(\phi^{-1}U) \\
\downarrow & & \downarrow \\
\mathcal{C}(V) & \xrightarrow{\phi_V^*} & \mathcal{C}(\phi^{-1}V)
\end{array}$$

where all the maps are maps of k-algebras.

The collection of maps is denoted by ϕ^* we call it the *pullback map* of k-valued functions. The pullback maps are *functorial* in the sense that if ϕ and ψ are composable continuous maps, it holds true that $(\psi \circ \phi)^* = \phi^* \circ \psi^*$. Note the change of order of the two involved maps; the pullback is *contravariant*, as one says.

Using this, we can define what it means to have a morphism of ringed spaces: they are maps $\phi : X \to Y$ so that ϕ^* behaves well with respect to the structure sheaves:

> **Definition** *A* morphism of ringed spaces $\phi : (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ *is a continuous map* $\phi : X \to Y$ *so that* $\phi^*(f) \in \mathcal{O}_X(\phi^{-1}U)$ *for each open* U *and each* $f \in \mathcal{O}_Y(U)$.

If X and Y are two ringed spaces, an *isomorphism* from X to Y is a morphism $\phi \colon X \to Y$ that has an inverse morphism; in other words, there is a morphism $\psi \colon Y \to X$ such that $\psi \circ \phi = \mathrm{id}_X$ and $\phi \circ \psi = \mathrm{id}_Y$. In more concrete terms, this boils down to ϕ being a homeomorphism such that $f \in \mathcal{O}_Y(U)$ if and only if $\phi^* f \in \mathcal{O}_X(\phi^{-1}U)$ for all open subsets $U \subseteq Y$.

[3] following the principle of step-by-step learning, and shielding neophytes from the whole gale of the axiomatic abstraction

Example 3.23 Consider $X = \mathbb{C}$ with sheaves $\mathcal{C}, \mathcal{H}$ defined by

$$\mathcal{C}(U) = \mathcal{C}(U, \mathbb{C}) \text{ and } \mathcal{H}(U) = \{\text{holomorphic } f : U \to \mathbb{C}\}$$

Then the identity map $\mathbb{C} \to \mathbb{C}$ induces a morphism of ringed spaces $(\mathbb{C}, \mathcal{C}) \to (\mathbb{C}, \mathcal{H})$: this is just saying that a holomorphic function is also continuous. But there is no morphism of ringed spaces $(\mathbb{C}, \mathcal{H}) \to (\mathbb{C}, \mathcal{C})$ going the other way. ☆

3.24 In the definition of a morphism of ringed spaces, the condition is that ϕ^* maps $\mathcal{O}_Y$ into $\mathcal{O}_X$. There is a very natural way to check this condition: if it is true locally, it holds globally:

Lemma 3.25 *If there is a basis* $\{U_i\}_{i \in I}$ *for the topology of* Y *such that* ϕ^* *takes* $\mathcal{O}_Y(U_i)$ *into* $\mathcal{O}_X(\phi^{-1}U_i)$*, then* ϕ^* *maps* $\mathcal{O}_Y$ *into* $\mathcal{O}_X$.

Proof: Let $U \subseteq X$ be any open, and take an element $f : U \to k$ in $\mathcal{O}_X(U)$. The pullback of f is the section $\phi^*(f) \in \mathcal{C}(\phi^{-1}U)$. Now, there is subset J of the index set I so that $\{U_j\}_{j \in J}$ is an open covering of U, and by assumption, $\phi^*(f|_{U_j})$ lies in $\mathcal{O}_Y(\phi^{-1}U_j)$. Since $\phi^*(f|_V) = \phi^*(f)|_{\phi^{-1}V}$ for each open V, these functions coincide on the intersections $\phi^{-1}U_j \cap \phi^{-1}U_{j'} = \phi^{-1}(U_j \cap U_{j'})$ of preimages of any two members of the covering $\{\phi^{-1}U_j\}_{j\in J}$ of $\phi^{-1}U$, and consequently they glue together to an element of $\mathcal{O}_Y(\phi^{-1}U)$. ❏

The sheaf of regular functions on algebraic sets

We are now ready to define *the sheaf $\mathcal{O}_X$ of regular functions* on X, with the assumption that X be an irreducible closed algebraic set in $\mathbb{A}^n$.

Ringed space
ringet rom (ringa rom)

DEFINITION *The* structure sheaf $\mathcal{O}_X$ *is the sheaf defined by*

$$\mathcal{O}_X(U) = \{\, f \in k(X) \mid f \text{ is regular in } U \,\} = \bigcap_{x\in U} \mathcal{O}_{X,x},$$

for each open set $U \subset X$.

All the rings $\mathcal{O}_X(U)$ are subrings of the function field $k(X)$; $\mathcal{O}_X(U)$ merely picks out which rational functions in $k(X)$ are regular in U. Moreover, when $U \subseteq V$ are two open subsets, the restriction map from V to U is just the inclusion $\bigcap_{x\in V} \mathcal{O}_{X,x} \subseteq \bigcap_{x\in U} \mathcal{O}_{X,x}$.

PROPOSITION 3.27 *Let X be an irreducible closed algebraic set. Then $\mathcal{O}_X$ is a sheaf.*

PROOF: We only need to verify the Gluing axiom. Assume first that f and g are regular on non-empty open subsets U and V respectively, and that $f|_{U\cap V} = g|_{U\cap V}$. Then $f = g$ as elements in $k(X)$ (because $U \cap V \neq \emptyset$). Next, let $\{U_i\}$ be a covering of U and assume given sections f_i of $\mathcal{O}_X$ over U_i coinciding on the pairwise intersections. Since all the intersections $U_j \cap U_i$ are non-empty, the f_i's all correspond to the same element $f \in k(X)$, and since the U_i's cover U, that element is regular in U. ❏

3.28 Notice that Proposition 3.7 on page 49 when interpreted in the context of sheaves, says that the global sections of the structure sheaf $\mathcal{O}_X$ is the coordinate ring $A(X)$; in other words, one has $\mathcal{O}_X(X) = A(X)$. In particular, when $X = \mathbb{A}^m$, one has $\mathcal{O}_{\mathbb{A}^m}(\mathbb{A}^m) = k[x_1, \ldots, x_m]$.

More generally, Proposition 3.7 implies the following:

PROPOSITION 3.29 *For an irreducible closed algebraic set X, we have*

i) *For each $f \in A(X)$, we have*

$$\mathcal{O}_X(D(f)) = A(X)_f$$

ii) *For each $p \in X$, we have*

$$\mathcal{O}_{X,p} = A(X)_{\mathfrak{m}_p}$$

EXERCISE 3.3 Let $X = \mathbb{R}^n$ with the usual topology, and let $\mathcal{O}_X(U)$ denote the subring of $\mathcal{C}(U)$ consisting of the *bounded continuous functions*. Show that $\mathcal{O}_X$ is not a sheaf. ★

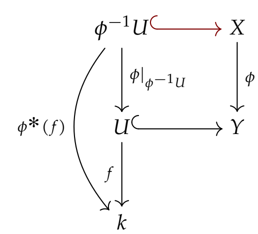
3.3 *Towards the definition of a variety*

In this section we introduce the main objects of study in this course, namely the varieties. We begin by defining what an affine variety is, and subsequently the affine varieties will serve as building blocks for the general varieties. The general definition may appear rather theoretical, but soon, when we come to projective varieties, there will be many examples illustrating why it is necessary and how it works in practice.

Affine varieties

Here come the definition of an affine variety:

DEFINITION *An* affine variety *is a ringed space $(X, \mathcal{O}_X)$ which is isomorphic (as a ringed space) to an irreducible closed subset of some $\mathbb{A}^n$ equipped with its sheaf of regular functions.*



In more detail, this means that there is an irreducible algebraic set $X_0 \subseteq \mathbb{A}^n$ and a homeomorphism $\phi\colon X \to X_0$, so that the map $\phi^*\colon \mathcal{C}_{X_0} \to \mathcal{C}_X$ induces an isomorphism between $\mathcal{O}_X$ and $\mathcal{O}_{X_0}$. Thus for all open subsets $U \subseteq X_0$, the map ϕ^* takes $\mathcal{O}_{X_0}(U)$ isomorphically into $\mathcal{O}_X(\phi^{-1}(U))$.

3.31 It might be tempting to define affine varieties as irreducible closed algebraic subsets — after all, these already come equipped with a sheaf of regular functions. However, there are certain advantages with including the word 'isomorphism' into the definition, which will be clear as affine varieties will form open coverings of general vartieties. For instance, the distinguished open subsets $D(f)$ of an algebraic set X are *a priori* open subsets of X, so they are *per se* not closed algebraic sets. However, they are naturally ringed spaces, endowed with the restriction $\mathcal{O}_X|_{D(f)}$ of the sheaf of regular functions as sheaf of rings, and the next proposition tells us that they in fact turn out to be affine varieties:

Proposition 3.32 *Let X be an irreducible closed algebraic set and let $f \in A(X)$. Then the pair $(D(f), \mathcal{O}_X|_{D(f)})$ is an affine variety.*

Proof: We need to exhibit a closed algebraic set W and a homeomorphism $\phi\colon W \to D(f)$ inducing an isomorphism between the sheaves of rings.

To this end, assume that $X = Z(\mathfrak{a})$ for an ideal $\mathfrak{a} \subseteq k[x_1, \ldots, x_n]$. The desired set W will be the closed algebraic subset $W = Z(\mathfrak{b}) \subseteq \mathbb{A}^n \times \mathbb{A}^1 = \mathbb{A}^{n+1}$ defined by the following ideal[4]:

Isomorphisms
isomorfier

$$\mathfrak{b} = \mathfrak{a}k[x_1, \ldots, x_{n+1}] + (1 - f \cdot x_{n+1}).$$

The subset W is contained in inverse image $X \times \mathbb{A}^1$ of X under the projection onto $\mathbb{A}^n$, and consists of those points where $x_{n+1} = 1/f(x_1, \ldots, x_n)$. We let ϕ denote the restriction of the projection to W. This is a homeomorphism onto $D(f)$: the inverse is given by the map α sending $(x_1, \ldots, x_n)$ to $(x_1, \ldots, x_n, 1/f(x_1, \ldots, x_n))$.

The salient point is that ϕ^* and α^* are mutually inverse homomorphisms between $A(W)$ and $\mathcal{O}_{D(f)}(D(f))$, and as α and ϕ are mutually inverse, this amounts to verifying that $A(W)$ and $\mathcal{O}_{D(f)}(D(f))$ are mapped into each other.

The sheaf $\mathcal{O}_X$ of regular functions
knippet av regulære funksjoner

A regular function g on W is a polynomial in the coordinates $x_1, \ldots, x_{n+1}$, and substituting $1/f(x_1, \ldots, x_n)$ for x_{n+1}, gives a regular function on $D(f)$ since $\mathcal{O}_{D(f)}(D(f)) = A(X)_f$ (this is Proposition 3.7 on page 49). So α^* takes $A(W)$ into $\mathcal{O}_{D(f)}(D(f))$.

Affine varieties
affine varieteter

Similarly, if g is regular on $D(f)$, it is expressible in the form a/f^r where a is a polynomial in $x_1, \ldots, x_n$, and therefore $g \circ \phi$ is regular on W; indeed, it holds true that

$$a(\phi(x_1, \ldots, x_{n+1}))/f(\phi(x_1, \ldots, x_{n+1}))^r = a(x_1, \ldots, x_n)x_{n+1}^r.$$

To finish the proof, we have to show that ϕ^* takes the sheaf of rings $\mathcal{O}_{D(f)}$ into the sheaf of regular functions $\mathcal{O}_W$, but because of Lemma 3.25 on page 53, it suffices to show that for any distinguished open set $D(g) \subseteq D(f)$ and any regular function h on $D(g)$, the composite $g \circ \phi$ is regular on $\phi^{-1}(D(g)) = W_{\phi^* g}$; but this is now obvious since ϕ^* is an isomorphism between $A(W)$ and $\mathcal{O}_{D(f)}(D(f))$, and $\mathcal{O}_{D(g)}(D(g)) = \big(\mathcal{O}_{D(f)}(D(f))\big)_g$ and $\mathcal{O}_{W_{\phi^* g}}(W_{\phi^* g}) = A(W)_{\phi^*(g)}$. ❏

3.33 The set W in the proof is nothing but the *graph* of the function $1/f$ embedded in $\mathbb{A}^{n+1}$. Two simple examples of the situation are depicted in the margin (Figure 3.2 and 3.3) in both cases $X = \mathbb{A}^1$. In the first figure the function f is given as $f(x) = x(x-1)(x-2)$, and in the second, $f(x) = x$. In the latter case $D(f) = \mathbb{A}^1 \backslash \{0\}$, and W is the hyperbola $xy = 1$.

We have a corresponding statement for irreducible closed sets:

Proposition 3.34 *Let X be an affine variety and let $Z \subset X$ be an irreducible closed set. Then Z is an affine variety.*

Proof: As X is an affine variety, there is an isomorphism ι from X onto an irreducible closed set $X_0 \subset \mathbb{A}^n$. The map ι restricts to an isomorphism from Z onto an irreducible closed subset $W \subset X_0$, which must itself be a closed irreducible subset of $\mathbb{A}^n$. Thus Z is an affine variety, being isomorphic to $(W, \mathcal{O}_W)$. ❑

Prevarieties

Before we can define varieties, we will need the more general notion of *prevarieties*.

> Definition *A* prevariety *is an irreducible ringed space so that there is a finite open covering* $\{X_i\}$ *of* X *with each* $(X_i, \mathcal{O}_X|_{X_i})$ *being an affine variety.*

A *morphism* $\phi : X \to Y$ between two prevarieties is just a morphism of the underlying ringed spaces, *i.e.* we require $f \circ \phi$ to be regular on $\phi^{-1}(U)$ for each $f \in \mathcal{O}_Y(U)$. And, naturally, ϕ is an *isomorphism* if there is a morphism $\psi\colon Y \to X$ such $\phi \circ \psi = \mathrm{id}_X$ an $\psi \circ \phi = \mathrm{id}_Y$.

The sheaf $\mathcal{O}_X$ is called *the structure sheaf* of X, and the sections of $\mathcal{O}_X$ over an open subset U are called *regular functions* on U.

If $f \in \mathcal{O}(U)$ is a regular function, we write $D(f)$ or U_f for the *distinguished open set* $\{\, x \in U \mid f(x) \neq 0 \,\}$.

> Definition *For a prevariety* X*, we define its* function field $k(X)$ *to be the function field* $K = k(U)$ *of any affine open subset* $U \subset X$.

Note that this is well-defined: Given two open affine subsets $U, V \subset X$ the function fields $k(U)$ and $k(V)$ are canonically isomorphic (they both equal $k(W)$ where $W \subset U \cap V$ is a common affine subset. As before, we regard elements of $k(X)$ as rational functions; that is functions that are regular on some open subset of X.

Being a morphism is a *local property* of a continuous map $\phi\colon X \to Y$ between two prevarieties; that is, one can check it being a morphism on appropriate open coverings. One has:

Lemma 3.37 *Let* X *and* Y *be two prevarieties and let* $\phi\colon X \to Y$ *be a continuous map. Suppose one can find open coverings* $\{U_i\}_{i \in I}$ *and* $\{V_i\}_{i \in I}$ *of respectively* X *and* Y *such that* ϕ *maps each* U_i *into* V_i*, and such that* $\phi|_{U_i}$ *is a morphism between* U_i *and* V_i*, then* ϕ *is a morphism.*

Proof: Let f be a regular function on some open $V \subseteq Y$ and let $U \subseteq X$ be an open subset with $\phi(U) \subseteq V$. To see that $f \circ \phi|_U$ is regular in U, it suffices by the patching property of sheaves, to show that its restriction to each $U_i \cap U$ is regular. But $U_i \cap U$ maps into V_i, and by hypothesis $f \circ \phi|_{U_i}$ is regular, and

$$\begin{array}{ccc} \mathcal{C}_{X_0}(U) & \xrightarrow[\simeq]{\phi^*} & \mathcal{C}_X(\phi^{-1}(U)) \\ \uparrow & & \uparrow \\ \mathcal{O}_{X_0}(U) & \xrightarrow[\simeq]{} & \mathcal{O}_X(\phi^{-1}(U)) \end{array}$$

[4] We already came across a similar ideal when performing the Rabinowitsch trick, except that in the present situation f does not belong to $\mathfrak{a}$.

$$\begin{array}{ccccc} W & & \hookrightarrow & & \mathbb{A}^{n+1} \\ \alpha\uparrow\downarrow\phi & & & & \downarrow \\ D(f) & \hookrightarrow & X & \hookrightarrow & \mathbb{A}^n \end{array}$$

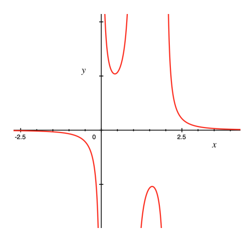


Figure 3.2: The set W is the graph of the function $1/f$ with $f(x) = x(x-1)(x-2)$.

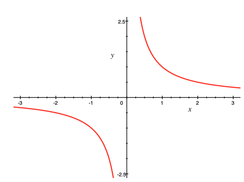


Figure 3.3: Projections onto the x-axis, makes the hyperbola $xy = 1$ is isomorphic to $\mathbb{A}^1 \backslash \{0\}$.

Prevarieties
prevarieteter

because restrictions of regular functions are regular, it follows that $f \circ \phi|_{U_i \cap U}$ is regular. ❑

Exercises

3.4 Show that the composition of two composable morphisms is a morphism. Show that morphisms having $\mathbb{A}^1$ as target are just the regular functions.

3.5 Let f be a regular function without zeros on the prevariety X. Show that $1/f$ is a regular function.

3.6 Show that if one replaces the word morphism by the word isomorphism throughout Lemma 3.37, the lemma stays true.

★

Open and closed sub(pre)varieties

3.38 Assume $U \subseteq X$ is an open subset of a prevariety X. We may endow U with the restriction of the structure sheaf $\mathcal{O}_X$ to U; that is, we put $\mathcal{O}_U = \mathcal{O}_X|_U$. Then $(U, \mathcal{O}_U)$ will be a prevariety. In fact, this follows from the slightly more general statement:

Proposition 3.39 *A prevariety X has a basis for the topology consisting of open affine subsets.*

Proof: Let $\{X_i\}$ be any open affine covering of X. If $U \subseteq X$ is an open subset of X, the sets $U_i = U \cap X_i$ form an open covering of U. The U_i's will not necessarily be affine, but we know that the distinguished open sets in X_i form a basis for its topology, and by Proposition 3.32 on page 56 above they are affine varieties. Hence we can cover each of the U_i's, and thereby U, by affine opens. ❑

3.40 Thus open subsets in a prevariety inherit a structure of prevariety from the surrounding space. The same hods true for irreducible closed subsets, however this is slightly more subtle.

Let X be a prevariety and let $Z \subset X$ be an irreducible closed subset. Of course, if $\{X_i\}$ is an open cover of X by affine varieties, the intersections $X_i \cap Z$ form an open cover of Z, and according to Proposition 3.34 they are affine varieties. The delicate point is to define the sheaf of rings $\mathcal{O}_Z$ on Z. For an open set $U \subset Z$, we define $\mathcal{O}_Z(U)$ as the set of elements $f : U \to k$ that are locally restrictions of a regular function $X \to k$. More precisely: for every point $p \in U$ there is an open neighbourhood V of p in X and a regular function g_V on V such that $g_V|_{V \cap Z} = f|_{V \cap Z}$. We leave it to the reader to check that the sheaf axioms are satisfied, so we obtain sheaf of rings $\mathcal{O}_Z$ on Z.

PROPOSITION 3.41 (CLOSED SUB-PREVARIETIES) *Assume that $Z \subseteq X$ is an irreducible closed subset of the prevariety X. Endowed with the sheaf of rings $\mathcal{O}_Z$ just defined, Z becomes a prevariety.*

PROOF: Most has been proven; the only thing still to be verified is that when $\mathcal{O}_Z$ is restricted to the affines $V_i = X_i \cap Z$, it becomes isomorphic with the sheaf $\mathcal{O}_{V_i}$ of regular functions on V_i. It suffices to check this statement on each distinguished open set $D(g) \subseteq V_i$. For those it is clear from Lemma 3.7 on page 49. ❑

EXERCISE 3.7 The x-axis with the origin deleted is a closed and irreducible subset Z of $\mathbb{A}^2 \backslash \{(0,0)\}$. Exhibit regular functions on Z that are not restrictions of regular functions on $\mathbb{A}^2 \backslash \{(0,0)\}$. This illustrates why the definition of $\mathcal{O}_Z$ is delicate; functions can locally be extended to ambient space without being globally extendable. HINT: Regular functions on then punctured affine plane $\mathbb{A}^2 \backslash \{(0,0)\}$ are polynomials (see Example 3.11 on page 50). ★

EXERCISE 3.8 Let X be a prevariety and let $Y \subseteq X$ be a closed irreducible subset. For any open $U \subseteq X$, let $\mathcal{I}_Y(U)$ be the subset of regular functions on U that vanish on $Y \cap U$. Show that $\mathcal{I}_Y(U)$ is an *ideal* in $\mathcal{O}_X(U)$. Show that if $V \subseteq U$ are two open sets, then ρ_{UV} takes $\mathcal{I}_Y(U)$ into $\mathcal{I}_Y(V)$. Show that $\mathcal{I}_Y$ is a sheaf (of abelian groups, or if you want, a sheaf of rings without unit element). ★

Maps into affine space

An important fact in algebraic geometry is that morphisms having an affine variety as target can be described by regular functions on the source. In the particular case that the target is the affine space $\mathbb{A}^n$ itself, such morphisms are naturally determined by the n components. This paves the way for describing maps into any affine variety $Y \subseteq \mathbb{A}^n$ – the components have to satisfy all the defining equations of Y – and this leads to the main theorem of the section (Theorem 3.45).

3.42 Suppose we are given a prevariety X and a collection $f_1, \ldots, f_n$ of regular functions on X. Letting the f_i's serve as component functions, one builds a mapping

$$\begin{aligned} \phi \colon X &\to \mathbb{A}^n \\ x &\mapsto (f_1(x), \ldots, f_n(x)) \end{aligned}$$

The map ϕ is obviously continuous, and as would be expected, it is a *morphism*. Indeed, since being a morphism is a local property (Lemma 3.37 above), it suffices to check the defining property on the distinguished open subsets of $\mathbb{A}^n$. So let $U = D(b) \subseteq \mathbb{A}^n$ be one. Regular functions on U are (Proposition 3.7 on page 49) of the form $g = a/b^r$ where a is a polynomial and r a non-negative integer. For points x in the inverse image $\phi^{-1}(U)$, it holds true that $b(f_1(x), \ldots, f_n(x)) \neq 0$, hence

$$(g \circ \phi)(x) = a(f_1(x), \ldots, f_n(x))/b(f_1(x), \ldots, f_n(x))^r$$

is regular in $\phi^{-1}(U)$.

On the other hand, if $\phi\colon X \to \mathbb{A}^n$ is a morphism, the component functions f_i of ϕ being the compositions $f_i = x_i \circ \phi$ of the morphism ϕ with the coordinate functions, are morphisms from X to $\mathbb{A}^1$. Hence we have proved that prevarieties have the very natural property that morphisms into the affine n-space $\mathbb{A}^n$ are determined by giving regular component functions:

Proposition 3.43 (Morphisms into $\mathbb{A}^n$) *Assume that X is a prevariety. Sending ϕ to ϕ^* sets up a one-to-one correspondence between morphisms $\phi\colon X \to \mathbb{A}^n$ and k-algebra homomorphisms $\phi^*\colon k[x_1, \ldots, x_n] \to \mathcal{O}_X(X)$.*

3.44 The Proposition 3.43 has an immediate generalization. We may replace the affine n-space with any affine variety:

Theorem 3.45 (Morphisms into affine varieties) *Assume that X is a prevariety and Y an affine variety. The assignment $\phi \mapsto \phi^*$ sets up a one-to-one correspondence between morphisms $\phi\colon X \to Y$ and k-algebra homomorphisms $\phi^*\colon A(Y) \to \mathcal{O}_X(X)$.*

Proof: Suppose that $Y \subseteq \mathbb{A}^n$. Giving a morphism $\phi\colon X \to Y$ is the same as giving a morphism $\phi\colon X \to \mathbb{A}^n$ that factors through Y. By Proposition 3.43 above, giving a $\phi\colon X \to \mathbb{A}^n$ is the same as giving the algebra homomorphisms $\phi^*\colon k[x_1, \ldots, x_n] \to \mathcal{O}_X(X)$, and ϕ takes values in Y if and only if $f(\phi(x)) = 0$ for all polynomials $f \in I(Y)$; that is, the composition map ϕ^* vanishes on the ideal $I(Y)$. Hence ϕ takes values in Y if and only if ϕ^* factors through the quotient $A(Y) = k[x_1, \ldots, x_n]/I(Y)$. ❑

Example 3.46 The ring of regular functions on a prevariety X is a finitely generated k-algebra, and is therefore the coordinate ring of an *affine* variety X_{aff}. By Theorem 3.45 above, there is a morphism $X \to X_{\text{aff}}$ that induces the identity on the ring of regular functions. ☆

3.47 Specializing the prevariety X to be affine as well, we get the following corollary, the main theorem for morphism of affine varieties:

Theorem 3.48 (Main theorem for affine varieties) *Assume that X and Y are two affine varieties. Then $\phi \mapsto \phi^*$ is a one-to-one correspondence between morphisms $\phi\colon X \to Y$ and k-algebra homomorphisms $\phi^*\colon A(Y) \to A(X)$.*

An immediate corollary is the following:

THEOREM 3.49 *Let X and Y be two affine varieties and $\phi\colon X \to Y$ a morphism. Then ϕ is an isomorphism if and only if ϕ^* is an isomorphism. In particular, X and Y are isomorphic if and only if $A(X)$ and $A(Y)$ are isomorphic as k-algebras.*

A priori there could be many ways of viewing an affine variety X as a closed algebraic set in some $\mathbb{A}^n$, thus there are a priori many different coordinate rings $A(X) = k[x_1,\ldots,x_n]/I(X)$. However, according to the theorem, all these in fact give the same coordinate ring, namely $A(X) = \mathcal{O}_X(X)$. In other words, $A(X)$ is an intrinsic invariant of X which does not depend on any chosen embeddings into affine space.

EXAMPLE 3.50 Consider the morphism $\phi : \mathbb{A}^1 \to \mathbb{A}^3$ given by $t \mapsto (t, t^2, t^3)$, which defines the affine twisted cubic $C \subset \mathbb{A}^3$. If x, y, z are the usual coordinates on $\mathbb{A}^3$, we have $\phi^*x = t$, $\phi^*y = t^2$ and $\phi^*z = t^3$. Thus ϕ corresponds to the ring map

$$\begin{aligned} \phi^* : k[x,y,z] &\quad\to\quad k[t] \\ x &\mapsto t \\ y &\mapsto t^2 \\ z &\mapsto t^3 \end{aligned}$$

Note also that the coordinate ring of the twisted cubic, $A(C) = k[x,y,z]/(y - x^2, z - x^3)$ is isomorphic to $k[t]$, the cooridnate ring of $\mathbb{A}^1$. ☆

EXAMPLE 3.51 *(A non-affine prevariety)* The prevariety $\mathbb{A}^n \backslash \{0\}$ is not affine if $n \geqslant 2$. Indeed, by Example 3.11 on page 50 the inclusion $\iota\colon \mathbb{A}^n \backslash \{0\} \to \mathbb{A}^n$ induces an isomorphism i^* between the spaces of global sections of the two structure sheaves. If $\mathbb{A}^n \backslash \{0\}$ were affine, the inclusion would therefore be an isomorphism after Theorem 3.49 above, but this is of course not the case. ☆

3.52 In categorical terms, Proposition 3.45 above says in view of exercise 2.24 on page 46, that the category of affine varieties is equivalent to the category of finitely generated k-algebras that are integral domains. So the study of the affine varieties is in this sense equivalent to the study of domains finitely generated as k-algebras.

3.53 As an example of applications of the diverse "morphism-theorems" in the previous paragraphs, we prove a criterion for a prevariety to be affine: it should be contrasted with Example 3.51 above.

PROPOSITION 3.54 *Let X be a prevariety and assume that there are regular functions $f_1, \ldots, f_r$ on X that generate the unit ideal in $\mathcal{O}_X(X)$. Assume also that each X_f is affine. Then X is affine.*

PROOF: Let $A = \mathcal{O}_X(X)$, and let f be a regular function on X. Recall the open set X_f where f does not vanish. We claim that $\mathcal{O}_X(X_f) = A_f$. The inclusion $A_f \subseteq \mathcal{O}_X(X_f)$ is clear. So assume that g is a regular function on X_f, we need to show that for some $n \in \mathbb{N}$ the function $f^n \cdot g$ is regular on X. Since X is cover by a finite number of open affine subsets, it suffices to see that for each open affine U of X, it holds that $f^n \cdot g|_{X_f \cap U}$ is regular on U for some $n \in \mathbb{N}$. But this is clear since $X_f \cap U = D(f|_U)$ so that $\mathcal{O}_X(X_f \cap U) = \mathcal{O}_X(U)_f$ (after Proposition 3.32 on page 56).

Consider then the map $\phi\colon X \to X_{\text{aff}}$ as in Example 3.46 above, where X_{aff} is the affine variety whose coordinate ring equals A. Restricted to an affine $X_f \subseteq X$, the motphism ϕ induces an isomorphism bewteen X_f and $(X_{\text{aff}})_f$ since both have A_f as coordinate ring. The X_{f_i}'s cover X by hypothesis, and since the f_i's generate the unit ideal in A, they cover X_{aff} as well. It follows that ϕ is an isomorphism. ❏

3.4 Varieties

The Hausdorff axiom is the third and final axiom required of varieties. Zariski topologies are, as we have seen, far from being Hausdorff, but some properties[5] of Hausdorff spaces can be salvaged by this third axiom. In some sense the third axiom is a remedy for the topologies not being Hausdorff.

Morphism of prevarieties
morfi mellom prevariteter

3.55

DEFINITION *A prevariety $(X, \mathcal{O}_X)$ is called a* variety *if the following condition is fulfilled.*

- *For any two morphisms $\phi, \psi\colon Z \to X$ whose source Z is a prevariety, the set of points where ϕ and ψ coincide is closed; that is, the subset*

$$\{ x \in Z \mid \phi(x) = \psi(x) \}$$

is closed.

Of course, one may as well require the set of points where ϕ and ψ assumes distinct values to be open. The subset $\{ x \in Z \mid \phi(x) = \psi(x) \}$ is frequently called the *equalizer* of ϕ and ψ.

Isomorphism of prevarities
isomorfi av prevariteter

3.57 The first thing to check, is that affine varieties deserve having the term variety as part of their name:

PROPOSITION 3.58 *Any affine variety is a variety.*

PROOF: To begin with, observe that if f and g are two regular functions on a prevariety Z, the set where they coincide is closed. Indeed, the diagonal in $\mathbb{A}^1 \times \mathbb{A}^1 = \mathbb{A}^2$ being the zero locus of the polynomial $x_1 - x_2$ is closed, and the map $Z \to \mathbb{A}^2$ given as $x \mapsto (f(x), g(x))$ is continuous. Since preimages of closed sets by continuous maps are closed, it follows that $\{ x \mid f(x) = g(x) \}$ is closed.

Next, let $X \subseteq \mathbb{A}^m$ be affine, and assume that ϕ and ψ map Z into X. If $y_1, \ldots, y_m$ denote the coordinate functions on $\mathbb{A}^m$, the compositions $y_i \circ \psi$ and $y_i \circ \phi$ are regular functions on Z. The set where ϕ and ψ coincide is the intersection of the sets where each pair $y_i \circ \psi$ and $y_i \circ \phi$ coincide, and by the initial observations each of these subsets is closed, hence their intersection is closed. ❑

The following criterion is often useful for proving that a prevariety is a variety:

Lemma 3.59 *Assume that X is a prevariety such any two different points are contained in a common open affine subset. Then X is a variety.*

Proof: Let Z be a prevariety and ϕ and ψ two maps from Z to X. Let $x \in Z$ be a point such that $\phi(x) \neq \psi(x)$. By assumption there is an open affine set U in X containing both $\phi(x)$ and $\psi(x)$, and then $V = \phi^{-1}(U) \cap \psi^{-1}(U)$ is an open set in Z in which x lies. Now, U being affine is a variety by Proposition 3.58 above, and consequently the set $W \subseteq V$ where the two maps ψ and ϕ coincide is closed; but this means that $V \backslash W$ is an open set in Z containing x that is entirely contained in the set $\{ z \in Z \mid \phi(z) \neq \psi(z) \}$. It follows that $\{ z \in Z \mid \phi(z) \neq \psi(z) \}$ is open since x was an arbitrary point. ❑

Structure sheaves
strukturknipper

Example 3.60 *(A prevariety which is not a variety)* This is an example of a prevariety X for which the Hausdorff axiom is not satisfied, so X is a prevatiety that is not a variety. These "*non-separated prevarieties*", as they often are called, exist on the fringe of the algebro-geometric world, you would very seldom meet them – although now and then they materialize from the darkness and serve useful purposes. Anyhow, this is the only place in this course where such a creature will appear, and the only reason to include it, is to convince you that the Hausdorff axiom is needed.

The intuitive way to think about X is as an affine line with "two origins". It does not carry enough functions that the two origins can be separated – if a function is regular in one, it is regular in the other and assumes the same value.

Regular functions
regulære funksjoner

The underlying topological space is the set $X = (\mathbb{A}^1 \backslash \{0\}) \cup \{0_1, 0_2\}$ endowed with the topology of finite complements. It has two copies of the affine line $\mathbb{A}^1$ lying within it; either with one of the twin origins as origin; that is, $A_1 = (\mathbb{A}^1 \backslash \{0\}) \cup \{0_1\}$ and $A_2 = (\mathbb{A}^1 \backslash \{0\}) \cup \{0_2\}$. Both these sets are open sets and their intersection A is given as $A = A_1 \cap A_2 = \mathbb{A}^1 \backslash \{0\}$. Obviously, the Hausdorff axiom is not satisfied, because the two inclusions of $\mathbb{A}^1$ in X are equal on $\mathbb{A}^1 \backslash \{0\}$ which is not closed.

To tell what regular functions X carries, let $U \subseteq X$ be any open subset, and define $\mathcal{O}_X(U)$ to be the set of rational functions $a(x)/b(x)$ in one variable x with $b(x) \neq 0$ for all $x \in U \cap A_1 \backslash \{0_1, 0_2\}$, and if either of the twin origins, or both, belong to U we require aditionally that $b(0) \neq 0$. Letting the restriction maps be the obvious ones, we obtain a subprsheaf of $\mathcal{C}_X$ which is a sheaf since lying in $\mathcal{O}_X$ is a local condition.

With this definition, the subsets $U\backslash\{0_1\}$ and $U\backslash\{0_2\}$ carry the *same* regular functions whenever U is an open set containing both 0_1 and 0_2.

Now, the salient point of the example is that this definition, in spite of being rather peculiar, really gives X the structure of a prevariety. Indeed, the two copies A_1 and A_2 of $\mathbb{A}^1$ form an affine open cover of X, and by construction, it holds true that $\mathcal{O}_X|_{\mathbb{A}_i} = \mathcal{O}_{\mathbb{A}^1}$. ☆

Exercises

3.9 *(For fringy people.)* Let X be any closed algebraic set and let $Y \subseteq X$ be a proper closed subset. Construct a prevariety $X_{\amalg}$ containing unseparable twin copies of Y and two different open subsets both isomorphic to X that intersect along $X\backslash Y$.

3.10 Mimic the construction of "the bad guy" above with $\mathbb{A}^2$ and the origin to get an "even worse guy" X. Exhibit two affine open subsets of X whose intersection equals $\mathbb{A}^2\backslash\{0\}$. Conclude that the intersection of two affine open subsets of a prevariety is not necessarily affine. This is contrary to what is true for varieties, and shows that the hypothesis that X be a variety in Proposition 3.78 on page 70 can not be skipped. HINT: Example 3.11 might be of interrest.

3.11 Formulate a universal property that characterizes the equilizer of two morphisms, meaningful in any category.

★

3.5 Products of varieties

In this section, we will describe the construction of the product of two (pre)varieties X and Y. The product will be characterized by a certain universal property, which has a meaningful formulation in any category, and in this generality it is not a priori that the product exists. Nevertheless, to show that products exists in the category of varieties, we will define it first for affine varieties, then for general prevarieties, and then finally we show that the product $X \times Y$ is in fact a variety whenever X and Y are.

The universal property of a product

3.61 The product of two prevarieties X and Y consists of a prevariety $X \times Y$ together with two morphisms $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ called *the projections*, such that the three comply with the following condition:

Note that we are using that any open subset of an irreducible space is dense

❏ For any prevariety Z and any pair of morphisms $\phi_X\colon Z \to X$ and $\phi_Y\colon Z \to Y$, there is a morphism $\phi\colon Z \to X \times Y$ such that $\phi_X = \pi_X \circ \phi$ and $\phi_Y = \pi_Y \circ \phi$. Moreover, ϕ is uniquely defined by these conditions.

In a more laid-back language, the defining property is that giving a morphism into the product is the same thing as giving the two components.

As usual with objects defined by universal properties, the product is uniquely defined, if it exists:

Proposition 3.62 *The product is unique up to a unique isomorphism.*

Proof: Observe first that a morphism $\phi\colon X \times Y \to X \times Y$ such that $\pi_X \circ \phi = \pi_X$ and $\pi_Y \circ \phi = \pi_Y$ must, by the uniqueness part of the definition, be equal to the identity $\mathrm{id}_{X\times Y}$. Assume then that W and W' with projections π_X, π_Y and π'_X, π'_Y are two products of X and Y – *i.e.* they both have the universal property. By the existence part there is a unique morphism $\phi\colon W \to W'$ such that $\pi'_x \circ \phi = \pi_X$ and $\pi'_Y \circ \phi = \pi_Y$ and a unique morphism $\psi\colon W' \to W$ with $\pi_X \circ \psi = \pi'_X$ and $\pi_Y \circ \psi = \pi'_Y$.

From the observation at the top of the proof it ensues that $\psi \circ \phi = \mathrm{id}_{W'}$ and that $\phi \circ \psi = \mathrm{id}_W$. Indeed, by symmetry it suffices to check the first. One has $\pi_X \circ \psi \circ \phi = \pi'_X \circ \phi = \pi_X$ and ditto $\pi_Y \circ \psi \circ \phi = \pi_Y$, so we can conclude that $\psi \circ \phi = \mathrm{id}_{W'}$. ❏

[5] Properties expressed in terms of morphisms not in terms continuous maps.

3.63 The main theorem of this section follows next, and the proof occupies the rest this chapter. As already explained, the second part will be established first.

Theorem 3.64 (Existence of products) *Any two prevarieties X and Y has a product $X \times Y$. It is a prevariety whose underlying set is the Cartesian product of X and Y, and together with the two projections π_X and π_Y it satisfies the universal property.*

i) If X and Y are varieties, the product $X \times Y$ is a variety.

ii) When X and Y are affine varieties, the product $X \times Y$ will be affine, and it holds true that the coordinate ring is given as $A(X \times Y) = A(X) \otimes_k A(Y)$.

The product of affine varieties

We start out by proving that products exist in the category of affine varieties. The first step is to show that we have products in the category of closed algebraic sets and polynomial maps, and subsequently that these turn out to be irreducible – that is, affine varieties – when both factors are irreducible.

3.65 Let $X \subseteq \mathbb{A}^n$ and $Y \subseteq \mathbb{A}^m$ be two closed algebraic sets, and choose coordinates $x_1, \ldots, x_n$ in $\mathbb{A}^n$ and $y_1, \ldots, y_m$ in $\mathbb{A}^m$. The product of X and Y will be the closed algebraic subset of $\mathbb{A}^{n+m}$ given by the ideal $\mathfrak{a}$ in the polynomial ring $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ generated by the ideals $I(X)$ and $I(Y)$ of polynomials vanishing on respectively X and Y. That is, we put

$$\mathfrak{a} = (f_1(x), \ldots, f_r(x), g_1(y), \ldots, g_s(y)),$$

where the f_i's are generators for $I(X)$ and the g_j's for $I(Y)$. Moreover, we shall temporarily use the notation $W = Z(\mathfrak{a})$. Notice that by the general theory of tensor products of algebras, it holds true that

$$k[x_1, \ldots, x_n, y_1, \ldots, y_m]/\mathfrak{a} \simeq A(X) \otimes_k A(Y),$$

which comes close to the statement about coordinate rings in the theorem; it remains to be seen that $\mathfrak{a}$ is a radical ideal and that $\mathfrak{a}$ is prime when $I(X)$ and $I(Y)$ both are prime[6].

Varieties
varieteter

For the moment W is just a closed algebraic subset, and in the final step of the construction, which is the hardest part, it will turn out to be an affine variety when X and Y are.

The first, and easy, step of the construction is to show that the subset W of $\mathbb{A}^{n+m}$ satisfies the universal property among closed algebraic sets and polynomial maps.

The two projection π_X and π_Y are induced from the natural linear projections p_n and p_m mapping $\mathbb{A}^{n+m}$ onto $\mathbb{A}^n$ and $\mathbb{A}^m$ respectively. They clearly send points in W to respectively $X \subseteq \mathbb{A}^n$ and $Y \subseteq \mathbb{A}^m$.

Lemma 3.66 *The subset W has the Cartesian product $X \times Y$ as underlying set, and together with the two maps $\pi_X = p_n|_W$ and $\pi_Y = p_m|_W$ satisfies the universal property of a product in the category of closed algebraic sets and polynomial maps.*

Proof: If the point $(x_1, \ldots, x_r, y_1, \ldots, y_s)$ in $\mathbb{A}^{n+m}$ belongs to W, it holds true that $f_i(x_1, \ldots, x_r) = 0$ for all i and $g_j(y_1, \ldots, y_s) = 0$ for all j by the definition of the ideal $\mathfrak{a}$. Hence $(x_1, \ldots, x_r) \in X$ and $(y_1, \ldots, y_s) \in Y$, and we can conclude that W coincides with the Cartesian product $X \times Y$.

Given two polynomial maps ϕ_X and ϕ_Y from a closed algebraic set Z into respectively X and Y. Since $X \subseteq \mathbb{A}^n$ and $Y \subseteq \mathbb{A}^m$, the two maps take values in $\mathbb{A}^n$ and $\mathbb{A}^m$, and consequently the map $\phi\colon Z \to \mathbb{A}^{n+m}$ defined by $z \mapsto (\phi_X(z), \phi_Y(z))$ is a polynomial map $Z \to \mathbb{A}^{n+m}$ with values in W. One easily convinces oneself that it solves the universal problem; indeed, by definition $\pi_X \circ \phi = \phi_X$ and $\pi_Y \circ \phi = \phi_Y$. ❑

3.67 The next step is to establish that W is irreducible whenever both X and Y, and that it serves as the product of X and Y in the category of prevarieties. The first lemma, about W being irreducible, holds for a large class of topologies on the Cartesian product $X \times Y$; the salient hypothesis being that all the sets $\{x\} \times Y$ and $X \times \{y\}$ are closed.

Lemma 3.68 *Assume that X and Y are two irreducible topological spaces. Assume $X \times Y$ is equipped with a topology such that all sets of the form $\{x\} \times Y$ and $X \times \{y\}$ are closed and irreducible. Then $X \times Y$ is irreducible as well.*

Proof: Assume the product $X \times Y$ can be expressed as a union $X \times Y = Z_1 \cup Z_2$ of two closed subsets Z_1 and Z_2. Let $X_i = \{\, x \in X \mid \{x\} \times Y \subseteq Z_i \,\}$. For each $y \in Y$ let $\iota_y \colon X \to X \times Y$ be the inclusion of $X \times \{y\}$ into $X \times Y$, which is continuous since $X \times \{y\}$ is closed (and hence inverse images of closed sets will be closed). It holds true that $X_i = \bigcap_{y \in Y} \iota_y^{-1}(X \times \{y\} \cap Z_i)$, and consequently the X_i's are both closed sets.

For every $x \in X$ the set $\{x\} \times Y$ is contained in either Z_1 or Z_2 since Y is irreducible, and it follows that $X = X_1 \cup X_2$. Now, X is assumed to irreducible, so either it holds that $X_1 = X$, and therefore that $Z_1 = X \times Y$, or $X_2 = X$, and $Z_2 = X \times Y$. ❑

3.69 We now know that W is irreducible and may hence infer from Lemma 3.66 above that W is the product of X and Y in the category of affine varieties. A straightforward gluing argument applied to maps into W, extends this to the category of all prevarieties, and shows that W, indeed, is the product of X and Y in the category of prevarieties:

Lemma 3.70 *The set W together with the projections $\pi_X = p_n|_W$ and $\pi_Y = p_m|_W$ is the product of X and Y in the category of prevarieties.*

Proof: As already observed, the closed algebraic set W is irreducible and it merely remains to establish the universal property.

Given two morphisms ϕ_X and ϕ_Y from a prevariety Z into respectively X and Y. Cover Z by open affine sets Z_i and consider the restrictions $\phi_X|_{Z_i}$ and $\phi_Y|_{Z_i}$. Since W satisfies the universal property among affine varieties, they give rise to morphisms $\phi_i \colon Z_i \to W$ such that $\pi_X \circ \phi_i = \phi_X|_{Z_i}$ and $\pi_Y \circ \phi_i = \phi_Y|_{Z_i}$.

On the intersections $Z_{ij} = Z_i \cap Z_j$ the morphism ϕ_i and ϕ_j must agree; indeed, both are solutions to the universal problem posed by the morphism $\phi_X|_{Z_{ij}}$ and $\phi_Y|_{Zij}$, and this solution being unique, it holds that $\phi_i|_{Z_{ij}} = \phi_j|_{Z_{ij}}$.

The different ϕ_i's therefore patch together, and we obtain a morphism ϕ with the requested property that $\pi_X \circ \phi = \phi_X$ and $\pi_Y \circ \phi = \phi_Y$. ❑

3.71 The last thing to establish about the affine products is that the coordinate rings are as announced:

Lemma 3.72 *In the present setting $A(X \times Y) = A(X) \otimes_k A(Y)$.*

Proof: Any element f in the tensor product $A(X) \otimes_k A(Y)$ can be represented as a finite sum $f = \sum g_i \otimes h_i$ of decomposable tensors where $g_i \in A(X)$ and $h_i \in A(Y)$, and we may assume that the h_i's are linearly independent over k; indeed, this will be the case if we use a sum with a minimal number of terms.

Assume that f is nilpotent and fix a point x_0 in X. Considered as a function of y, the element $f(x_0, y) \in A(Y)$ will be nilpotent, and hence $f(x_0, y) =$

$\sum_i g_i(x_0)h_i(y) = 0$. Since the h_i's are linearly independent, it follows that $g_i(x_0) = 0$ for all i. Now, the point x_0 was an arbitrary point in X so that $g_i = 0$ as functions on X, and we are done. ❑

EXERCISE 3.12 Lemma 3.72 is a special case of the following result from commutative algebra. If A and B are two reduced k-algebras finitely generated over the algebraically closed field k, then $A \otimes_k B$ is reduced. Show this. HINT: Adapt the proof of lemma 3.72. ★

Products of prevarieties

Let X and Y be two prevarieties. We shall work with affine open coverings $\{X_i\}$ and $\{Y_i\}$ of respectively X and Y.

The definition of the product as a prevariety requires the specification of an underlying set, a topology on that set and a sheaf of rings on that topological space, all constructed in a manner that the resulting space has an open covering by affine varieties.

The underlying set of the product will be nothing but the Cartesian product $X \times Y$. To introduce the topology, we observe that $X \times Y$ is the union of the sets $U_{ij} = X_i \times Y_j$, and requiring these to form an open covering, we obtain a topology. Indeed, one declares a subset U to be open when $U \cap U_{ij}$ is open in U_{ij} for each pair of indices i and j.

EXERCISE 3.13 Show that this gives a topology on $X \times Y$. Show that the induced topologies on the sets U_{ij} coincide with their original topologies, and that the projections onto X and Y are continuous. ★

It remains to define the structure sheaf on $X \times Y$. This is also pretty straightforward. We simply say that a function f which is continuous on an open set U is regular at a point $p \in U$ if the restriction $f|_{U_{ij}}$ is regular at p for one (hence for all, see Lemma 3.73 below) of the affine subsets U_{ij} that contain p. Next, we let $\mathcal{O}_{X\times Y}$ be the subsheaf of $\mathcal{C}_{X\times Y}$ whose elements $\mathcal{O}_{X\times Y}(U)$ over an open U consists of the functions regular at all points p in U. Since $\mathcal{C}_{X\times Y}$ is a sheaf, and since being regular is a local property, one gets for free that $\mathcal{O}_{X\times Y}$ is a sheaf of rings.

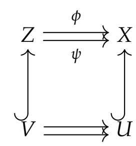
The following lemma ensures that $\mathcal{O}_{X\times Y}|_{U_{ij}} = \mathcal{O}_{X_i \times Y_j}$, so that U_{ij} is an affine cover of $X \times Y$.

LEMMA 3.73 *The setting is as above. If $p \in U_{ij} \cap U_{kl}$, then $f|_{U_{ij}}$ is regular at p if and only if $f|_{U_{kl}}$ is regular at p.*

PROOF: Since the functions $f|_{Uij}$ and $f|_{U_{kl}}$ are both continuous, and since being regular is a local property, we can finish the proof by observing that $(X_i \cap X_l) \times (Y_j \cap Y_k)$ is an open neighbourhood of p both in U_{ij} and in U_{kl} on which $f|_{U_{ij}}$ and $f|_{U_{kl}}$ coincide. ❑

The final point in establishing that $X \times Y$ is a product, is to verify that the two projections are morphisms, and that the universal property is satisfied. The underlying continuous map ϕ associated to two maps ϕ_X and ϕ_Y, is the obvious one, namely the one defined by $\phi(z) = (\phi_X(z), \phi_Y(z))$, and it is a matter of simple verifications to check that ϕ so defined is a morphism. And as usual, we leave the work to the zealous students.

EXERCISE 3.14 Show that the projections π_X and π_Y are morphisms as is the map ϕ described in the text above. ★



Consequences

The *diagonal* Δ_X of a space X is the subset $\Delta_X = \{ (x,x) \mid x \in X \}$ of $X \times X$. More generally if $\phi\colon X \to Y$ is a morphism, one has the notion of the *graph* of ϕ is the subset $\Gamma_\phi = \{ (x,y) \mid \phi(x) = y \}$ of the product $X \times Y$. Putting $\phi = \mathrm{id}_X$ we see that $\Delta_X = \Gamma_{\mathrm{id}_X}$, and clearly Γ_ϕ is the preimage of the diagonal Δ_Y along the map $\phi \times \mathrm{id}_Y\colon X \times Y \to Y \times Y$.

Equalizers
ekvalisatorer

3.74 In topology a space is Hausdorff if and only if the diagonal is closed in the product topology. This hinges on the observation that two open neighbourhoods U and V of points x and y respectively, are disjoint precisely when $U \times V \subseteq X \times Y$ does not meet the diagonal. In algebraic geometry a corresponding statement holds. A prevariety X satisfies the Hausdorff axiom if and only if the diagonal is closed in $X \times X$, but this time in the Zariski topology.

PROPOSITION 3.75 *A prevariety X is a variety if and only if the diagonal $\Delta_X \subseteq X \times X$ is closed.*

PROOF: The diagonal being the equalizer of the two projections, it will be closed when X is a variety. Assume then that Δ_X is closed and let $\phi, \psi\colon Z \to X$ be two morphisms. Their equalizer $\{ z \mid \phi(z) = \phi(z) \}$ is the inverse image of Δ_X by the morphism $Z \to X \times Y$ whose components are ϕ and ψ. Hence it is closed. ❑

As an immediate corollary one has

COROLLARY 3.76 *When X and Y are varieties and $\phi\colon X \to Y$ is a morphism, the graph Γ_ϕ is closed in the product $X \times Y$.*

3.77 The second application of the fact that products of varieties are varieties illustrates the use of a general principle often referred to as "reduction to the diagonal". When we come to the study of intersections of subvarieties in affine and projective space, this principle will turn out to be a very useful tool. In its simplest form – formulated for sets – it is the observation that the intersection $U \cap V$ of two subsets U and V of a set X, is naturally bijective to the intersection $U \times V \cap \Delta_X$. This relation persists when U and V are open subsets of a prevariety

X, but of course with the annotation "bijective" replaced with "isomorphic". A consequence is that the intersection of two open affines in a variety is affine:

Proposition 3.78 *Assume that U and V are open affine subsets of the variety X, then the intersection $U \cap V$ is affine*

Proof: The intersection $U \cap V$ is isomorphic to the intersection $U \times V \cap \Delta_X$, but the product $U \times V$ is affine and since Δ_X is closed, X being a variety, it ensues that $U \times V \cap \Delta_X$ is closed in $U \times V$. Hence affine. ❑

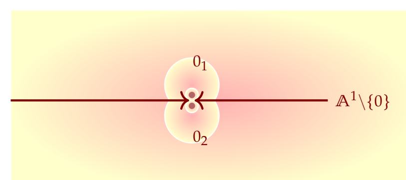
The proposition does not persist for prevarieties that are not varieties. In Exercise 3.10 on page 64 you were asked to make an example by gluing two copies of $\mathbb{A}^2$ together along $\mathbb{A}^2 \backslash \{(0,0)\}$.

Exercise 3.15 Show that $U \cap V$ is isomorphic to $U \times V \cap \Delta_X$ whenever U and V are open subsets of a prevariety X. ★

Exercise 3.16 Let X be a variety and $Y \subseteq X$ a closed sub-prevariety (defined as in the introduction to Proposition 3.41 on page 59). Show that Y is a variety. ★

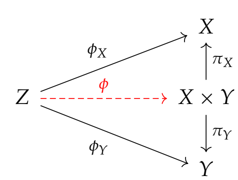
Exercise 3.17 Let X be the bad guy from Example 3.60 on page 63. Show that $X \times X$ can be covered by four open subsets each one isomorphic to $\mathbb{A}^2$. Verify explicitely that the diagonal in $X \times X$ is not closed. ★

Exercise 3.18 *(Fibre products.)* Let $\phi\colon X \to S$ and $\psi\colon Y \to S$ be two morphisms between varieties. Denote by $X \times_S Y$ be the subset of $X \times Y$ defied as $\{ (x,y) \mid \phi(x) = \psi(y) \}$. Show that $X \times_S Y$ is a closed subset by showing it equals the equalizer of the two maps $\phi \circ \pi_X$ and $\psi \circ \pi_Y$. Formulate a universal property that characterizes the fibre product $X \times_S Y$. ★



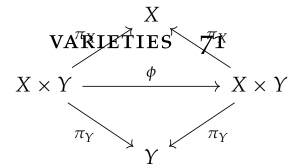
Exercises

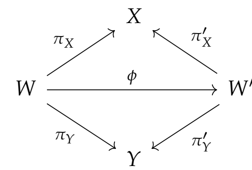
3.19 *(Rational cusp.)* Consider the curve C in $\mathbb{A}^2$ whose equation is $y^2 - x^3$. Show that C can be parametrized by the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^2$ defined as $\phi(t) = (t^2, t^3)$. Describe the map $\phi^*\colon A(C) \to A(\mathbb{A}^1)$. Show that ϕ is bijective but not an isomorphism. Show that the function field of C equals $k(t)$.



3.20 *(Rational node.)* In this exercise we let C be the curve in $\mathbb{A}^2$ whose equation is $y^2 - x^2(x+1)$. Define a map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^2$ by $\phi(t) = (t^2 - 1, t(t^2 - 1))$. Show that $\phi(\mathbb{A}^1) = C$, and describe the map $\phi^*\colon A(C) \to A(\mathbb{A}^1)$. Show that ϕ is not an isomorphism, but induces an isomorphism $\mathbb{A}^1 \backslash \{\pm 1\} \to C\backslash\{0\}$. Show that the function field of C equals $k(t)$.

3.21 Let C be one of the curves from the two previous exercises. Show that, except for finitely many, every line through the origin intersects C in exactly one other point. What are the exceptional lines in the two cases? Use this to give a geometric interpretation of the parameterizations in the previous exercises.





3.22 *(An acnode.)* Consider the curve D given by $y^2 = x^2(x-1)$ in $\mathbb{A}^2$. Make a sketch of the real points of D (see the figure in the margin); notice that the origin is isolated among the real points – such a point is called an *acnode*. Show that $(t^2+1, t(t^2+1))$ is a parameterization of D. Exhibit a complex linear change of coordinates in $\mathbb{A}^2$ that brings D on the form in problem 3.20 above.

3.6 An epilogue – about Zariski topologies.

As an epilogue, we remind you that a variety has two ingredients: a topological space X and the structure sheaf $\mathcal{O}_X$. Among the two the structure sheaf is the main player, the Zariski topology having a more supportive role. For instance, if X is an irreducible and Noetherian space whose only closed irreducible sets are the points, the closed sets, apart form the entire space, are precisely the finite subsets. This means that all such spaces are homeomorphic as longs as their cardinality is the same. So for instance, the affine lines $\mathbb{A}^1$ over different countable[7] fields are homeomorphic, and they are even homeomorphic to the bad guy – the line with a double origin – we just constructed.

[6] There is a purely algebraic proof of these facts, see Proposition ?? in CA, but below we shall give geometric versions

Later on, after having introduced the concept of dimension, we shall see that any Noetherian irreducible space of dimension one falls in this category, so they are all homeomorphic (as long as they live over equipotent ground fields). But there is an extremely rich fauna of such varieties!

In higher dimensions the Zariski topologies play a more decisive role, but still they do not distinguish varieties very well. One may, for instance, prove that the affine planes $\mathbb{A}^2(\bar{\mathbb{F}}_p)$ and $\mathbb{A}^2(\bar{\mathbb{F}}_q)$ over the algebraic closures of the finite fields $\mathbb{F}_p$ and $\mathbb{F}_q$ are homeomorphic even when $p \neq q$. However, if such a homeomorphisms preserves the degree of curves (*i.e.* takes lines to lines, conics to conics etc..) it follows that $p = q$. (If these matters interest you, see the [8].)

The diagonal
diagonalen

3.23 Show that the algebraic closure of a countable field is countable.

3.24 Wiegand and Krauter proved that given a curve in $\mathbb{A}^2(\bar{\mathbb{F}}_p)$ and a subset S of C meeting every irreducible component of C, then there exists an irreducible curve D in $\mathbb{A}^2$ such that $C \cap D = S$. Let $\bar{\mathbb{F}}_p$ and $\bar{\mathbb{F}}_{p'}$ be the algebraically closure of two prime fields.

a) Show that there are only countably many irreducible curves in $\mathbb{A}^2(\bar{\mathbb{F}}_p)$ (and in $\mathbb{A}^2(\bar{\mathbb{F}}_{p'})$).
b) Assume that $C \subseteq \mathbb{A}^2(\bar{\mathbb{F}}_p)$ and $C' \subseteq \mathbb{A}^2(\bar{\mathbb{F}}_{p'})$ are two curves. Use the statement above to prove that any homeomorphism $\phi\colon C \to C'$ extends to a homeomorphisms $\mathbb{A}^2(\bar{\mathbb{F}}_p) \simeq \mathbb{A}^2(\bar{\mathbb{F}}_{p'})$. Hint: Number the irreducible curves in $\mathbb{A}^2(k)$ and use induction.
c) Conclude that $\mathbb{A}^2(\bar{\mathbb{F}}_p)$ and $\mathbb{A}^2(\bar{\mathbb{F}}_{p'})$ are homeomorphic.

★

Chapter 4

Projective varieties

TOPICS IN CHAPTER 4: *Projective spaces – homogeneous coordinates – closed projective sets – homogenous ideals and closed projective sets – projective Nullstellensatz – distinguished open sets – Zariski topology and regular functions – projective varieties – global regular functions on projective varieties – morphism from quasi-projective varieties – linear projections.*

Projective geometry arose in the wake of the discovery of the perspective by iItalian renaissance painters like Leonardo da Vinci, Albrecht Dürer and Filippo Brunelleschi to mention some. In a perspective drawing, one considers bundles of light rays emanating from or meeting at a point (the observers eye) or meeting at an apparent point at infinity, the so-called vanishing point, when rays are parallel. Figures are perceived the same if one is the projection of the other.

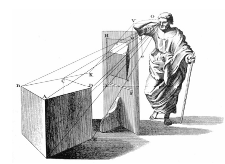


Figure 4.1: The discovery of perspective in art.

In the beginning, projective geometry was purely a synthetic geometry (no coordinates, no functions, merely points and lines). The properties of the different figures that where studied were properties invariant under projection from a point. Subsequently, an analytic theory developed and eventually became the basis for the projective geometry as we know it in algebraic geometry today.

The projective spaces and the projective varieties are in some sense the algebro-geometric counterparts to compact spaces, with which they share many nice properties.

Non-compact spaces are on the other hand typically difficult to handle; if you discard a bunch of points in an arbitrary manner from a compact space (for instance, a sphere) it is not much you can say about the result unless you know the way the discarded points were chosen, and moreover, functions can tend to infinity near the missing points which sometimes can be problematic. Compact spaces and projective varieties are in some sense complete, they do not suffer from the deficiencies of these 'punctured' spaces — hence their importance and popularity.

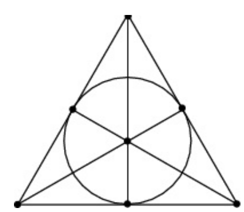


Figure 4.2: The Fano plane; the projective plane over the field with two elements.

4.1 We shall begin with getting more acquainted with the projective spaces, and subsequently equip $\mathbb{P}^n$ with a variety structure. This amounts to endowing it with a topology (which naturally will be called the *Zariski topology*) and telling what functions on $\mathbb{P}^n$ are regular; *i.e.* defining the sheaf $\mathcal{O}_{\mathbb{P}^n}$ of regular functions. Finally, we shall introduce the larger class of *projective varieties*. They will be the closed subvarieties of $\mathbb{P}^n$; that is, irreducible subsets of $\mathbb{P}^n$ with the topology induced from the Zariski topology on $\mathbb{P}^n$ and equipped with the sheaf of rings of functions which locally are restrictions of regular functions.

4.1 The projective spaces $\mathbb{P}^n$

4.2 Let n be a non-negative integer. The underlying set of *the projective n-space* $\mathbb{P}^n$ over k is the set of lines passing through the origin in $\mathbb{A}^{n+1}$; or in other words, the set of one-dimensional vector subspaces of k^{n+1}.

The projective n-spaces $\mathbb{P}^n$
de projektive rommene

The ground field k does not appear in the notation, but is tacitly understood. If need be, it will be indicated with a subscript: $\mathbb{P}^n_k$. Projective spaces over the complex numbers are commonly denoted as $\mathbb{CP}^n$.

EXAMPLE 4.3 There is merely one line through the origin in $\mathbb{A}^1$, so $\mathbb{P}^0$ is just a point. ☆

To each point $(a_0, \ldots, a_n) \in \mathbb{A}^n$ other than the origin, we can associate the point in $\mathbb{P}^n$ corresponding to the line spanned by it. It is clear that two points $a = (a_0, \ldots, a_n)$ and $b = (b_0, \ldots, b_n)$ in $\mathbb{A}^{n+1} - \{0\}$ span the same line if and only if $b = t \cdot a$, i.e.,

$$(b_0, \ldots, b_n) = (ta_0, \ldots, ta_n) \tag{4.1}$$

for some $t \in k^*$. Thus one may as well consider $\mathbb{P}^n$ as the set of equivalence classes in $\mathbb{A}^{n+1} - \{0\}$ under the equivalence relation (4.1). We write $(a_0 : \ldots : a_n) \in \mathbb{P}^n$ for the equivalence class of the point $(a_0, \ldots, a_n) \in \mathbb{A}^{n+1} - 0$.

We write

$$\pi\colon \mathbb{A}^{n+1} - \{0\} \to \mathbb{P}^n,$$

for the quotient map, i.e., the map that sends $(a_0, \ldots, a_n)$ to $(a_0 : \ldots : a_n)$.

Homogeneous coordinates and basic open sets

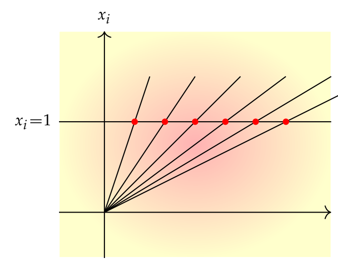
4.4 Coordinates are of course very useful and desirable tools, but on $\mathbb{P}^n$ there are no *global* coordinates. However, there is a good substitute. If $[x] \in \mathbb{P}^n$ corresponds to the line through the point $x = (x_0, \ldots, x_n)$, we say that $(x_0 : \ldots : x_n)$ are *homogeneous coordinates* of the point $[x]$ — notice the use of colons to distinguish them from the usual coordinates in $\mathbb{A}^{n+1}$. The homogeneous coordinates of $[x]$ depend on the choice of the point x in the line $[x]$ and are not unique; they are only defined up to a scalar multiple, so that $(x_0 : \ldots : x_n) = (tx_0 : \ldots : tx_n)$ for all elements $t \in k^*$. Be aware that $(0 : \ldots : 0)$ is forbidden; it does not correspond to any line through the origin and hence is not the coordinate of any point in $\mathbb{P}^n$.

Homogeneous coordinates
homogene koordinater

4.5 Visualizing projective spaces can be quite challenging, but the following is one way of thinking about them. This description of $\mathbb{P}^n$ is an invaluable tool when working with projective spaces and will be important in parts of the subsequent theoretical development.

Fix one of the coordinates, say x_i, and let $D_+(x_i)$ denote the set of lines $[x] = (x_0 : \ldots : x_n)$ for which $x_i \neq 0$. These sets are called the *basic open subsets* of $\mathbb{P}^n$ like their affine cousins. The subvariety $A_i = Z(x_i - 1)$ of $\mathbb{A}^{n+1}$ where $x_i = 1$ will also be useful. Every line $[x]$ with $x_i \neq 0$ intersects the subvariety A_i in precisely one point, namely the point $(x_0/x_i, \ldots, x_n/x_i)$. Thus there is a natural bijection α_i between the subsets $D_+(x_i)$ of $\mathbb{P}^n$ and A_i of $\mathbb{A}^{n+1}$. Now, obviously the subvariety A_i is isomorphic to affine n-space $\mathbb{A}^n$ (the projection that forgets the i-th coordinate gives an isomorphism of varieties); hence $D_+(x_i)$ is in a natural bijective correspondence (later on we shall see it is an isomorphism) with $\mathbb{A}^n$. To avoid unnecessary confusion, let us denote the coordinates on A_i by t_j with j running from 0 to n but staying different from i. Then the bijection α_i from $D_+(x_i)$ to A_i is given by the assignment $t_j = x_j/x_i$; it has the restriction $\pi|_{A_i}$ of the canonical projection as inverse.

The basic open subsets
prinsipale åpne mengder



4.6 The complement of the distinguished open subset $D_+(x_i)$ consists of the lines lying in the subvariety of $\mathbb{A}^{n+1}$ where $x_i = 0$; that is, the subvariety $Z(x_i)$. This is an affine n-space with coordinates[1] $(x_0, \ldots, \hat{x}_i, \ldots, x_n)$, and so the complement $\mathbb{P}^n \backslash D_+(x_i)$ is equal to the projective space $\mathbb{P}^{n-1}$ of lines in that affine space. It is called the *hyperplane at infinity*.

[1] A hat (or *circumflex* which is the proper name) indicates that a variable is missing.

Hyperplane at infinity
hyperplanet i det uendelige

Be aware that the 'hyperplane at infinity' is a *relative* notion; it depends on the choice of the coordinates. In fact, given any non-zero linear functional $\lambda(x)$ in the x_i's, one may chose coordinates so that the hyperplane $\lambda(x) = 0$ is the hyperplane at infinity.

Examples

4.7 *(The projective line)* When $n = 1$, we have the *projective line* $\mathbb{P}^1$. It consists of a 'big' subset isomorphic to $\mathbb{A}^1$ to which one has added a point at infinity. Every point can be made the point at infinity by an appropriate coordinate change — chose coordinates $(x_0 : x_1)$ so that the point is $(1 : 0)$; then $D_+(x_1)$ will be the big $\mathbb{A}^1$ and the hyperplane at infinity $Z_+(x_1)$ will consist of $(1 : 0)$ alone.

The projective line over the complex numbers, endowed with the Euclidean topology, is the good old *Riemann sphere* known from courses in complex analysis. Indeed, let $(x_0 : x_1)$ be the homogeneous coordinates on $\mathbb{P}^1$. In the set $D_+(x_1)$ — which is isomorphic to $\mathbb{A}^1$; that is, to $\mathbb{C}$ — one uses $z = x_0/x_1$ as coordinate, where as in $D_+(x_0)$ the coordinate is $z^{-1} = x_1/x_0$; and the patching on $D_+(x_0) \cap D_+(x_1)$ is the same as the one used to construct the Riemann sphere.

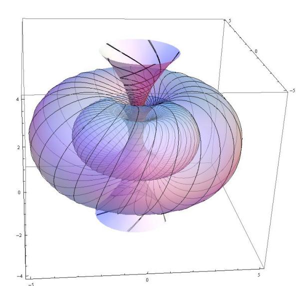
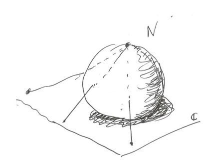


Figure 4.3: The Hopf fibration

The projection map $\pi \colon \mathbb{C}^2 \backslash \{0\} \to \mathbb{CP}^1$ is interesting. Restricted to the unit sphere $S^3 \subset \mathbb{C}^2$ it becomes the illustrious map $S^3 \to S^2$ which goes under the name of the *Hopf fibration*. It is easy to see that the fibres are circles, so that π is a fibration in circles of the three-sphere S^3 over then two-sphere S^2.



The projective line over the reals $\mathbb{R}$ is topologically just a circle; note that there is only one point at infinity. One uses lines and not rays through the origin, and so there is no distinction between ∞ and $-\infty$.

4.8 *(The projective plane)* The variety $\mathbb{P}^2$ is called *the projective plane*. It has, for each i, a basic open subset $\mathbb{A}^2 = D_+(x_i)$ with a projective line at infinity 'wrapped' around it.

The projective plane contains many subsets that are in a natural one-to-one correspondence with the projective line $\mathbb{P}^1$. The set of one dimensional subspaces contained in a fixed two dimensional vector subspace of $\mathbb{A}^3$ is such a $\mathbb{P}^1$, and of course any two dimensional subspace will do. These subsets are called *lines* in $\mathbb{P}^2$.

By linear algebra, two different two dimensional vector subspaces of $\mathbb{A}^3$ intersect along a line through the origin. This leads to the fundamental observation that the two corresponding lines in $\mathbb{P}^2$ intersect in a *unique* point. Two lines do not necessarily meet in the 'finite part'; that is, in the affine 2-space $\mathbb{A}^2$ where $x_i \neq 0$. This occurs if and only they have a common intersection with the line at infinity, and then one says that the two lines meet at infinity. And naturally, when they do not meet in the finite part, they are experienced to be parallel; hence 'parallel' lines meet at a common point at infinity.

The projective plane over the reals, is a subtle creature. After having picked one of the coordinates x_i, we find a cell of shape $D_+(x_i)$ in $\mathbb{P}^2$, which is a copy of $\mathbb{R}^2$, enclosed by the line at infinity; a circle bordering the affine world like the Midgard Serpent. Again, be aware that in constructing $\mathbb{P}^2$ one uses lines through the origin and not rays emanating from the origin. This causes $\mathbb{P}^2$ to be non-orientable – tubular neighbourhoods of the lines are in fact Möbius bands.

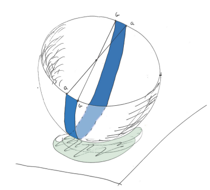


Figure 4.4: A Möbius band in the real projective plane.

4.9 *(The Fano plane)* Figure 4.2 on page 73 illustrates the 'synthetic' Fano plane with seven points and seven lines. This configuration is realized in the present context as the set of points in the projective plane $\mathbb{P}^2_{\bar{\mathbb{F}}_2}$ over the field[2] $\bar{\mathbb{F}}_2$ whose coordinates belong to $\mathbb{F}_2$; that is, points whose coordinates are either 0 or 1.

[2] An algebraic closure of the field $\mathbb{F}_2$ with two elements

There are 7 such point (there are 8 in $\mathbb{F}_2^3$, but $(0:0:0)$ is forbidden). There are also 7 non-zero linear forms $\lambda_1 x_0 + \lambda_1 x_1 + \lambda_2 x_2$ whose coefficients λ_i belong to $\mathbb{F}_2$, and the seven lines these define together with the seven points constitute a Fano plane.

We leave it to the interested students to verify (easy linear algebra in characteristic two) that this indeed is a (synthetic) projetive plane (*i.e.* two of the seven points lie on exactly one of the seven lines, and two of the seven lines meet in exactly one of the seven point; in both cases 'exactly' comes for free).

☆

4.2 *The Zariski topology and projective Nullstellensatz*

The Zariski topology

One may use the projection map $\pi\colon \mathbb{A}^{n+1}\backslash\{0\} \to \mathbb{P}^n$ to equip $\mathbb{P}^n$ with a topology: a subset in $X \subseteq \mathbb{P}^n$ is declared to be closed if and only if the inverse image $\pi^{-1}(X)$ is closed in $\mathbb{A}^{n+1}\backslash\{0\}$. Since the operation of forming inverse images behaves well with respect to intersections and unions (*i.e.* commutes with them), these sets are easily seen to fulfil the axioms for the closed sets of a topology. Naturally, this topology is called the *Zariski topology* on $\mathbb{P}^n$. It is the quotient topology with respect to the equivalence relation on $\mathbb{A}^{n+1}\backslash\{0\}$ giving $\mathbb{P}^n$.

The Zariski topology
Zariski topologien

4.10 Polynomials on $\mathbb{A}^{n+1}$ do not descend to functions on $\mathbb{P}^n$ unless they are constant — non-constant polynomials are not constant along lines through the origin. However, if F is a *homogeneous* polynomial, it holds true that $F(tx) = t^d F(x)$ where d is the degree of F and t any scalar, so if F vanishes at a point x, it vanishes along the entire line joining x to the origin. Hence it is makes sense to say that F is *zero* at a point $[x] \in \mathbb{P}^n$; and it is meaningful to talk about the zero locus in $\mathbb{P}^n$ of a set of homogeneous polynomials. A homogeneous ideal $\mathfrak{a}$ in $k[x_0, \ldots, x_n]$ is generated by homogeneous polynomials, and we can speak about the *zero locus* $Z_+(\mathfrak{a})$ in $\mathbb{P}^n$ as the common set of zeros of the generators, which one verifies also equals the common set of zeros of all elements from $\mathfrak{a}$.

The zero locus of a homogeneous ideal
nullpunktsmengden til et homogent ideal

In this story there is one ticklish point. The maximal ideal $\mathfrak{m}_+ = (x_0, \ldots, x_n)$ vanishes only at the origin in $\mathbb{A}^{n+1}$, and so it defines the *empty set* in $\mathbb{P}^n$; indeed, for no point in $\mathbb{P}^n$ do all the coordinates x_i vanish. Hence $\mathfrak{m}_+$ goes under the name of *the irrelevant ideal*.

The irrelevant ideal
de irrelevante idealet

A subset X of type $Z_+(\mathfrak{a})$ is called a *closed projective set*, and we shall shortly see that all closed subsets of $\mathbb{P}^n$ are of this kind. The topology induced on X from the Zariski topology on $\mathbb{P}^n$ is called the *Zariski topology* on X. If additionally X is an *irreducible* space, X is said to be a *projective variety* (it will in due course be equipped with a sheaf of regular functions), and an open subset of a projective variety is said to be a *quasi-projective variety*. Since π by definition is continuous, the Zariski topology makes the projective spaces $\mathbb{P}^n$ irreducible, so they are projective varieties.

Closed projective set
lukkede projektive mengder

Projective varieties
projektive variteter

Quasi-projective varieties
kvasiprojektive variteter

EXERCISE 4.1 Show that if $\mathfrak{a}$ and $\mathfrak{b}$ are two homogeneous ideals, then $\mathfrak{a} \cdot \mathfrak{b}$ and $\mathfrak{a} + \mathfrak{b}$ are homogeneous, and it holds true that $Z_+(\mathfrak{a} \cdot \mathfrak{b}) = Z_+(\mathfrak{a}) \cup Z_+(\mathfrak{b})$ and $Z_+(\mathfrak{a} + \mathfrak{b}) = Z_+(\mathfrak{a}) \cap Z_+(\mathfrak{b})$. Show that the radical $\sqrt{\mathfrak{a}}$ is homogeneous and that $Z_+(\sqrt{\mathfrak{a}}) = Z_+(\mathfrak{a})$. ★

The projective Nullstellensatz

4.11 The correspondence between homogeneous ideal in the polynomial ring $k[x_0, \ldots, x_n]$ and closed subsets of the projective space $\mathbb{P}^n$ is as in the affine case governed by a Nullstellensatz.

There are however, some differences. In the projective case the ideals must of course be homogeneous, and there is a slight complication concerning the ideals with empty zero locus. In the affine case, the zero locus of an ideal $\mathfrak{a}$ is empty if and only if $1 \in \mathfrak{a}$. In $\mathbb{P}^n$ one must exclude the ideal $\mathfrak{m}_+ = (x_0, \ldots, x_n)$; there is no point in $\mathbb{P}^n$ on which all the homogeneous coordinates are zero, and the same applies to any $\mathfrak{m}_+$-primary ideal: the elements of an ideals $\mathfrak{a}$ whose radical equals the irrelevant ideal $\mathfrak{m}_+$, do not have common zeros in $\mathbb{P}^n$, and so $Z_+(\mathfrak{a}) = \emptyset$.

4.12 A simple and geometric way of thinking about the interplay between the affine and the projective Nullstellensatz, is via the *affine cone* $C(X)$ over a closed set $X \subseteq \mathbb{P}^n$. It is defined as $C(X) = \pi^{-1}X \cup \{0\}$ and is a cone in the sense that it contains the line joining any one of its points to the origin; or phrased differently, if $x \in C(X)$, then $tx \in C(X)$ for all scalars $t \in k$ (recall that the projection map $\pi\colon \mathbb{A}^{n+1}\backslash\{0\} \to \mathbb{P}^n$ sends a point to the line joining it to the origin).

The affine cone over a projective variety
affine kjegler

The inverse image $\pi^{-1}(X)$ will now and then be called the *punctured cone* over X and denoted by $C_0(X)$; so that $C_0(X) = C(X) \cap (\mathbb{A}^{n+1}\backslash\{0\})$. The affine cone $C(X)$ is closed, since $C_0(X)$ is closed in $\mathbb{A}^{n+1}\backslash\{0\}$ and 0 lies in its closure in $\mathbb{A}^{n+1}$. More explicitly, if $X = Z_+(\mathfrak{a}) \subset \mathbb{P}^n$, then $C(X) = Z(\mathfrak{a}) \subset \mathbb{A}^{n+1}$.

The punctured cone
den punkterte kjeglen

This sets up one-to-one correspondence between closed non-trivial[3] cones in $\mathbb{A}^{n+1}$ and non-empty closed subsets in $\mathbb{P}^n$:

[3] Formally, a cone in $\mathbb{A}^{n+1}$ is a subset closed under homothety; that is, if $x \in C$, then $tx \in C$ for all scalars $t \in k$. Clearly the singleton $\{0\}$ comply with this definition, so $\{0\}$ is a cone. It is called the *trivial cone* or the *null cone*.

LEMMA 4.13 *Associating the affine cone $C(X)$ to X gives a bijection between closed non-empty subsets of $\mathbb{P}^n$ and closed non-trivial cones in $\mathbb{A}^{n+1}$. The bijection respects inclusions, intersections and unions.*

PROOF: Let $C \subseteq \mathbb{A}^{n+1}$ be a non-trivial cone and denote by C_0 the punctured cone $C_0 = C(X)\backslash\{0\}$. There are two points to notice; firstly, C_0 is nonempty, and if it is closed in $\mathbb{A}^{n+1}\backslash\{0\}$, its closure in $\mathbb{A}^{n+1}$ clearly satisfies $\overline{C_0} = C$, and secondely, $\pi^{-1}\pi(C_0) = C_0$. It follows that C is closed in $\mathbb{A}^{n+1}$ if and only if C_0 is closed in $\mathbb{A}^{n+1}\backslash\{0\}$, and by the definition of the Zariski topology on $\mathbb{P}^n$, we infer that C is closed if and only if $\pi(C_0)$ is closed in $\mathbb{P}^n$. This shows that the correspondence of the lemma is surjective, and it is injective since it holds true that $\pi(\pi^{-1}C) = C$ because π is surjective.

The last statement in the lemma follows from general features of inverse images. ❑

4.14 Next comes the translation between cones and homogeneous ideals. The vanishing locus of a homogeneous ideal is, as we already have noted, a cone, and the converse holds true as well. To any closed subset $X \subseteq \mathbb{P}^n$, we let $I(X)$ be the ideal in the polynomial ring $k[x_0, \ldots, x_n]$ generated by all homogeneous polynomials that vanish in X. It is an homogeneous ideal, and one has $I(X) = I(C(X))$:

LEMMA 4.15 *If C is a closed cone in $\mathbb{A}^{n+1}$, then the ideal $I(C)$ is homogeneous.*

PROOF: Let $f \in I(C)$ be of degree d and write $f = \sum_{i \geqslant 0} f_i$ for the decomposition of f into homogeneous elements. We need to show that each f_i lies in $I(C)$. But since C is a cone, if $x \in C$, then also $f(tx) = 0$ for any $t \in k$. This implies that the polynomial in t

$$f(tx) = f_0(x) + f_1(x)t + f_2(x)t^2 + \ldots + f_d(x)t^d$$

has infinitely many roots (k is infinite) and consequently is the zero-polynomial. In particular, all the coefficients are zero;*i.e.* each $f_i(x) = 0$. ❑

4.16 Combining the bijection in Lemma 4.13 on the preceding page above with the bijection between homogeneous radical ideals and closed cones from Lemma 4.15 and the affine Nullstellensatz, one arrives at the following version of the Nullstellensatz in a projective setting:

Proposition 4.17 (Projective Nullstellensatz) *Assume that $\mathfrak{a}$ is a homogeneous ideal in $k[x_0, \ldots, x_n]$.*

- *i) Then $Z_+(\mathfrak{a})$ is empty if and only if $1 \in \mathfrak{a}$ or $\mathfrak{a}$ is $\mathfrak{m}_+$-primary; that is, $\mathfrak{m}_+^N \subseteq \mathfrak{a}$ for some N.*
- *ii) If $Z_+(\mathfrak{a}) \neq \emptyset$, it holds true that $I(Z_+(\mathfrak{a})) = \sqrt{\mathfrak{a}}$.*
- *iii) Associating $I(X)$ with X sets up a bijection between closed non-empty subsets $X \in \mathbb{P}^n$ and proper, radical homogeneous ideals I in $k[x_0, \ldots, x_n]$ different from the irrelevant ideal.*
- *iv) The subset $Z_+(\mathfrak{a})$ is irreducible if and only if the radical $\sqrt{\mathfrak{a}}$ is a prime ideal.*

Proof: We already have argued for most of the statements; what remains, is to clarify when $Z_+(\mathfrak{a})$ is empty, and this happens precisely when $Z_+(\mathfrak{a}) \subseteq \{0\}$. There are two cases: either $Z(\mathfrak{a}) = \emptyset$ or $Z(\mathfrak{a}) = \{0\}$, which by the Affine Nullstellensatz correspond to respectively $1 \in \mathfrak{a}$ or $\sqrt{\mathfrak{a}} = I(\{0\}) = \mathfrak{m}_+$.

For the last statement, it is quite clear (and left to the zealous students to verify) that X is irreducible if and only if the cone $C(X)$ over X is irreducible. ❑

The distinguished open subsets

In the same spirit as in the definition of the basic open sets $D_+(x_i)$, one defines the *distinguished open subset* $D_+(F) = \{ [x] \in \mathbb{P}^n \mid F(x) \neq 0 \}$ for any homogeneous polynomial F. All these sets are open in $\mathbb{P}^n$ simply because their complements are the closed sets $Z_+(F)$.

Distinguished open subsets
prinsipale, utpekte, særskilte åpne mengder

4.18 In Paragraph 4.5, we introduced the subsets A_i of $\mathbb{A}^{n+1}$ where the i-th coordinate equals 1, and showed that the restriction $\pi|_{A_i}$ is a bijection between A_i and $D_+(x_i)$; now we go one step further:

Proposition 4.19 *The restriction $\pi|_{A_i}$ of π is a homeomorphism between A_i and $D_+(x_i)$.*

In the proof we need the process of *homogenization* of a polynomial which, is a systematic way of producing a homogeneous polynomial f^h from a polynomial f. It is not canonical, but depends on the choice of one of the variables x_i. If d is the degree of f, one puts

Homogenization of polynomials
homogenisering av polynomer

$$f^h(x_0, \ldots, x_n) = x_i^d f(x_0 x_i^{-1}, \ldots, x_n x_i^{-1}). \tag{4.2}$$

Formally this seems like a rational expression, but when expanded, all terms on the right side of (4.2) will have denominators whose degrees are at most d, and so f^h will be a genuine polynomial. For example, if $f = x_1 x_2^3 + x_3 x_0 + x_0$, one finds that relative to the variable x_0 one has

$$f^h(x_0, x_1, x_2, x_3) = x_0^4\big(x_1 x_0^{-1}(x_2 x_0^{-1})^3 + x_3 x_0^{-1} + 1\big) = x_1 x_2^3 + x_3 x_0^3 + x_0^4.$$

The net effect of the homogenization process is to fill up all the monomial terms with the chosen variable to make them have the same degree. This shows that f^h is homogeneous of degree d; which, by the way, also is clear since the fractions $x_j x_i^{-1}$ are invariant when the variables are scaled by t, and the front factor x_i^d changes by the factor t^d.

The important relation $f|_{A_i} = f^h|_{A_i}$ is easy to establish (just put $x_i = 1$). On the other hand, if one starts with a homogeneous polynomial F and considers the restriction $F|_{A_i}$ (which amounts to setting $x_i = 1$) one will loose any factor of F which is a power of x_i; hence it does not hold in general that $(F|_{A_i})^h = F$.

PROOF OF PROPOSITION 4.19: Now, we come back to the proof of the proposition. The restriction $\pi|_{A_i}$ is, as already observed, continuous, so our task is to show that the also inverse is continuous; or equivalently, that $\pi|_{A_i}$ is a closed map.

Since any closed subset of A_i is the intersection of sets of the form $Z = Z(f) \cap A_i$, and π being bijective takes intersections to intersections, it suffices to demonstrate that $\pi(Z(f) \cap A_i)$ is closed in $D_+(x_i)$ for any polynomial f on A_i. But this is precisely what the homogenization f^h is constructed for: the subset $Z(f^h)$ of $\mathbb{A}^{n+1}$ is a closed cone satisfying $Z(f^h) \cap A_i = Z(f) \cap A_i$, and this means that $\pi(Z(f) \cap A_i) = Z_+(f^h) \cap D_+(x_i)$. ❏

Exercises

4.2 Let $V \subseteq \mathbb{A}^{n+1}$ be a linear vector subspace of dimension $m+1$. Show that V is a cone and that the corresponding projective variety $\mathbb{P}(V)$ is isomorphic to $\mathbb{P}^m$. It is called a *linear subvariety* of $\mathbb{P}^n$. Show that if W is another linear subspace of dimension $m'+1$ and $m + m' \geqslant n$, then $\mathbb{P}(V)$ and $\mathbb{P}(W)$ has a non-empty intersection, which is a linear subvariety. Show that if $m + m' = n$ and V and W are generic; that is, $V + V = k^{n+1}$, then $\mathbb{P}(V) \cap \mathbb{P}(W)$ is a point.

Linear subvarieties
linrære undervarieteter

4.3 Let $(a_{10} : a_{11} : a_{12})$ and $(a_{20} : a_{21} : a_{22})$ be two different points in $\mathbb{P}^2$. Show that the line through them has equation

$$\det \begin{pmatrix} x_0 & x_1 & x_2 \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix} = 0.$$

Dually, if $\lambda_{10}x_0 + \lambda_{11}x_1 + \lambda_{12}x_2 = 0$ and $\lambda_{20}x_0 + \lambda_{21}x_1 + \lambda_{22}x_2 0 =$ are equations for two different lines, show that the coordinates of their intersection point are the 2×2-minors of the matrix

$$\begin{pmatrix} \lambda_{10} & \lambda_{11} & \lambda_{12} \\ \lambda_{20} & \lambda_{21} & \lambda_{22} \end{pmatrix}.$$

4.4 Show that n hyperplanes in $\mathbb{P}^n$ always have a common point of intersection. Show that n linearly independent hyperplanes meet in exactly one point.

4.5 Let $p_0, \ldots, p_n$ be $n+1$ points in $\mathbb{P}^n$ and let $v_0, \ldots, v_n$ be non-zero vectors lying on the corresponding lines in $\mathbb{A}^{n+1}$. Show that the p_i's lie on a hyperplane if and only if the v_i's are linearly dependent.

4.6 Let $f(x_0, x_1, x_2, x_3) = x_0^3 x_2 + x_3^2 x_1 + 1$. Determine f^h with respect to each of the four variables.

4.7 In this exercise $n = 2$ and the coordinates are x_0 and x_1. Let $f(x_0) = (x_0 - a)(x_0 - b)$. Determine f^h and make a sketch of $Z(f) \cap \mathbb{A}^1$ and the cone $Z(f^h)$

4.8 Show that any affine variety is quasi-projective.

4.9 How doe the circles and the ellipses fit into the picture described in Examples 4.34 and 1.27?

★

4.3 *Regular functions on projective varieties*

The sheaf of regular functions on projective varieties

Although polynomials on $\mathbb{A}^{n+1}$ do not descend to functions on $\mathbb{P}^n$, certain rational functions do. To describe these, assume that a and b are two polynomials both homogeneous of the same degree, say d. Although none of them define a function on the projective space $\mathbb{P}^n$, their fraction do, at least at points where the denominator does not vanish. Indeed, for $t \neq 0$ we have

$$\frac{a(tx)}{b(tx)} = \frac{t^d a(x)}{t^d b(x)} = \frac{a(x)}{b(x)},$$

whenever $b(x) \neq 0$. The function $a(x)/b(x)$ thus takes the same value at all points on the line $[x]$, and this common value is the value of $a(x)/b(x)$ at $[x]$.

4.20 This observation leads to the definition of the *sheaf of regular functions* on $\mathbb{P}^n$, or more generally to the notion of regular functions on any closed projective set $X \subseteq \mathbb{P}^n$; hence to the sheaf $\mathcal{O}_X$ of regular functions on X.

The sheaf of regular functions on $\mathbb{P}^n$
knippet av regulære funksjoner

A function f on an open subset U of X is said to be *regular* at a point $p \in U$ if there exists an open neighbourhood $V \subseteq U$ of p in X and homogeneous polynomials a and b of the same degree such that $b(x) \neq 0$ throughout V and such that the equality

Regular functions
regulære funksjoner

$$f(x) = \frac{a(x)}{b(x)}$$

holds for $x \in V$.

EXAMPLE 4.21 For each index i fixed, the functions x_j/x_i are regular on the basic open subset $D_+(x_i)$ of $\mathbb{P}^n$. ☆

DEFINITION *Let $X \subseteq \mathbb{P}^n$ be closed. We define the* structure sheaf *of X by setting, for each open $U \subset X$*

$$\mathcal{O}_X(U) = \{\, f : U \to k \mid f \text{ is regular for every point in } U \,\}.$$

Sums and products of regular functions are regular, and og clurse constants are regular, so $\mathcal{O}_X(U)$ is a k-algebra, and the restriction maps are k-algebra homomorphisms. Moreover, a regular function in $\mathcal{O}_X(U)$ is invertible if and only if it does not vanish at any point in U.

The Gluing axiom for $\mathcal{O}_X$ is easy to establish: if a regular function is given for each member of an open covering $\{U_i\}$ of an open subset of X and if they coincide on the intersection $U_i \cap U_j$, they patch together as continuous functions into $\mathbb{A}^1$, and since being regular is a local condition, the resulting function is regular (it restricts to regular functions on each of the members of the open covering $\{U_i\}$). Hence $\mathcal{O}_X$ is indeed a sheaf on X.

EXERCISE 4.10 Show in detail that $\mathcal{O}_X(U)$ is a k-algebra. ★

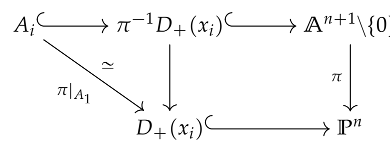
EXERCISE 4.11 Given an open $U \subseteq X$ and a continuous function $f\colon U \to \mathbb{A}^1$. Let $C_0(U)$ be punctured cone over U and denote by $\pi_U\colon C_0(U) \to U$ the (restriction of the) projection. Show that f is regular if and only if the composition $f \circ \pi$ is regular on $C_0(U)$. ★

4.23 It is of interest to compare regular functions on closed subsets of a closed projective set and on the surrounding set. At least locally there is a reasonable answer. Assume that $Y \subseteq X$ is a Zariski closed subset of the closed projective set X and that $U \subseteq X$ is open. The following lemma is almost tautological:

LEMMA 4.24 (RESTRICTION AND LOCAL EXTENSION) *If f is a regular function in the open set $U \subseteq X$ and $Y \subseteq X$ is closed, the restriction $f|_{U \cap Y}$ is regular in $Y \cap U$. Any functions on $U \cap Y$ regular at a point $p \in Y$ extends to a regular function on some open neighbourhood V of p in X.*

Projective varieties are varieties

In the end of this section we shall establish that projective varieties are varieties. For this, there will be two steps, and in the first we content ourself to see that they are *prevarieties*. The projective varieties have been equipped with a topology and a sheaf of rings, so what is left, is to check they are locally affine. To this end, we shall show that the basic open subsets $D_+(x_i)$ are isomorphic to affine n-space $\mathbb{A}^n$; more precisely, the regular functions x_j/x_i will serve as affine coordinates. This resolves the matter for the projective space itself, and for a closed subset $X \subseteq \mathbb{P}^n$ the sets $X \cap D_+(x_i)$ will do (closed subsets of affine space are affine).



4.25 In Paragraph 4.5 on page 74, we introduced the subvariety A_i of $\mathbb{A}^{n+1}$ where the i-th coordinate x_i equals one, and there we established that $\pi|_{A_i}$ is a homeomorphism between A_i and $D_+(x_i)$ (Proposition 4.19). Here we shall accomplish the description and prove it is an isomorphism of varieties. In Paragraph 4.5 we introduced the natural inverse map α_i, which sends a point a point $[x]$ with homogeneous coordinates $(x_0 : \cdots : x_n)$ to the point $\alpha_i([x]) = (x_0/x_i, \ldots, x_n/x_i)$ which obviously is independent of any scaling of the x_i's. The i-th coordinate equals one, and hence $\alpha_i([x])$ lies in A_i.

Proposition 4.26 *The projection $\pi|_{A_i}$ is an isomorphism between A_i and $D_+(x_i)$. The inverse map is the map α_i above; that is, trhe one with $\alpha_i(x_0 : \cdots : x_n) = (x_0/x_i, \ldots, x_n/x_i)$.*

In particular, this means that the basic open subset $D_+(x_i)$ is an affine n-space on which the n functions $x_0/x_i, \ldots, x_n/x_i$ serve as coordinates (of course, $1 = x_i/x_i$ is not included)

Proof: The two maps $\pi|_{A_i}$ and α_i are mutually inverse and both are homeomorphisms (Proposition 4.19 on page 79), so what is left, is to check that they are morphisms.

That $\pi|_{A_i}$ is a morphism is almost trivial: let f be regular at p and represent f in some open neighbourhood U of p as $f(x) = a(x)/b(x)$ with a and b homogeneous polynomials of the same degree and with b being non-zero throughout U. One simply has $f \circ \pi|_{A_i} = a/b|_{A_i}$, which is regular on $A_i \cap \pi^{-1}(U)$ as b does not vanish along $\pi^{-1}U$.

To prove the α_i is a morphism, let f be regular on an open set $U \subseteq A_i$ and represent f as $f = a/b$ with a and b being polynomials and b not vanishing in U. To obtain $f \circ \alpha_i$ one simply plugs in $x_j x_i^{-1}$ in the j-th slot (this automatically inserts a one in slot i) and one arrives at the expression

$$f \circ \alpha_i(x_0 : \cdots : x_n) = a(x_0 x_i^{-1}, \ldots, x_n x_i^{-1})/b(x_0 x_i^{-1}, \ldots, x_n x_i^{-1}).$$

We already observed that the fractions $x_j x_i^{-1}$ are regular throughout $D_+(x_i)$, and as the regular functions form a ring with non-vanishing functions being invertible, it follows that $f \circ \alpha_i$ is regular in U. ❑

4.27 This was the warm up for $\mathbb{P}^n$, and the general case of a projective variety is not very much harder – in fact it follows immediately:

Proposition 4.28 *Assume that $X \subseteq \mathbb{P}^n$ is an irreducible closed projective set. Then $V_i = D_+(x_i) \cap X$ equipped with the sheaf $\mathcal{O}_X|_{V_i}$ is an affine variety.*

Proof: Closed irreducible subsets of affine varieties are affine varieties. ❑

We have almost established the following very important theorem:

Theorem 4.29 *Irreducible, closed projective sets are varieties when endowed with the Zariski topology and the sheaf of regular functions.*

Proof: Let $X \subseteq \mathbb{P}^n$ be the set in question. The only thing that remains to be proven is that X satisfies the Hausdorff axiom. By Lemma 3.59 on page 63, if suffices to exhibit an open affine subset containing any two given points. This is no big deal: Given two distinct points in X, there is linear form λ on $\mathbb{A}^{n+1}$ that does not vanish at either. Hence both lie in the distinguished open subset $D_+(\lambda) \cap X$, which is affine by Proposition 4.28 above. ❑

The following corollary is, with Proposition 4.26 above in mind, merely an observation

Corollary 4.30 *Let F be a homogeneous form on $\mathbb{A}^{n+1}$. The open set $D_+(F) \cap D_+(x_i)$ is affine. When the coordinates x_j/x_i on $D_+(x_i)$ are used, $D_+(F) \cap D_+(x_i)$ corresponds to the distinguished open set $D(F^d)$ of $\mathbb{A}^n$ where $F^d = F(x_0/x_i, \dots, x_n/x_i)$.*

The affine open sets A_i allows us to do computations with the structure sheaf $\mathcal{O}_X$:

Example 4.31 On $X = \mathbb{P}^1$, we have $D_+(x_0) \simeq \mathbb{A}^1$, which has affine coordinate ring equal to a polynomial ring in one variable, so

$$\mathcal{O}_X(D_+(x_0)) = k[\frac{x_1}{x_0}]$$

Likewise, $\mathcal{O}_X(D_+(x_1)) = k[\frac{x_0}{x_1}]$. On the intersection, $D_+(x_0) \cap D_+(x_1) = D_+(x_0x_1)$, we have

$$\mathcal{O}_X(D_+(x_0x_1)) = k[\frac{x_1}{x_0}, \frac{x_0}{x_1}]$$

☆

The projective closure

4.32 The homogenization process described in Paragraph 4.18, which associated f^h with f, extends to ideals, and this yields an algebraic description of the closure in $\mathbb{P}^n$ of subvarieties of the basic open set $D + (x_i)$.

Assume that X is a subvariety of one of the basic open sets $D_+(x_i)$ and let $\mathfrak{a}$ be the ideal in $k[x_0/x_i, \ldots, x_n/x_i]$ defining it. Let $\mathfrak{a}^h$ be the homogeneous ideal

$$\mathfrak{a}^h = \{ f^h \mid f \in \mathfrak{a} \}$$

in $k[x_0, \ldots, x_n]$ where f^h denotes the homogenization of a polynomial f from $k[x_0/x_i, \ldots, x_n/x_i]$; that is, $f^h = x_i^d f$ where d is the degree of f.

LEMMA 4.33 *$Z_+(\mathfrak{a}^h)$ is the closure of X in $\mathbb{P}^n$.*

PROOF: One has $Z_+(f^h) \cap D_+(x_i) = Z(f)$ and hence

$$Z_+(\mathfrak{a}^h) \cap D_+(x_i) = \bigcap_{f \in \mathfrak{a}} Z(f^h) \cap D_+(x_i) = \bigcap_{f \in \mathfrak{a}} Z(f) = Z(\mathfrak{a}).$$

❑

EXAMPLE 4.34 *(Conics in the projective plane)* When trying to understand a variety in $\mathbb{P}^n$, it is sometimes useful to consider the different 'affine pieces' $X \cap D_+(x_i)$. As we shall see shortly, $D_+(x_i)$ is isomorphic to an affine space $\mathbb{A}^n$, and affine techniques apply to study of $X \cap D_+(x_i)$. And one may turn the sword around and study affine varieties by studying their closures in a projective space. There are stronger theorems about projective varieties, and this sometimes gives insight in the affine situation.

For example, this technique sheds considerable light on plane conics. The projective conic $xy - z^2 = 0$ becomes the *hyperbola* $(x/z)(y/z) = 1$ in the $\mathbb{A}^2$ which equals $D_+(z)$ and have x/z and y/z as coordinates, but it materializes as the *parabola* $y/x - (z/x)^2$ in $D_+(x)$, which is an $\mathbb{A}^2$ with coordinates y/x and z/x. So the difference between the hyperbola and the parabola is just a matter of perspective! They are both affine parts of the same projective curve. In other words and bearing Example 1.27 on page 17 in mind, *all conics are up to the choice of coordinates the same* when considered in the projective plane $\mathbb{P}^2$ over the complex numbers $\mathbb{C}$ (or any algebraically closed field). ☆

Global regular functions on projective varieties

One of the fundamental theorem of affine varieties states that the ring $\mathcal{O}_X(X)$ of global sections of the structure sheaf of an affine variety X — that is, the space of globally defined regular functions — is equal to the coordinate ring $A(X)$. This ring is quite large and in many ways completely determines the structure of the variety.

For projective varieties the situation is quite different. The only globally defined regular functions turn out to be the constants (Theorem 4.43 below).

True, one has the coordinate ring $S(X) = A(C(X))$ of the *cone* over X, but most elements there are not functions on X, not even the homogeneous ones, as they are not invariant under scaling.

4.35 By assumption X will be irreducible, and the same is then true for the cone $C(X)$. The ring $S(X) = A(C(X))$ is therefore an integral domain and has a fraction field which we shall denote by K. One calls $S(X)$ the *homogeneous coordinate ring* of X. For $X = \mathbb{P}^n$ we have $S(X) = k[x_0, \ldots, x_n]$, and if $X \subset \mathbb{P}^n$ is a projective variety, we have $S(X) = k[x_0, \ldots, x_n]/I(X)$.

Homogeneous coordinate rings
den homogene koordinatringen

Note that $S(X)$ is a graded ring because the ideal $I(X)$ is homogeneous, and thus it has a decomposition into homogeneous parts $S(X) = \bigoplus_{i \geqslant 0} S(X)_i$, where $S(X)_i$ denotes the subspace of elements of degree i. The fraction field K of $S(X)$ is not graded, but fractions of two homogeneous elements from $S(X)$ has a degree, namely $\deg ab^{-1} = \deg a - \deg b$.

Any regular function defined on some open set in $C(X)$ gives an element of K, and two such elements are equal as functions on an open if and only if they are the same element in K. The ground field k is contained in K as the constant functions on $C(X)$.

EXAMPLE 4.36 Consider the conic $X = Z_+(x_0x_2 - x_1^2)$ in $\mathbb{P}^2$. Its homogeneous coordinate ring is given by

$$S(X) = k[x_0, x_1, x_2]/(x_0x_2 - x_1^2).$$

We will soon show that X is in fact isomorphic to $\mathbb{P}^1$ (whose homogeneous coordinate ring equals $k[u_0, u_1]$). Thus unlike the case of affine varieties, the homogeneous coordinate ring $S(X)$ is dependent on the particular embedding into projective space. ☆

EXERCISE 4.12

a) Let R be a graded ring. Show that the set T consisting of the homogeneous elements in R is a multiplicative system and that the localization $T^{-1}R$ is a graded ring. Show $T^{-1}R$ is an integral domain when R is, and in that case, the homogeneous piece of degree zero $(T^{-1}R)_0$ is a field.
b) Let $R = k[x_0, x_1]$ and let T be the multiplicative system $T = \{ x_1^i \mid i \in \mathbb{N} \}$. Show that the homogeneous piece $(T^{-1}R)_0$ of degree zero of $T^{-1}R$ equals $k[x_0x_1^{-1}]$. Show furthermore that the decomposition of $T^{-1}R$ into homogeneous pieces is given as

$$T^{-1}R = \bigoplus_{i \in \mathbb{Z}} k[x_0x_1^{-1}] \cdot x_1^i.$$

c) If $X \subseteq \mathbb{P}^n$ is a projective variety, show with notation as in Problem 4.12, that the rational function field $k(X)$ equals $(T^{-1}S(X))_0$

★

4.37 While the elements of the homogeneous coordinate ring does not correspond to functions per se, it is quite convenient for describing the structure sheaf of a projective variety. The following proposition follows from the distinguished open sets:

Proposition 4.38 *Let $X \subset \mathbb{P}^n$ be a projective variety with homogeneous coordinate ring $S = S(X)$. Then*

$$\mathcal{O}_X(D_+(x_i)) = \left(S_{x_i}\right)_0$$

Example 4.39 Consider the cubic surface $X = Z_+(xw^2 - yz^2) \subset \mathbb{P}^3$. Then

$$\begin{aligned}\mathcal{O}_X(D_+(w)) &= \left(\frac{k[x,y,z,w]_w}{(xw^2 - yz^2)}\right)_0 \\ &= \left(\frac{k[x,y,z,w^{\pm 1}]}{(x - w^{-2}yz^2)}\right)_0 \\ &= \left(k[y,z,w^{\pm 1}]\right)_0 \\ &= k[y/w, z/w]\end{aligned}$$

☆

4.40 As a gentle beginning, let us consider the case of the projective space $\mathbb{P}^n$ itself. So let f be a global regular function on $\mathbb{P}^n$. Composing it with the projection $\pi\colon \mathbb{A}^{n+1}\backslash\{0\} \to \mathbb{P}^n$ we obtain a regular function $\pi^* f = f \circ \pi$ on $\mathbb{A}^{n+1}\backslash\{0\}$. In Example 3.11 on page 50, we showed that every regular function on $\mathbb{A}^{n+1}\backslash\{0\}$ is a polynomial, hence $\pi^* f$ is a polynomial. But $\pi^* f$ is also constant on lines through the origin, and therefore it must be constant. We thus arrive at the following:

Proposition 4.41 *The only global regular functions on $\mathbb{P}^n$ are the constants, i.e., $\mathcal{O}_{\mathbb{P}^n}(\mathbb{P}^n) = k$.*

For $\mathbb{P}^1$, we can see this result in light of the computations in Example 4.31. The restriction of a regular function $f\colon \mathbb{P}^1 \to k$ to $D_+(x_0)$ is a regular function on $\mathbb{A}^1$, i.e., a polynomial in x_1/x_0. Lilkewise $f|_{D_+(x_1)}$ is a polynomial in x_0/x_1. These two polynomials have to match up in $\mathcal{O}_X(D_+(x_0) \cap D_+(x_1)) = k[x_0/x_1, x_1/x_0]$ – but this is only possible if the polynomials are constants. So Proposition 4.41 basically comes from the fact that non-constant polynomials defined on $\mathbb{A}^1 \subset \mathbb{P}^1$ must blow up at the point at infinity.

One may compare this result to Liouville's theorem in complex analysis: any holomorphic function $f\colon \mathbb{C}\mathbb{P}^1 \to \mathbb{C}$ on the Riemann sphere restricts to a holomorphic function $f|_{\mathbb{C}}\colon \mathbb{C} \to \mathbb{C}$ which, being holomorphic at ∞, must be bounded, and hence f is constant.

4.42 For a general projective variety the same conclusion holds, but it is somewhat more difficult to prove. One has

Theorem 4.43 *The only globally defined regular functions on a projective variety $X \subseteq \mathbb{P}^n$ are the constants. In other words, it holds true that $\mathcal{O}_X(X) = k$.*

PROOF: Let f be a global regular function on X. Composed with the projection, it gives a global regular function $\pi^* f$ on the punctured cone $C(X)\setminus\{0\}$ which we shall denote by F. Note that F is an element in the function field K of $C(X)$.

For each index i denote by D_i the basic open set $D(x_i)$ of the cone $C(X)$ where $x_i \neq 0$. We know that the coordinate ring $A(D_i)$ satisfies $A(D_i) = T_i^{-1}S(X)$ where T_i is the multiplicative system $\{ x_i^r \mid r \in \mathbb{N} \}$, so that for each index i the function F has a representation $F = g_i/x_i^{r_i}$ for some $g_i \in S(X)$ and some natural number r_i. As the function F is constant along lines through the origin, it must be homogeneous of degree zero; in other words, g_i is homogeneous and $\deg g_i = r_i$.

So we have $x_i^{r_i} F = g_i$, and the salient point is that g_i lies in the homogeneous part $S(X)_{r_i}$. It follows that $hx_i^{r_i}F \in S(X)_{r_i+j}$ for all elements h of $S(X)$ that are homogeneous of degree j.

Now, chose an integer r so big that $r > \sum_i r_i$. Then any monomial M of degree r contains at least one of the variables, say x_i, with an exponent larger than r_i, and consequently $MF \in S(X)_r$. In other words, multiplication by F leaves the finite dimensional vector space $S(X)_r$ invariant. It is a general fact (for instance, use the Cayley-Hamilton theorem), that F then satisfies a relation of the type

$$F^m + a_{m-1}F^{m-1} + \ldots + a_1 F + a_0 = 0$$

where the a_i are elements in the ground field k. This shows that F, which is an element of the function field K, is algebraic over k, and since k is algebraically closed by assumption, F lies in k and is constant. ❑

4.44 An important consequence of the theorem is that morphisms of projective varieties into affine varieties are necessarily constant. Indeed, if $X \subseteq \mathbb{P}^n$ is projective and $Y \subseteq \mathbb{A}^m$ is affine, the component functions of a morphism $\phi\colon X \to Y \subseteq \mathbb{A}^m$ must all be constant according to the theorem. Hence we have:

COROLLARY 4.45 *Any morphism from a projective variety to an affine one is constant.*

Another consequence is the following:

COROLLARY 4.46 *A variety X which is both projective and affine, is a point.*

PROOF: The coordinate functions are regular functions on a subvariety $X \subseteq \mathbb{A}^n$, and according to the theorem they must be constant when X is projective. ❑

4.4 *Morphisms of quasi-projective varieties*

When it comes to morphisms between affine varieties the picture is completely clear: the main theorem (Theorem 3.48 on page 60) tells us they are just given as k-algebra homomorphism between the coordinate rings; or if the target variety is contained in the affine space $\mathbb{A}^n$, a morphism is simply given by n regular functions on the source.

The picture is less clear when one considers morphisms between projective or quasi-projective varieties. Many morphisms however, are easily defined as set-theoretical maps, and for experienced geometers it is pretty obvious that they are morphisms, but at least once in a lifetime one should check it in detail. So we offer a little simple lemma in that direction.

Rational maps
rasjonale avbidninger

Several of the examples we shall present are examples of the important notion of *rational maps* between two varieties. Such 'maps' are not genuine globally defined morphism; they are only defined on open subsets, but still they convey a great amount of information about the relationship between the two varieties. The rational function on a variety are all examples of rational maps with values in $\mathbb{A}^1$. A rational map ϕ between two varieties X and Y will be indicated by a dashed arrow $\phi\colon X \dashrightarrow Y$, and not a solid none, as a reminder that it is not a globally defined morphism.

Birational maps
birasjonale avbildninger
Birationally equivalent
birasjonal ekvivalens

A rational map having a rational inverse is called *birational*, and the two varieties are said to be *birationally equivalent*. The study of varieties up to birational equivalence has during certain historic periods been an important part of algebraic geometry (and still is, though to a less extent), and we shall give them a deeper treatment in Chapter 6.

A simple criterion for being a morphism

4.47 Let X and Y be two quasi-projective varieties and let $\phi\colon X \to Y$ be a continuous map. Assume we know that ϕ fits into a diagram shaped like

$$\begin{array}{ccc} C_0(X) & \xrightarrow{\Phi} & C_0(Y) \\ {\scriptstyle \pi_X}\downarrow & & \downarrow{\scriptstyle \pi_Y} \\ X & \xrightarrow[\phi]{} & Y \end{array} \tag{4.3}$$

where $C_0(X)$ and $C_0(Y)$ are the punctured cones over respectively X and Y, and with the two vertical maps being the usual projections, and where Φ is a morphism; in other words, ϕ lifts to a morphism between the punctured cones. Then it follows that ϕ is a morphism. Usually it is not difficult to check this case by case, but it is worthwhile doing it once and for all, hence the following little lemma:

Lemma 4.48 *With the setting as in the diagram (4.3) above, the map ϕ is a morphism.*

Proof: Assume that X is open in a subvariety of the projective space $\mathbb{P}^n$ and that $(x_0 : \ldots : x_n)$ are homogenous coordinates in $\mathbb{P}^n$. Then the sets $U_i = D_+(x_i) \cap X$ form an open covering of X, and it suffices to see that each restriction $\phi|_{U_i}$ is a morphism.

Earlier, in Paragraph 4.5, we defined a natural map $\alpha_i\colon D_+(x_i) \to \mathbb{A}^{n+1}\backslash\{0\}$, by setting $\alpha_i(x_0 : \ldots : x_n) = (x_0/x_i, \ldots, x_n/x_i)$. It is a section of the canonical projection π over $D_+(x_i)$ whose basic property is being an isomorphism between $D_+(x_i)$ and the closed subvariety A_i of $\mathbb{A}^{n+1}$ where $x_i = 1$. Restricted to the open subset U_i of X the section α_i gives an isomorphism $\beta_i\colon U_i \simeq A_i \cap C_0(X)$ satisfying $\pi_X \circ \beta_i = \mathrm{id}_X|_{U_i}$. It follows that

$$\begin{array}{ccccc} A_i \cap C_0(X) & \hookrightarrow & C_0(X) & \xrightarrow{\Phi} & C_0(Y) \\ \beta_i \uparrow \simeq & & \pi_X \downarrow & & \downarrow \pi_Y \\ U_i & \hookrightarrow & X & \xrightarrow[\phi]{} & Y \end{array}$$

$$\phi|_{U_i} = \phi \circ \pi_X \circ \beta_i = \pi_Y \circ \Phi \circ \beta_i,$$

and since the three maps to the right are morphisms, $\phi|_{U_i}$ is one as well. ❑

4.49 The main example to have in mind is when the map Φ is given as $\Phi(x) = (f_0(x), \ldots, f_m(x))$ with the components f_i's being homogeneous polynomials of the same degree. The set X can be the open subset of $\mathbb{P}^n$ where the f_i's do not vanish simultaneously; that is, $X = \mathbb{P}^n\backslash Z_+(f_0, \ldots, f_m)$, or any quasi-projective set contained therein. And Y might be the entire projective space $\mathbb{P}^m$, or any quasi-projective subvariety Y so that the cone $C_0(Y)$ contains the image of Φ. In all cases ϕ will be a rational map $\mathbb{P}^n \dashrightarrow Y$.

On X, the morphism Φ descends to the map $\phi([x]) = (f_0(x) : \ldots : f_m(x))$ between the quasi-projective varieties X and Y. Because the f_i's all have the same degree, say d, it holds true that

$$(f_0(tx), \ldots, f_n(tx)) = (t^d f_0(x), \ldots, t^d f_n(x))$$

for any non-zero scalar t, and therefore $\phi([x])$ does not depend on the representative of $[x]$. Moreover, the homogeneous coordinates $(f_0(x) : \ldots : f_m(x))$ are legitimate because in X not all of the f_i's vanish at the same time.

4.50 It may happen that different morphisms Φ fit into the diagram (4.3) paired with the same map ϕ: the components of Φ may be changed by a common factor without altering the map ϕ. Notice that the set where ϕ is defined; that is, the closed set X in diagram (4.3), also is susceptible to change. It might grow, or it might shrink, depending on the behaviour of the common factor. Certainly, a common factor might introduce common zeros, in which case the set X will shrink. But the situation might also improve. When the f_i's are rational functions, the morphism Φ is not defined where one of them has a pole, but multiplying through by the least common multiple of their denominators, will yield a lifting Φ whose components are polynomials and thus extends ϕ beyond the set of poles. We will see a few examples of this in the next sections.

Linear projections

As one can guess from the names, projections are central players in projective geometry, and they are prominent examples of rational maps coming out of the scenario of the little lemma.

4.51 A *projection* is a surjective, rational map $\pi\colon \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ which in the staging of the little lemma 4.48 is induced by a surjective, linear map $\Phi\colon \mathbb{A}^{n+1} \to \mathbb{A}^{m+1}$. It is common usage to reserve the name *projection* for the cases when that $m < n$, since if $m = n$ the map Φ will be an isomorphism (it is assumed to be surjective) and one would rather call ϕ a *linear isomorphism* or a *linear automorphism* of $\mathbb{P}^n$. The set where the projection π is not defined, is the projective linear space $\mathbb{P}(\ker \Phi)$ corresponding to the kernel of Φ. This subspace is usually called the *centre* of the projection.

Projection *projeksjoner*

Centre of a projection *projeksjonssenter*

For a closed subvariety $X \subseteq \mathbb{P}^n$ it may ceranly happen that $\mathbb{P}\ker(\Phi)$ meets X, so that π is not defined on the whole X. But in the case that X is not entirely contained in $\mathbb{P}(\ker \phi))$, one has a subvariety of $\mathbb{P}^m$ which reasonably can be called the projection of X: the restriction $\pi|_X$ is well-defined on a non-empty open $U \subseteq X$. The closure of $\pi|_X(U)$ in the target $\mathbb{P}^m$ will be the *projection* of X.

Projection of a subvariety *projektsjonen av en undervarietet*

The archetype of a projection is induced by the map $\mathbb{A}^{n+1} \to \mathbb{A}^{m+1}$ that forgets some $n - m$ of the coordinates, say the last ones. The projection is then described by the assignment $(x_0 : \ldots : x_n) \mapsto (x_0 : \ldots : x_m)$, and the common zeros of the components is the linear subspace V where the $m+1$ first coordinates vanish; that is, where $x_0 = \ldots = x_m = 0$.

The simplest example is the projection from a point, where a projection simply forgets one of the coordinates, *e.g.* sending $(x_0 : \cdots : x_n)$ to $(x_0 : \cdots : x_{n-1})$, which is well defined away from the point $(0\colon \ldots : 0\colon 1)$.

It is common practice to identify the target space $\mathbb{P}^m$ with the linear subvariety of $\mathbb{P}(W)$ of $\mathbb{P}^n$, where $W \subseteq \mathbb{A}^{n+1}$ is given by the equations $x_{m+1} = x_{m+2} = \ldots = x_n = 0$. Note, that this a complementary subspace to the centre V; that is, $k^{n+1} = V \oplus W$.

The geometric interpretation of the projection from a point p onto a $\mathbb{P}(W)$, is as follows. Take a point in $\mathbb{P}^n$ not in the centre $\mathbb{P}(V)$; that is, a one-dimensional subspace L of k^{n+1} not lying within V. The subspace $V + L$ of k^{n+1} spanned by V and L intersects W in a line. and $\pi(L)$ is that intersection. This follows from the classical dimension formula from linear algebra which yields, since $W + V + L = k^{n+1}$, that $\dim(V + L) \cap W = \dim(V + L) + \dim W - (n+1) = 1$.

In particular, if one projects from a point p, the target variety is a hyperplane on which p does not lie, and the image of a point $x \in \mathbb{P}^n$ is the intersection of the line through x and p with H.

4.52 When one wants to study a projective variety X by means of projections, it is of course important to be able to describe the projected variety. If the variety X is given by ideal ideal $I \subset k[x_0, \ldots, x_n]$, this involves computing the *elimination ideal* $I \cap k[x_0, \ldots, x_m]$, which can be quite a difficult task in general. However, if X is given on parametric form, it is trivial to describe the projection. Let us give a few examples.

Examples

4.53 *(Automorphisms of $\mathbb{P}^1$)* As explained in in Paragraph 4.51 above every linear automorphism of k^{n+1} induces an automorphism of $\mathbb{P}^n$. Two proportional automorphisms induce the same map on $\mathbb{P}^n$, so it is the projective linear group $\mathrm{PGL}(n+1, k)$ acting.

Later in the course (Theorem 12.1 on page 239) we shall see that $\mathrm{PGL}(n+1, k)$ comprises all the automorphisms of $\mathbb{P}^n$; however, with the technical arsenal so far developed, we can at present merely accomplish this for the projective line. We are to show that each automorphismis $g\colon \mathbb{P}^1 \to \mathbb{P}^1$ is induced by a linear coordinate shift. The image of the singleton $Z_+(x_1)$ is a singleton, and with appropriate coordinats $(y_0 : y_1)$ on the target $\mathbb{P}^1$ we may assume it equals $Z_+(y_1)$. Then g maps $D_+(x_1)$ isomorphically onto $D_+(y_1)$, and by 2.54 the restriction $g|_{D_+(x_0/x_1)}$ is induced by an algebra-isomorphism $k[x_0/x_1] \to k[y_0/y_1]$. Now, such an isomorphism is determined by the action on the generator x_0/x_1 which must be send to a *linear*[4] expression $ay_0/y_1 + b$, and we are through: the coordinate shift $y_1 \mapsto y_1$ and $y_0 \mapsto ay_0 + by_1$ does the job.

[4] If α and β is a pair of mutually inverse maps between $k[t]$ and $k[u]$, write $\alpha(t) = at^n + l.t.$ and $\beta(u) = bt^m + l.t.$ (where $l.t$ is shorthand for 'lower terms'). Then $t = \beta(\alpha(t)) = a^m bt^{nm} + l.t.$

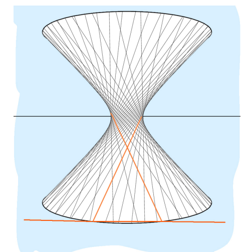
4.54 *(Projecting the twisted cubic)* The twisted cubic C is the image of $\mathbb{P}^1$ under the map $(u : v) \to (u^3 : u^2v : uv^2 : v^3)$ (which is a morphism according to Lemma 4.48). Consider the projection from $\mathbb{P}^3$ to $\mathbb{P}^2$ with centre $(0 : 0 : 0 : 1)$ which just forgets the last coordinate; note that the centre lies on C. The image of a point $(u^3 : u^2v : uv^2 : v^3)$ is $(u^3 : u^2v : uv^2)$ when $u \neq 0$, but when $u = 0$ the point coincides with the centre and the projection is not defined. However, one may discard the common factor u and obtain a genuine parameterization $(u : v) \to (u^2 : uv : v^2)$ of the projection of C. Thus the projection of the twisted cubic is the conic with equation $y^2 = xz$.

Observe that in thia case the projection decreases the degree by one, which is due to the fact that the centre of projection lies on C. Notice also that the parameterization of the conic is a globally defined on $\mathbb{P}^1$, so that the projection, which *a priori* is merely a rational map, extends to a global morphism.

4.55 *(The twisted cubic again)* We continue with the twisted cubic, but change the centre of the projection to $(0 : 0 : 1 : 0)$; that is, the projection forgets the third coordinate. The effect on a point $(u^3 : u^2v : uv^2 : v^3)$ is to send it to the point $(u^3 : u^2v : v^3)$ in $\mathbb{P}^2$. This time the projection is defined all along $\mathbb{P}^1$, and one easily checks that the equation of the image is $y^3 = x^2z$; that is, the image is the well-known cuspidal cubic curve. Notice that the degree is conserved, and that the image aquires a singular point.

4.56 *(The quadric in $\mathbb{P}^3$)* In this example we project the quadric Q in $\mathbb{P}^3$ with equation $xy - zw = 0$ from the point $p = (0 : 0 : 0 : 1)$, which lies on Q. The lines $z = x = 0$ and $z = y = 0$ both lie on Q, and they intersect in the point p. The entire first line is mapped to the point $p_1 = (0 : y : 0) \in \mathbb{P}^2$ and the second to the point $p_2 = (x : 0 : 0)$. So each of these two lines are collapsed to a point.

Off these two lines the projection is one-to-one: If $q = (x : y : z)$ is a point in $\mathbb{P}^2$ with $z \neq 0$, there is exactly one point on the quadric Q projecting to q, namely $(x : y : z : xy/z)$.



To summarize, projecting a quadric Q from a point p on it collapses the two lines L_1 and L_2 on Q passing through p to two different points p_1 and p_2 in $\mathbb{P}^2$. The projection induces an isomorphism from $Q \backslash L_1 \cup L_2$ to $\mathbb{P}^2 \backslash L$; that is, $\mathbb{P}^2$ without the line L through p_1 and p_2. So the projection is an example of a birational map.

Note that the image of $Q \backslash \{p\}$ includes the two points p_1 and p_2 as well and equals $(\mathbb{P}^2 \backslash L) \cup \{p_1, p_2\}$. It is neither open nor closed.

☆

Exercises

4.13 Show that each rational function in one variable defines a rational map $\mathbb{P}^1 \dashrightarrow \mathbb{P}^1$ which extends to a morphism $\mathbb{P}^1 \to \mathbb{P}^1$. HINT: Express the rational function as $P(t)/Q(t)$ with P and Q polynomials without common factors and consider the assignment $(x_0 : x_1) \mapsto (x_0^d P(x_1/x_0) : x_0^d Q(x_1/x_0))$ where $d = \max(\deg P, \deg Q)$.

4.14 Let the projection $\mathbb{P}^3$ to $\mathbb{P}^2$ be given by the assignment $(x : y : z : w) \mapsto (x : x + z : w + y)$. Determine the centre and describe the projection of the twisted cubic parametrized as $(u : v) \mapsto (u^3 : u^2v : uv^2 : v^3)$. HINT: The key words is "rational node" (see Problem 3.20 on page 70).

4.15 Find points in $\mathbb{P}^4$ such that the projection of the rational normal quartic $(u^4 : u^3v : u^2v^2 : uv^3 : v^4)$ projects onto a twisted cubic.

4.16 Describe (by giving an equation) the image of the rational normal quartic under the projection $\mathbb{P}^4 \dashrightarrow \mathbb{P}^2$ that forgets the third and the forth coordinate. Accomplish the same task but with the projection that forgets the second and the forth coordinate.

4.17 Let C_d be the rational normal curve in $\mathbb{P}^d$ whose parametrization is

$$\phi_d(u : v) = (u^d : u^{d-1}v : \ldots : uv^{d-1} : v^d).$$

Let $\pi \colon \mathbb{P}^d \backslash \{q\} \to \mathbb{P}^{d-1}$ be the projection with centre $q = (0 : 0 : \ldots : 0 : 1)$. Prove that $q \in C_d$ and that the closure in $\mathbb{P}^{d-1}$ of $\pi(C_d \backslash \{q\})$ is equal to C_{d-1}.

★

Chapter 5

Segre and Veronese varieties

TOPICS IN CHAPTER 5: *Closed embeddings – rational normal curves – Segre embeddings and Segre varieties – the homogeneous ideals of the Segre varieties – Veronese embeddings and Veronese varieties – the homogeneous ideals of the Veronese varieties – conics and the Veronese surface – bihomogeneous ideals – geometry of polynomials*

So far we have produced and described examples of closed subvarieties of affine or projective space by giving explicit equations. There is another important way of obtaining and studying closed subvarieties, namely by means of so-called *closed embeddings*. In precise terms, it is a morphism $\iota\colon X \to Y$ having a closed image and inducing an isomorphism between X and the image $\iota(X)$ (here, of course, $\iota(X)$ is endowed with the variety structure as a closed subvariety of the ambient space Y, as in Paragraph 3.40 on page 58).

Closed embeddings
lukkede embeddinger

The *Segre* and the *Veronese varieties* are two important families of projective varieties defined by natural closed embeddings. They are omnipresent in algebraic geometry. In addition to furnishing non-trivial examples of projective varieties, they are of a great theoretical interest, not only in algebraic geometry, but in several other areas of mathematics.

Is is not too hard to write down the Segre and Veronese maps and show they are closed embeddings, but to determine their homogeneous prime ideals is more delicate, and illustrates well the subtlety of determining if a given ideal is prime.

Mainly for pedagogical reasons, we treat the rational normal curves separately before proceeding with the Segre and the Veronese embeddings, even though they are a special kind of Veronese varieties. At the end of the chapter we have included a section on bihomogeneous polynomials of great use when working with closed subvarieties of the products $\mathbb{P}^n \times \mathbb{P}^m$.

5.1 Closed embeddings

Working with closed embeddings is a welcome counterpart to working with equations. It can be extremely difficult to work out the geometry of vanishing locus of an ideal, but on the other hand, finding the equations for the image of a closed embedding can also be quite challenging. However, it is fair to say that the embedding approach is the more fruitful, as one can utilize both geometric and algebraic arguments.

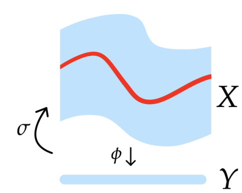
5.1 As the next lemma shows, being a closed embedding is a property of a morphism local on the target. This can be quite useful when checking that a given morphism is a closed embedding, and we'll apply it both for the Segre and Veronese embeddings.

Lemma 5.2 *If $\phi\colon X \to Y$ is a morphism and $\{U_i\}_{i\in I}$ is a family of open sets in Y that covers the closure $\overline{\phi(X)}$ of the image, then ϕ is a closed embedding if and only the restriction $\phi|_{\phi^{-1}(U_i)}\colon \phi^{-1}(U_i) \to U_i$ is a closed embedding for each i.*

Proof: If ϕ is a closed embedding, it is straightforward to verify that all the restrictions $\phi|_{\phi^{-1}(U_i)}$ are closed embeddings, so our task is to prove the converse.

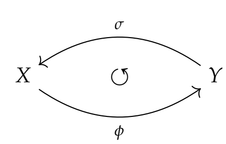
Without loss of generality, we may assume $Y = \overline{\phi(X)}$. Assume that a covering $\{U_i\}$ as in the lemma is given. The map ϕ is then obviously injective. Next, we verify it is a closed map, so let $Z \subseteq X$ be closed. It will suffice to see that each intersection $\phi(Z) \cap U_i$ is closed in U_i since a subset of a topological space is closed if and only it meets every member of an open covering in a closed set. But $\phi(Z) \cap U_i = \phi(Z \cap \phi^{-1}(U_i))$ which is closed in U_i because the restriction $\phi|_{\phi^{-1}(U_i)}$ is assumed to be a homeomorphism. So ϕ is closed map (in particular the image $\phi(X)$ will be closed) and hence is a homeomorphism between Y and $\phi(Y)$.

It remains to see that the inverse map $(\phi|_{\phi(X)})^{-1}$ is a morphism of varieties. To that end, let $U \subseteq X$ be an open set and f a regular function on U. We must check that $f \circ \phi^{-1}|_{\phi(U)}$ is regular on $\phi(U)$. But being a regular function is a local property, and U may be replaced by $U \cap \phi^{-1}(U_i)$. But in that case $f \circ \phi^{-1}|_{\phi(U)}$ is regular simply because $\phi|_{\phi^{-1}(U_i)}$ is assumed to be an isomorphism so that $\phi^{-1}|_{\phi(U)}$ is a morphism. ❑



5.3 In some examples, one finds that a morphism into a projective space locally admit left sections that are projections. This will be the case both for the Veronese and the Segre embeddings, and it is worthwhile to settle both of these cases with one lemma. The proof is close to being trivial, and the lemma is a special case of a much more general lemma (which we give as an exercise, see Problem 5.2 below). We begin with recalling the notion of a section.

Given a map $\phi\colon X \to Y$. A map $\sigma\colon Y \to X$ the other way around is said to be a *left section* of ϕ if $\sigma \circ \phi = \mathrm{id}_X$ and one says that σ is a *right section* if $\phi \circ \sigma = \mathrm{id}_Y$. When ϕ is a morphism, we will also require σ to be a morphism.



$\sigma \cdot \phi = \mathrm{id}_X$

Sections of morphisms
seksjoner til morfier

Lemma 5.4 *Assume that the morphism $\phi\colon \mathbb{A}^n \to \mathbb{A}^{n+m}$ has a left section which is a projection. Then ϕ is a closed embedding.*

Proof: Since the section σ is a projection, coordinates $(x_1, \ldots, x_{n+m})$ on $\mathbb{A}^{n+m}$ may be chosen such that $\sigma(x_1, \ldots, x_{n+m}) = (x_1, \ldots, x_n)$. In these coordinates

$\phi(\mathbb{A}^n)$ appears as the graph of the morphism $\psi\colon \mathbb{A}^n \to \mathbb{A}^m$ being the composition $\pi \circ \phi$ where $\pi(x_1, \ldots, x_{n+m}) = (x_{1+n}, \ldots, x_{n+m})$. Hence $\phi(\mathbb{A}^n)$ is closed, and one easily verifies that $\sigma|_{\phi(\mathbb{A}^n)}$ serves as an inverse to ϕ. ❑

Exercises

5.1 Assume X and Y are two affine varieties and that $\phi\colon X \to Y$ is a morphism. Show that ϕ is a closed embedding if and only if the map $\phi^*\colon A(Y) \to A(X)$ between the coordinate rings is surjective. HINT: The Main Theorem for Affine Varieties (Theorem 3.48 on page 60) may be useful.

5.2 Let X and Y be two varieties and $\iota\colon X \to Y$ a morphism that has a left section σ. The aim of the exercise is to show that ι is a closed embedding.

a) Show that one may assume that X and Y are affine.
b) When X and Y are affine, show that ι is a closed embedding. HINT: σ^* will be a right section for ι^*, then cite Problem 5.1.

5.3 Being a closed embedding is a not a local property on the source. Consider the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^2$ given as $\phi(t) = \big(t^2 - 1, t(t^2 - 1)\big)$.

a) Show that ϕ is a closed map, but not a closed embedding.
b) Exhibit an open covering $\{U_i\}$ of $\mathbb{A}^1$ such each restriction $\phi|_{U_i}$ is a closed embedding into some open subset V_i of $\mathbb{A}^2$.

5.4 Assume that the ground field k is of positive characteristic 2. Show that the morphism $\mathbb{P}^1 \to \mathbb{P}^2$ that sends $(x_0 : x_1)$ to $(x_0^4 : x_0^2 x_1^2 : x_1^4)$ is a closed map which is a homeomorphism onto its image, but which is not a closed embedding.

★

5.2 *Rational normal curves*

The *rational normal curves* are by far the most frequently met Veronese varieties, possibly only contested by the famous Veronese surface. Their theory is simpler than the general one, and their treatment parallels both that of the general Segre and the general Veronese embeddings, so we find them well suited to serve as an introduction to the whole story.

The affine version

5.5 Already in the first chapter of these notes (in Problem 1.11 on page 19) we met the affine versions of the rational normal curves. There they were describe by the following parameterizations $\phi_d \colon \mathbb{A}^1 \to \mathbb{A}^d$, one for each natural number d:

$$\phi_d(u) = (u, u^2, \dots, u^d). \tag{5.1}$$

With coordinates $z_1, \dots, z_d$ on the affine space $\mathbb{A}^d$ one was asked to show that image $C_d = \phi_d(\mathbb{A}^1)$ is given by the equations $z_{i+1} - z_i z_1$ for $1 \leqslant i \leqslant d-1$, but it is easy to see that the equations $z_i - z_1^i$ for $i = 2, \dots, d$ work as well. Another way of phrasing this, is to say that the projection onto the first factor of $\mathbb{A}^d$ — the one that sends $(z_1, \dots, z_d)$ to z_1 — serves as a left section for ϕ_d. Thus in light of Lemma 5.4, we conclude that ϕ_d is a closed embedding.

The projective version

5.6 The projective version of the *rational normal curve* of degree d is a curve in projective space $\mathbb{P}^d$, still denoted C_d, that has a parametrization $\Phi_d \colon \mathbb{P}^1 \to \mathbb{P}^d$ given as

Rational normal curves
rasjonale normale kurver

$$\Phi_d(x_0 : x_1) = (x_0^d : x_0^{d-1}x_1 : \dots : x_0^{d-i}x_1^i : \dots : x_0 x_1^{d-1} : x_1^d).$$

The components being of the same degree and without common zeros, this is a morphism (Paragraph 4.49 on page 90), and we shall verify it is a closed embedding with the help of Lemma 5.2. Let $t_0, \dots, t_d$ be homogeneous coordinates on $\mathbb{P}^d$ so that Φ_d is defined by $t_i = x_0^{d-i}x_1^i$.

The open covering required by Lemma 5.2 will, naturally, consist of the distinguished open sets $D_+(t_i)$. These are positioned in slightly different ways relative to the normal curve. The two opens $D_+(t_0)$ and $D_+(t_d)$ cover C_d, so if the image C_d is *a priori* known to be closed[1], these two will suffice. At present, however, we lack this insight and need the full covering.

5.7 The distinguished open set $D_+(t_0)$ of $\mathbb{P}^d$ has $D_+(x_0)$ as inverse image under Φ_d, and the restriction $\Phi_d|_{D_+(x_0)}$ is nothing but the parametrization ϕ_d of the affine normal curve described in (5.1) above: setting $u = x_1 x_0^{-1}$ and $z_i = t_i t_0^{-1}$ makes this clear. Thus $\Phi_d|_{D_+(x_0)}$ is a closed embedding. Likewise, is it as clear that the inverse image of the distinguished open set $D_+(t_d)$ equals $D_+(x_1)$ and that the restriction $\Phi_d|_{D_+(x_1)}$ is a closed embedding (just reorder the coordinates in Paragraph 5.5 above).

5.8 The inverse images under Φ_d of the other distinguished open sets $D_+(t_\nu)$ of $\mathbb{P}^d$ — that is, those with $0 < \nu < d$ — all equal the distinguished open set $D_+(x_0 x_1)$. This is the set $\mathbb{A}^1 \backslash \{0\}$ where both $u = x_0/x_1$ and $u^{-1} = x_1/x_0$ acts as a coordinate. To determine the restriction[2] of Φ_d to $D_+(x_0 x_1)$ we invert the ν-the coordinate $t_\nu = x_0^{d-\nu}x_1^\nu$ and find for the i-th affine coordinate:

$$t_i/t_\nu = x_0^{\nu-i}x_1^{i-\nu}.$$

The coordinates with $0 \leqslant i < \nu$ yield the map $\mathbb{A}^1 \to \mathbb{A}^\nu$, which is given by the

[1] Images of projective varieties under morphisms are alway closed. This is a general theorem which we shall prove later (Theorem 10.18 on page 197).

[2] This is of course independent of d, but in order to apply Lemma 5.2 we need the image to be closed in each distinguished open set $D_+(t_i)$.

assignment

$$u \mapsto (u^{\nu}, u^{\nu-1}, \dots, u),$$

where $u = x_0 x_1^{-1}$. We recognize the parametrization ϕ_v of the affine normal curve C'_ν, but with components in reverse order. Those with $\nu < \iota \leqslant d$ make up the map $\mathbb{A}^1 \to \mathbb{A}^{d-\nu}$ given by

$$v \mapsto (v, v^2, \dots, v^{d-\nu}),$$

where $v = x_1 x_0$, and this is just the parameterization $\phi_{\nu-d}$ with image $C'_{\nu-d}$. Finally, we lie hands on the the restriction of Φ_d as the composition

$$D_+(x_0x_1) = \mathbb{A}^1 \backslash \{0\} \xrightarrow{\quad \eta \quad} \mathbb{A}^1 \times \mathbb{A}^1 \xrightarrow{\phi_v \times \phi_{d-\nu}} \mathbb{A}^\nu \times \mathbb{A}^{d-\nu} = D_+(t_\nu),$$

where $\eta \colon \mathbb{A}^1 \backslash \{0\} \to \mathbb{A}^1 \times \mathbb{A}^1 = \mathbb{A}^2$ is the map $u \mapsto (u, u^{-1})$. The image of $\phi_v \times \phi_{d-\nu}$ equals $C'_\nu \times C'_{d-\nu}$ which, being the product of two closed sets, is closed, and the image of η is closed as well. As sets closed in a closed subset are closed in the ambient space, we conclude that the restriction of Φ_d to $\Phi_d^{-1}(D_+(t_\nu))$ has closed image in $D_+(t_\nu)$.

The homogeneous ideal

5.9 When it comes to determining the homogeneous prime ideal $I(C_d)$ of the rational normal curve in the ring $R = k[t_0, \dots, t_d]$, we enter slightly delicate ground. Clearly all the quadratic polynomials $t_i t_j - t_l t_m$ with indices satisfying $i + j = l + m$, lie in $I(C_d)$, and in fact, as we'll prove in the next section, they generate the ideal $I(C_d)$. For convenience, we let $\mathfrak{a}$ denote the ideal in R generated by the quadrics $t_i t_j - t_l t_m$, and proceed to show that $I(C_d) = \mathfrak{a}$.

The two terms become respectively $x_0^{2d-(i+j)} x_1^{i+j}$ and $x_0^{2d-(l+m)} x_1^{l+m}$ when evaluated along C_d

The ideal $I(C_d)R_{t_0}$ in the localization R_{t_0} of R (where t_0 is inverted), equals the prime ideal of the affine rational normal curve $I(C_d \cap D_+(t_0))$ in $\mathbb{A}^d$. This ideal is generated by the above quadrics (indeed, the quadrics $z_i - z_{i+1} z_1$ in the coordinates $z_i = t_i t_0^{-1}$ are among them), and *mutatis mutandis* one checks that this holds true also when localizing in the other variables t_i. However — and here the delicacy shows up — this will not be sufficient to conclude since it does not exclude that

$$\mathfrak{a} = \mathfrak{q} \cap I(C_d),$$

where $\mathfrak{q}$ is a $(t_0, \dots, t_d)$-primary ideal. So we need to come up with a new idea.

PROPOSITION 5.10 *The homogeneous prime ideal $I(C_d)$ of the rational normal curve C_d in $\mathbb{P}^d$ is generated by the quadrics $t_i t_j - t_l t_m$ for all $i + j = l + m$.*

PROOF: The new idea is to exhibit a *normal form* modulo $\mathfrak{a}$ for any monomial in R. Specifically, every monomial in the t_i's may, modulo the ideal $\mathfrak{a}$, be brought on the form $t_s t_0^a t_d^b$ where $0 < s < d$ and a and b are non-negative integers. Indeed, among the quadrics in $\mathfrak{a}$ we find either $t_i^2 - t_0t_{2i}$ or $t_i^2 - t_dt_{2i-d}$ according to $2i \leqslant d$ or $2i > d$. Hence the square of any variable other than t_0 or t_d may, modulo $\mathfrak{a}$, be replaced by a product containing either t_0 or t_d, and by induction on the number of variables in the monomial different from t_0 and t_d, we are through.

Finally, the parameterization Φ_d on the level of homogeneous coordinate rings factors as the map $R/\mathfrak{a} \to k[x_0, x_1]$ that sends the class of t_i to $x_0^{d-i}x_1^i$. It follows that different classes of monomials in $R/\mathfrak{a}$ map to different monomials in $k[x_0, x_1]$; indeed, the monomial $t_st_0^at_d^b$ maps to $x_0^{ad+d-s}x_1^{bd+s}$ where a, b and s are unambiguously defined as respectively quotients and remainder upon division of the exponents of x_0 and x_1 by d.

A generating set of the k-vector space $R/\mathfrak{a}$ (the classes of the monomials) thus maps injectively into a linearly independent set of element in $k[x_0, x_1]$ (the monomials); consequently the map $R/\mathfrak{a} \to k[x_0, x_1]$ is injective, and so $\mathfrak{a}$ is a prime ideal. ❑

5.11 The quadrics in Proposition 5.10 above do not form a minimal generating set for the ideal $I(C_d)$ — there are far too many. A nice minimal set is described in the following proposition, which also realizes the rational normal curves as determinantal varieties.

Proposition 5.12 *The 2×2-minors of the matrix*

$$\begin{pmatrix} t_0 & t_1 & \dots & t_i & \dots & t_{d-1} \\ t_1 & t_2 & \dots & t_{i+1} & \dots & t_d \end{pmatrix}$$

form a minimal generating set for the ideal $I(C_d)$.

Proof: The quadrics $t_it_j - t_{i+1}t_{j-1}$ appear among the minors of the matrix. To see that also the other quadrics in Proposition 5.10 do, we resort to induction on $\max\{|i-j|, |l-m|\}$. Clearly we may assume that $i \leqslant j$ and $l \leqslant m$ and that $m - l \leqslant j - i$; in case of equality the constraint $i + j = l + m$ yields that $j = m$ and $i = l$, and we are finished. So we may assume that $m - l < j - i$. Then the equality

$$t_it_j - t_lt_m = (t_it_j - t_{i+1}t_{j-1}) + (t_{i+1}t_{j-1} - t_lt_m),$$

brings us through; indeed, since $m - l < j - i$, it is straightforward to see that $j - i \geqslant 2$, and then it holds that $|(i+1) - (j-1)| = |i - j + 2| = j - i - 2 < |j - i|$.

That the minors form a minimal set of generators follows from a dimension count. The map $\psi\colon R_2 \to k[x_0, x_1]_{2d}$ (where as usual subscripts indicate homogeneous parts of a given degree) induced from the algebraic incarnation $\Phi_d^*\colon R \to k[x_0, x_1]$ of the embedding, is surjective, so the dimension of its kernel is given as

$$\dim \ker \psi = \binom{d+2}{2} - (2d+1) = \binom{d}{2},$$

which is precisely the number of minors of the matrix in the proposition. ❑

EXAMPLE 5.13 Consider the twisted cubic $Y = C_3 \subset \mathbb{P}^3$ given by the three quadrics

$$q_0 = t_0t_2 - t_1^2, \quad q_1 = t_0t_3 - t_1t_2, \quad q_2 = t_1t_3 - t_2^2.$$

The ideal $I(Y)$ is not what one calls a *complete intersection*, since the ideal is not minimally generated by the same number of equations (three) as the codimension (two). On the other hand, we may write Y as the intersection of just two surfaces: it holds true that $Y = Z_+(f, g)$ where $f = q_0$ and $g = t_2q_2 - t_3q_1$. Indeed, the ideal (f, g) contains

Complete intersections
komplette snitt

$$q_1^2 = -t_2^2 f - t_0 g \quad \text{and} \quad q_2^2 = -t_3^2 f - t_2 g,$$

so $\sqrt{(f, g)} = I(Y)$. We say that Y is a *set-theoretic complete intersection*. ☆

Set-theoretic complete intersections
mengdeteoretiske komplett snitt

EXERCISE 5.5 Let $M_0, \ldots, M_d$ be a basis for the vector space $k[x_0, x_1]_d$ of homogeneous forms in x_0 and x_1 of degree d. Define a map by the assignment

$$\begin{aligned} \phi : \mathbb{P}^1 &\to \mathbb{P}^d \\ (x_0 : x_1) &\mapsto (M_0(x_0, x_1) : \ldots : M_d(x_0, x_1)). \end{aligned}$$

Show that there is a linear automorphism A of $\mathbb{P}^d$ so that $\psi = A \circ \Phi_d$. Conclude that $\psi(\mathbb{P}^1)$ is projectively equivalent to the rational normal curve C_d. ★

A geometric property

5.14 Any $d + 1$ points $p_0, \ldots, p_d$ on the rational normal curve are linearly independent. Indeed, we may assume that $p_i = \Phi_d(q_i)$ for $q_i = (1 : x_i) \in \mathbb{P}^1$. The coordinates of p_i appear as the rows of the matrix

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^d \\ 1 & x_2 & x_2^2 & \ldots & x_2^d \\ 1 & x_3 & x_3^2 & \ldots & x_3^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \ldots & x_m^d \end{pmatrix}$$

This is exactly the Vandermonde matrix with determinant $\prod_{i<j}(x_i - x_j)$, which is non-zero, so the points are linearly independent. We say that a collection of points in $\mathbb{P}^d$ are in *linearly general position* if no $d + 1$ of them lie in a hyperplane, which is a geometers preferred term for the points being linearly independent.

Linearly general position
lineært generell posisjon

5.15 The next proposition is a converse of the observation above, and it generalizes the classical fact that there is a unique conic through 5 linearly general points in $\mathbb{P}^2$.

Proposition 5.16 *Through any $d+3$ linearly general points in $\mathbb{P}^d$, there passes a unique rational normal curve.*

Proof: Let $a_0, \ldots, a_d, b_0, \ldots, b_d$, be non-zero and general elements from k, which in this context means that none[3] of the minors of the matrix

$$\begin{pmatrix} a_0 & a_1 & \ldots & a_d \\ b_0 & b_1 & \ldots & b_d \end{pmatrix} \tag{5.2}$$

[3] This means *all* minors, both the 2×2-ones and the 1×1-ones;*i.e.* the entries themselves

vanishes, and define polynomials

$$M_i = \frac{(x_0/a_0 - x_1/b_0) \cdot \cdots \cdot (x_0/a_d - x_1/b_d)}{(x_0/a_i - x_1/b_i)}, \qquad i = 0 \ldots d, \tag{5.3}$$

each of degree d. It holds that $M_j(a_i, b_i) = 0$ when $j \neq i$, and $M_i(a_i, b_i) \neq 0$ since all minors of the matrix (5.2) are non-zero. It ensues that the M_i's are linearly independent and hence form a basis for $k[x_0, x_1]_d$: if $\sum_{j=0}^{d} \lambda_i M_i = 0$ is a linear relation, then evaluating at $(a_i : b_i)$ gives $\lambda_i = 0$ for all i.

Now, the map $\Phi : \mathbb{P}^1 \to \mathbb{P}^d$ defined as in 5.5 using the polynomials in (5.2) sends the point $(a_i : b_i)$ to the i-th coordinate point e_i; that is:

$$\Phi(a_i : b_i) = (0 : \ldots : 1 : \ldots : 0),$$

where the 1 is located in the i-th slot. Evaluating Φ at the points $(0 : 1)$ and $(1 : 0)$ and noting that the enumerator will be a non-zero common factor for all the values of respectively the $M_i(0,1)$'s and the $M_i(1,0)$'s, we find

$$\Phi(1 : 0) = (a_0 : \ldots : a_d)$$
$$\Phi(0 : 1) = (b_0 : \ldots : b_d)$$

Finally, let $p_0, \ldots, p_{d+2}$ be $d+3$ points in linearly general position. Any $d+1$ of them are lineraøy independent, so we can after applying a linear coordinate change assume that $p_0, \ldots, p_d$ are the coordinate points $e_0, \ldots, e_d$. The condition that the p_i's are linearly general, means precisely that no minor of the $2 \times (d+1)$-matrix formed by their coordinates vanishes.[4] This means that we may feed the coordinates of p_{d+1} and p_{d+2} into the above construction and hence find the desired map Φ.

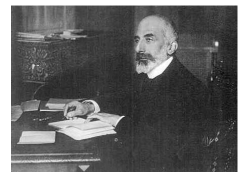
[4] This is seen by replacing the i-th and the j-th row in the $(d+1) \times (d+1)$ identity matrix with the coordinate vectors of p_{d+1} and p_{d+2}; the determinant of the resulting matrix does not vanish as any $d+1$ of the points are linearly independent).

It remains to verify that there is only one rational normal curve passing by the points $p_0, \ldots, p_{d+2}$. We leave that as an exercise. ❑

Exercise 5.6 Show that curve constructed above is unique. Hint: If a map Φ is given by polynomials N_i passes by the coordinate points and the points $(a_0 : \cdots : a_d)$ and $(b_0 : \cdots : b_d)$, the N_i's must be a scalar multiple of the M_i's. ★

5.3 The Segre varieties

The second kind of closed embeddings we shall describe are named after one of the great Italian geometers Corrado Segre. They are closed embeddings of the products $\mathbb{P}^n \times \mathbb{P}^m$ of two projective spaces into the projective spaces $\mathbb{P}^{nm+n+m}$. These products are thus projective, and consequently we get the important corollary that products of projective varieties are projective. We shall show that the images—the so-called Segre varieties—have ideals generated by certain minors of certain matrices, and establish a Nullstellensatz-like result describing the closed subsets of $\mathbb{P}^n \times \mathbb{P}^m$ as vanishing loci for ideals generated by so-called bihomogeneous forms.



Corrado Segre (1863–1924) Italian mathematician.

5.17 With $(x_0 : \ldots : x_n)$ and $(y_0 : \ldots : y_m)$ being homogeneous coordinates on the projective spaces $\mathbb{P}^n$ and $\mathbb{P}^m$ respectively, the *Segre maps* (also called *Segre embeddings* as they will turn out to be embeddings) are the maps whose component functions are all possible products of shape $x_i y_j$. In other words, they are the maps

Segre maps
Segre-avbildninger

$$\sigma_{n,m} \colon \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{nm+n+m}$$

Segre embeddings
Segre-embeddinger

defined by the assignements

$$\sigma_{n,m}([x],[y]) = (x_0 y_0 : \ldots : x_i y_j : \ldots : x_n y_m), \tag{5.4}$$

where $[x] = (x_0 : \ldots : x_n)$ and $[y] = (y_0 : \ldots : y_m)$ and the $x_i y_j$'s are listed in some order; to fix the ideas, we order them lexicographically. This presentation depends heavily on the choices of coordinates, but a canonical (and hence more abstract and less intuitive) approach avoiding these choices is sketched at the end of the section.

The definition of the Segre map $\sigma_{n,m}$ is legitimate since a simultaneous scaling of either the x_i's or the y_j's results in a simultaneous scaling of the products $x_i y_j$. Moreover, for every point of $\mathbb{P}^n \times \mathbb{P}^m$ at least one of the x_i's and one the y_j's do not vanish, so neither does their product vanish, and in (5.4) at least one component is non-zero. That $\sigma_{n,m}$ is a morphism is then clear from the observation in Paragraph 4.49 on page 90. The image of $\sigma_{n,m}$ is called a *Segre varity* and is denoted by $S_{n,m}$. And more generally, any subvariety projectively equivalent to $S_{n,m}$ goes under the name of a Segre variety.

Segre varieties
Segre-variteter

5.18 The double indexing is an invitation to represent the coordinates of the points in $S_{n,m}$ as the entires of an $(n+1) \times (m+1)$-matrix, *i.e.* the product of the column vector formed by the x_i's and the row vector formed by the y_j's:

$$(x_0, \ldots, x_n)^t (y_0, \ldots, y_m) = \begin{pmatrix} x_0 y_0 & x_0 y_1 & \ldots & x_0 y_m \\ x_1 y_0 & x_1 y_1 & \ldots & x_1 y_m \\ \vdots & \vdots & & \vdots \\ x_n y_0 & x_m y_1 & \ldots & x_n y_m \end{pmatrix}. \tag{5.5}$$

Subsequently, this leads us to introduce homogeneous coordinates t_{ij} in the projective space $\mathbb{P}^{nm+n+m}$ indexed by pairs i, j with $0 \leqslant i \leqslant n$ and $0 \leqslant j \leqslant m$ thinking about the t_{ij}'s as entries of the following $(n+1) \times (m+1)$-matrix M:

$$M = \begin{pmatrix} t_{00} & t_{01} & \dots & t_{0m} \\ t_{10} & t_{11} & \dots & t_{1m} \\ \vdots & \vdots & & \vdots \\ t_{n0} & t_{n1} & \dots & t_{nm} \end{pmatrix}. \tag{5.6}$$

Such a matrix whose entries are variables (or equivalently, elements algebraically independent over the ground field) is called a *generic matrix*.

Generic matrices
generiske matriser

It is clear that the matrix in (5.5) has rank one. Hence all the 2×2-minors of the generic matrix M in (5.6) vanish along the Segre variety $S_{n,m}$: at points in $S_{n,m}$ it holds true that $t_{ij} = x_i y_j$, and thus evaluating M there yields the matrix in (5.5), which we noticed is of rank one. In this way we obtain a collection of quadratic hypersurfaces $t_{ij}t_{lm} - t_{im}t_{lj}$ containing $S_{n,m}$. In fact, we shall see that the Segre variety is precisely the locus where those minors vanish, and even the stronger assertion that the homogeneous prime ideal $I(S_{n,m})$ is generated by these minors, holds true. Finally, $\sigma_{n,m}$ will be an isomorphism between $\mathbb{P}^n \times \mathbb{P}^m$ and $S_{n,m}$.

$$\begin{pmatrix} & \vdots & & \vdots & \\ \dots & t_{ij} & \dots\dots & t_{im} & \dots \\ & \vdots & & \vdots & \\ & \vdots & & \vdots & \\ \dots & t_{lj} & \dots\dots & t_{lm} & \dots \\ & \vdots & & \vdots & \end{pmatrix}$$

The embedding result

5.19 We have come to our main result in this section, that the Segre maps are closed embeddings.

> **Proposition 5.20** *The Segre map $\sigma_{n,m}$ is a closed embedding of the product $\mathbb{P}^n \times \mathbb{P}^m$ into $\mathbb{P}^{nm+n+m}$. The image of $\sigma_{n,m}$ is the locus where the all the 2×2-minors of the matrix M in (5.6) vanish.*

Proof: To ease the notation we will write σ for the Segre map $\sigma_{n,m}$.

Fix a pair of indices μ and ν with $0 \leqslant \mu \leqslant n$ and $0 \leqslant \nu \leqslant m$, and introduce the open subset $U = D_+(x_\mu) \times D_+(y_\nu)$ of the product $\mathbb{P}^n \times \mathbb{P}^m$ together with the distinguished open subset $D = D_+(t_{\mu\nu})$ of the projective space $\mathbb{P}^{nm+n+m}$. Since the relation $t_{\mu\nu} = x_\mu y_\nu$ holds, the coordinate $t_{\mu\nu}$ is non-zero if and only if both x_μ and y_ν are non-zero, and so the set U is the inverse image of D under σ. In view of Lemma 5.2 on page 96 it will therefore suffice to prove that the restriction $\sigma|_U$ is a closed embedding of U into D. Now, U is isomorphic to the affine space $\mathbb{A}^n \times \mathbb{A}^m$ which we equip with coordinates $x_i x_\mu^{-1}$ and $y_j y_\nu^{-1}$, the indices i and j running through the relevant[5] values. The distinguished open subset D is isomorphic to the affine space $\mathbb{A}^{mn+n+m}$, where we have the coordinates $t_{ij}t_{\mu\nu}^{-1}$. Moreover, the restriction of the Segre map to U, when expressed in the

[5] That is, $0 \leqslant i \leqslant n$ and $0 \leqslant j \leqslant m$, but $i \neq \mu$ and $j \neq \nu$.

coordinates $t_{ij}t_{\mu\nu}^{-1}$, takes the form

$$t_{ij}(\sigma([x],[y]))t_{\mu\nu}(\sigma([x],[y]))^{-1} = x_i y_j x_\mu^{-1} y_\nu^{-1}.$$

Setting $i = \mu$ in this relation, we recover $y_j y_\nu^{-1}$ and simillarily, setting $j = \nu$ we retrieve $x_i x_\mu^{-1}$. This means that the restriction $\sigma|_U$ is a map

$$\sigma|_U \colon \mathbb{A}^{n+m} \to \mathbb{A}^{nm+n+m}$$

that has a projection as a section; namely the one that forgets all coordinates but $t_{i\nu}t_{\mu\nu}^{-1}$ and $t_{\mu j}t_{\mu\nu}^{-1}$ for appropriate values of i and j. By Lemma 5.4 on page 96 it follows that $\sigma|_U$ is a closed embedding, and consequently σ is one as well.

It remains to prove that the Segre variety $S_{n,m}$ equals the locus where all the 2×2-minors of M vanish. To this end, consider such a point. At least one of its coordinates, say $t_{\mu\nu}$, is non-zero. Putting $x_i = t_{i\nu}t_{\mu\nu}^{-1}$ and $y_j = t_{\mu j}t_{\mu\nu}^{-1}$ for the pertinent values of i and j, one finds, since the minor $t_{ij}t_{\mu\nu} - t_{\mu j}t_{i\nu}$ vanishes at the point, that $x_i y_j = (t_{i\nu}t_{\mu\nu}^{-1}) \cdot (t_{\mu j}t_{\mu\nu}^{-1}) = t_{ij}t_{\mu\nu}^{-1}$, which precisely means that the point under consideration lies in $S_{n,m}$. ❑

$$\begin{pmatrix} & \vdots & & \vdots & \\ \cdots & t_{ij} & \cdots\cdots & t_{i\nu} & \cdots \\ & \vdots & & \vdots & \\ & \vdots & & \vdots & \\ \cdots & t_{\mu j} & \cdots\cdots & t_{\mu\nu} & \cdots \\ & \vdots & & \vdots & \end{pmatrix}$$

5.21 A direct application of the Segre maps being closed embeddings is the following result about products of projective varieties.

Corollary 5.22 *The product of two projective varieties is projective.*

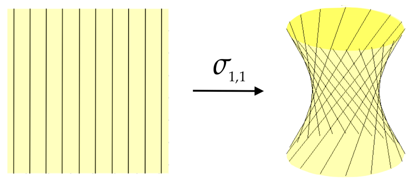
Proof: Let the two projective varieties be X and Y with $X \subset \mathbb{P}^n$ and $Y \subset \mathbb{P}^m$. The Zariski topology on a product is stronger than the product topology, and hence the product $X \times Y$ is closed in $\mathbb{P}^n \times \mathbb{P}^m$. The topological space $X \times Y$ carries both the sheaf of regular functions considered a product variety and the sheaf of regular functions considered a closed subvariety of $\mathbb{P}^n \times \mathbb{P}^m$, but of course, the two coincide. By the proposition $\mathbb{P}^n \times \mathbb{P}^m$ is isomorphic to a closed subvariety of $\mathbb{P}^{mn+n+m}$ via the Segre embedding, and consequently the product $X \times Y$ is as well isomorphic to a closed subvariety. ❑

Exercise 5.7 Verify that the two sheaves alluded to in the proof are equal. Hint: This is a local question. ★

Example 5.23 *(The quadric in $\mathbb{P}^3$)* The Segre variety $S_{n,m}$ will be a hypersurface only when $n = m = 1$, and in this case it will be a quadratic surface in $\mathbb{P}^3$. Indeed, the matrix M becomes the 2×2-matrix

$$\begin{pmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{pmatrix}$$

having the quadratic polynomial $t_{00}t_{11} - t_{01}t_{10}$ as only minor. Using a more commonplace notation, one recognizes this as the surface with equation $x_0x_3 - x_1x_2 = 0$. ☆



Determinantal varieties
determinantvarieteter

Generic determinantal varieties
generiske determinantvarieteter

5.24 The fact that the Segre varieties are cut out by the 2×2-minors of the generic matrix M in (5.6) places them in the broader perspective. They are members of the larger class of the so-called *determinantal varieties*. These are closed subvarieties whose equations are the minors of a given size of some matrix. In case the matrix is generic, as the matrix M in (5.6) is, the variety is said to be a *generic determinantal variety*.

For each triple $r \leqslant n \leqslant m$ of integer one has the closed subset $W_{r,n,m}$ of $\mathbb{P}^{nm+n+m}$ where the $(r+1) \times (r+1)$-minors of the generic $(n+1) \times (m+1)$-matrix M vanish. There is also an affine version, namely the cone $C(W_{r,n,m})$ in $\mathbb{A}^{(n+1)(m+1)}$ over $W_{r,n,m}$. The matrices belonging to $W_{r,n,m}$ are those whose rank is at most r, which explains the notation, and our Segre variety $S_{n,m}$ equals $W_{1,n,m}$.

The homogeneous ideal $I(S_{n,m})$

5.25 The last assertion in Proposition 5.20 may be paraphrased by saing that $Z_+(\mathfrak{a}) = S_{n,m}$, where $\mathfrak{a}$ is the ideal generated by the 2×2-minors of the matrix M, but allthough it turns out to be true in the end, the stronger statement that $I(S_{n,m}) = \mathfrak{a}$, is not at all *a priory* clear. It is not hard to see that the ideals of the intersections $S_{n,m} \cap D_+(t_{ij})$ in the affine spaces $D_+(t_{ij})$ are generated by the minors, but this is not enough since there might be an embedded component at the origin — so further effort is needed.

It is a general result that the ideals $I(W_{r,n,m})$ of polynomials vanishing along a generic determinantal variety are generated by the appropriate minors. In the general case this is a rather deep issue, and the result has the status as a theorem, frequently referred to as *The second fundamental theorem in invariant theory*. However, our present case of rank one maps and 2×2-minors, is not too complicated, and to the benefit of the curious students (and to stimulate their appetite for this classic area of mathematics) we offer a proof.

Proposition 5.26 *The ideal $I(S_{n,m})$ of the Segre-varity $S_{n,m}$ is generated by the 2×2-minors of the generic $(n+1) \times (m+1)$-matrix M in* (5.6).

Proof: It largely suffices to prove that the ideal $\mathfrak{a}$ generated by the 2×2-minors of M is a prime ideal.

In order to simplify the notation we introduce some abbreviations and write $A = k[t_{00}, \ldots, t_{nm}]/\mathfrak{a}$ and $R = k[x_0, \ldots, x_n, y_0, \ldots, y_m]$. On the level of homogeneous coordinate rings the Segre map materializes as the map from A to R which for each pair of indices i, j sends the class of t_{ij} to $x_i y_j$. The proof consists of showing that this map is injective, which certainly will imply that the ideal $\mathfrak{a}$ is prime.

The crucial observation is that since $t_{ij}t_{\mu\nu} - t_{i\nu}t_{\mu j}$ is a minor of M, the products $t_{ij}t_{\mu\nu}$ and $t_{i\nu}t_{\mu j}$ have the same image in A for all relevant indices i, j, μ and ν. In a product of two of the t_{ij}'s the second indices of the factors may thus be swapped without changing the class mod $\mathfrak{a}$.

Since any permutation can be expressed as a composition of transpositions, this means that the image of two monomials $t_{i_1 j_1} \cdot \ldots \cdot t_{i_r j_r}$ and $t_{i_1 j'_1} \cdot \ldots \cdot t_{i_r j'_r}$ where $(j'_1, \ldots, j'_r)$ is any permutation of $(j_1, \ldots, j_r)$, are equal in A. In other words, the class of a monomial $t_{i_1 j_1} \cdot \ldots \cdot t_{i_r j_r}$ is independent of the order of the indices – the i_t's and the j_t's may be independently permuted (the i_t's by a mere reordering of the factors) without the class changing.

From this ensues that the Segre map $A \to R$ is injective. Indeed, a monomial $t_{i_1 j_1} \cdot \ldots \cdot t_{i_r j_r}$ maps to $x_{i_1} \ldots x_{i_r} y_{j_1} \ldots y_{j_r}$ so if two have identical images in R, it follows from the polynomial ring R being a UDF that the sets of indices of the two are equal up to order, and combining the observation above with the fact that distinct monomials $x_{i_1} \ldots x_{i_r} y_{j_1} \ldots y_{j_r}$ are linearly independent in the polynomial ring R, we are through. ❑

The coordinate free approach

There is a coordinate free way of defining the Segre varieties which is not only worthwhile to describe, but in several contexts even is the preferable way of looking at the Segre varieties. The coordinate free approach links the Segre varieties to the tensor product of two vector spaces, and many properties of tensors have counterparts in geometric properties of the Segre varieties.

5.27 We start with two vector spaces E and F over k respectively of dimension n and m. Recall from the theory of multilinear algebra that there is a canonical bilinear map $E \times F \to E \otimes_k F$ which sends a pair (v, w) to the decomposable tensor $v \otimes w$. This map is a 'universally bilinear map' in the sense that any bilinear map $E \times F \to G$ factors through $E \otimes_k F$ by a linear map in a unique way. On the level of affine cones the Segre map is just this map: the set of decomposable tensors is identical to the cone over the Segre variety in the projective space $\mathbb{P}(E \otimes_k F)$.

To recover our earlier description of $S_{n,m}$, which depends on the choice of coordinates, chose bases $\{e_i\}$ and $\{f_j\}$ for E and F respectively, and write out the coordinates of a decomposable tensor in the induced basis $\{e_i \otimes f_j\}$ of $E \otimes F$. If $v = \sum_i x_i e_i$ and $w = \sum_j y_j f_j$, one finds that $v \otimes w = \sum_{i,j} x_i y_j e_i \otimes f_j$ by bilinearity, which fits with the definition of the Segre map—the coordinates of $v \otimes w$ are all possible products of the coordinates of v and those of w.

5.28 The manifestation of the Segre varieties as determinantal varieties appears in this context when E is replaced by the dual space E^* of linear functionals on E. A fundamental result in multilinear algebra asserts that there is a canonical isomorphism $E^* \otimes_k F \simeq \hom[k]EF$. And under this identification, decomposable tensors correspond to maps $E \to F$ of rank one: to a decomposable tensor $\phi \otimes w \in E^* \otimes_k F$ one associates the rank one map which is the composition

$$E \xrightarrow{\phi} k \xrightarrow{\iota_w} F$$

of the functional ϕ with the inclusion ι_w that sends a scalar α to $\alpha \cdot w$.

EXERCISE 5.8 *(A little linear algebra.)* Let E and F be two finite dimensional vector spaces over k.

a) Show that the map $E^* \times F \to \hom[k]EF$ sending a pair (ϕ, w) to the map $x \to \phi(x)w$ is bilinear and hence induces a linear map $E^* \otimes_k F \to \hom[k]EF$.
b) Chose a basis $\{e_i\}$ for E and let $\{\hat{e}_i\}$ denote the dual basis for E^* (the one with $\hat{e}_i(e_j) = \delta_{ij}$). Define a map $\hom[k]EF \to E^* \otimes_k F$ by the assigenment

$$\phi \mapsto \sum_i \hat{e}_i \otimes \phi(e_i).$$

Show that this map is a k-linear and serves an inverse to the map in point 5.8). In particular, the sum $\sum_i \hat{e}_i \otimes \phi(e_i)$ does not depend on the chosen basis.
c) The *rank* of the tensor $\tau \in E^* \otimes_k F$ is the least number of terms necessary to express τ as a sum of indecomposable tensors. Show that under the isomorphism $E^* \otimes_k F \simeq \hom[k]EF$ tensors of rank r correspond to linear maps of rank r.

The rank of tensors
rangen til en tensor

★

5.4 *A Nullstellensatz for* $\mathbb{P}^n \times \mathbb{P}^m$

5.29 Just like it is meaningful to speak about homogenous polynomials vanishing at points in projective space, it is meaningful to speak about bihomogeneous polynomials vanishing at points in the product $\mathbb{P}^n \times \mathbb{P}^m$. Indeed, because it holds true that $f(sx, ty) = s^a t^b f(x, y)$, if $f(x, y) = 0$ for one choice of homogeneous coordinates of $x = (x_0 : \cdots : x_n)$ and of $y = (y_0 : \cdots : y_m)$ for a point $x \times y$ in the product, it vanishes for all choices.

By the same reasoning it is meaningful to speak about *all* polynomials in a bihomogeneous ideal $\mathfrak{a}$ vanishing at a point. Hence the set $Z_+(\mathfrak{a})$ of points in $\mathbb{P}^n \times \mathbb{P}^m$ where all $f \in \mathfrak{a}$ vanish, is a well-defined subset of $\mathbb{P}^n \times \mathbb{P}^m$. And in the same manner, the ideal $I(Z)$ generated by all bihomogeneous polynomials that vanish along a subset $Z \subseteq \mathbb{P}^n \times \mathbb{P}^m$ is a well-defined ideal, so the main ingredients of a Nullstellensatz are available also in the present setting.

There is however one significant difference from the case of projective space. Instead of just one forbidden ideal, the irrelevant ideal $\mathfrak{m}_+$, there are now several. The ideals $\mathfrak{n}_+ = (x_0, \dots, x_n)$ and $\mathfrak{m}_+ = (y_0, \dots, y_m)$ are both *forbidden* since neither all the homogeneous coordinates of the first component of the point nor all those of the second are allowed to vanish. The two ideals are prime ideals, but they are not maximal, and any ideal containing one of them is forbidden.

Forbidden ideals
forbudte idealer

5.30 We shall lean on the Segre embedding (Proposition 5.20) to see that the projective Nullstellensatz for the ambient space $\mathbb{P}^{nm+n+m}$ entails a Nullstellensatz for the product $\mathbb{P}^n \times \mathbb{P}^m$. There is just one small obstacle. Since the substitutions $t_{ij} = x_i y_j$ are linear in both the x_i's and the y_j's, restriction of homogeneous polynomials $F(t_{00}, \dots, t_{nm})$ to $S_{n,m}$ yields bihomogeneous polynomials of the same degree in x_i's and y_j's; they are *balanced* as one says. However, all bihomogeneous polynomials are of course not balanced, so most ideals in $k[x_0, \dots, x_n, y_0, \dots, y_m]$ are not induced from homogeneous ideals in $k[t_{00}, \dots, t_{nm}]$.

Balanced bihomogeneous polynomials
balanserte bihomogene polynomer

To remedy this defect, we replace a bihomogeneous ideal $\mathfrak{a}$ by the ideal $\mathfrak{a}'$ generated by all balanced polynomials contained in $\mathfrak{a}$; the important point being that $\mathfrak{a}$ and $\mathfrak{a}'$ have the same zero locus: one inclusion is obvious, namely that $Z_+(\mathfrak{a}) \subseteq Z_+(\mathfrak{a}')$. Moreover, if $f \in \mathfrak{a}$ is bihomogeneous of bidegree (a, b), according to a or b being the larger, either all products $y_0^{a-b} f, \dots, y_m^{a-b} f$ or all products $x_0^{b-a} f, \dots, x_n^{b-a} f$ belong to $\mathfrak{a}'$, and since at any point of the product $\mathbb{P}^n \times \mathbb{P}^m$ at least one of the x_i's and one of y_j's do not vanish, it ensues that $Z_+(\mathfrak{a}') \subseteq Z_+(\mathfrak{a})$.

In fact, one has the stronger statement that the ideals $\mathfrak{a}$ and $\mathfrak{a}'$ share radicals; *i.e.* $\sqrt{\mathfrak{a}} = \sqrt{\mathfrak{a}'}$. Indeed, letting s be a sufficiently large integer[6] it holds true that $\mathfrak{m}^s \mathfrak{a} \subseteq \mathfrak{a}'$, with $\mathfrak{m}$ denoting the ideal $(x_0, \dots, x_n, y_0, \dots, y_m)$, and since every proper bihomogeneous ideal is contained in $\mathfrak{m}$, it follows that $\mathfrak{a}^{s+1} \subseteq \mathfrak{m}^s a \subseteq \mathfrak{a}'$ and hence

$$\sqrt{\mathfrak{a}} = \sqrt{\mathfrak{a}^s} \subseteq \sqrt{\mathfrak{a}'}.$$

Thus $\sqrt{\mathfrak{a}} = \sqrt{\mathfrak{a}'}$ as the other inclusion is obvious.

[6] Larger that $\max(n, m) \cdot |a - b|$ for all pairs (a, b) that occur as bidegrees of members of a set of generators for $\mathfrak{a}$.

THEOREM 5.31 (NULLSTELLENSATZ FOR A PRODUCT) *The assignments $Z_+(\mathfrak{a})$ and $I(Z)$ set up mutually inverse one-to-one correspondences between nonempty, closed subsets $Z \subseteq \mathbb{P}^n \times \mathbb{P}^m$ and radical, bihomogeneous ideals $\mathfrak{a} \subseteq k[x_0, \dots, x_n, y_0, \dots, y_m]$ not containing either of the forbidden ideals $\mathfrak{m}_+$ and $\mathfrak{n}_+$.*

PROOF: Just like for the original Nullstellensatz, the only point that needs a serious argument is the equality $I(Z_+(\mathfrak{a})) = \sqrt{\mathfrak{a}}$. Since both equalities $I(Z_+(\mathfrak{a})) = I(Z_+(\mathfrak{a}'))$ and $\sqrt{\mathfrak{a}} = \sqrt{\mathfrak{a}'}$ hold true, it will certainly suffice to prove the theorem for the case of $\mathfrak{a}'$; that is, for ideals generated by balanced polynomials. But in that case our theorem is easily reduced to the projective Nullstellensatz; indeed, let $\mathfrak{b}$ be a homogeneous ideal in $k[t_{00}, \ldots, t_{nm}]$ that induces $\mathfrak{a}'$, *i.e.* one such that $(\mathfrak{b} + \mathfrak{q})/\mathfrak{q} = \mathfrak{a}'$ where $\mathfrak{q}$ is the ideal of the Segre variety $S_{n,m} = \mathbb{P}^n \times \mathbb{P}^m$ in $\mathbb{P}^{nm+n+m}$. Consequently the zero locus $Z_+(\mathfrak{b} + \mathfrak{q})$ in $\mathbb{P}^{nm+n+m}$ equals the zero locus $Z_+(\mathfrak{a}')$ in $\mathbb{P}^n \times \mathbb{P}^m$. The projective Nullstellensatz yields the equality $I(Z_+(\mathfrak{b} + \mathfrak{q})) = \sqrt{(\mathfrak{b} + \mathfrak{q})}$, but clearly $\sqrt{(\mathfrak{b} + \mathfrak{q})/\mathfrak{q}} = \sqrt{\mathfrak{a}'}$, and we are through. ❑

Example 5.32 Consider points $(y_0 : y_1) \times (x_0 : x_1)$ in $\mathbb{P}^1 \times \mathbb{P}^1$. The prinsipal ideal (y_0) is not balanced, and the corresponding balanced ideal $\mathfrak{a}'$ equals (x_0y_0, x_1y_0). The zero locus of y_0 is the set $A = p \times \mathbb{P}^1$ with $p = (0 : 1)$; *i.e.* of points shaped like $(0 : 1) \times (x_0 : x_1)$. The zero loci of x_0y_0 and x_1y_0 are of the form $B_i = p \times \mathbb{P}^1 \cup \mathbb{P}^1 \times q_i$ where $q_0 = (0 : 1)$ and $q_1 = (1 : 0)$. Clearly $B_0 \cap B_1 = A$. ☆

Exercise 5.9 *(Nullstellensatz for a general product.)* Consider the product $\mathbb{P}^{n_1} \times \mathbb{P}^{n_2} \times \ldots \times \mathbb{P}^{n_r}$ of r projective spaces, and let $(x_{i0} : \ldots : x_{in_i})$ be homogeneous coordinates on $\mathbb{P}^{n_i}$. A polynomial $F(x_{ij})$ is said to be multi-homogeneous of degree $(\alpha_1, \ldots, \alpha_r)$ if it is homogeneous of degree α_i in each set of variables $(x_{i0}, \ldots, x_{in_i})$, and an ideal $\mathfrak{a} \in k[x_{ij} | 1 \leqslant i \leqslant r, 0 \leqslant j \leqslant n_i]$ is multi-homogeneous if it with any polynomial f contains the multi-homogeneous components of f. Formulate and prove a Nullstellensatz for the product. ✭

Hypersurfaces in $\mathbb{P}^n \times \mathbb{P}^m$

Recall that a hypersurface in $\mathbb{P}^n$ was defined as the $Z_+(F)$ for a homogeneous polynomial F.

5.33 Notice that, as in any polynomial ring, the height one prime ideals in the polynomial ring $k[x_0, \ldots, x_n, y_0, \ldots, y_m]$ are principal. If additionally a height one ideal is bihomogeneous, it will be generated by an irreducible bihomogeneous polynomial. Bearing the Nullstellensatz for $\mathbb{P}^n \times \mathbb{P}^m$ in mind, this means that any irreducible hypersurface Z (that is, a irreducible closed subset of codimension 1) is the zero locus of a bihomogeneous polynomial, which is unique up to scaling, and thus there is a *bidegree* (d, e) attached to Z.

Moreover, a general hypersurface; that is, any closed subset being the union of irreducible hypersurfaces is the zero-locus of a bihomogeneous polynomial without multiple factors, so every hypersurface has a bidegree.

5.34 Balanced hypersurfaces are precisely those whose bidegee is of the form (d, d) for some number d. A hypersurface on the Segre variety $S_{n,m}$ which is induced from $\mathbb{P}^{nm+n+m}$; that is, one which is obtained by intersecting $S_{n,m}$ with a hypersurface in $\mathbb{P}^{nm+n+m}$, is balanced. It will be of bidegree (d, d) if d is the degree of the intersecting hypersurface. Conversely, any balanced hypersurface is induced from the ambient $\mathbb{P}^{nm+n+m}$.

Example 5.35 Consider again $\mathbb{P}^1 \times \mathbb{P}^2$ coordinated so that points are shaped like $(y_0 : y_1) \times (x_0 : x_1 : x_2)$.

If α and β are two scalars, not both zero, the hypersurface $Z(\alpha y_0 - \beta y_1)$ is just the fibre over the point $(\beta : \alpha) \in \mathbb{P}^1$. A hypersurface of bidegree $(a, 0)$ is just the union of a fibres of the first projection (possibly counted with multiplicity); indeed, a polynomial of this bidegree only involves y_i's and is homogeneous of degree a. Hence it factors as a product of linear polynomials, and each vanishes along a fibre. ☆

EXAMPLE 5.36 *(Blow-up of a point in $\mathbb{P}^2$)* Let Z be the irreducible hypersurface $Z(y_0x_1 - y_1x_0)$ in $\mathbb{P}^1 \times \mathbb{P}^2$, and consider the restriction $\pi\colon Z \to \mathbb{P}^2$ of the second projection. If either $x_0 \neq 0$ or $x_1 \neq 0$, the fiber over $x = (x_0 : x_1 : x_2)$ is just the single point $(x_0 : x_1) \times (x_0 : x_1 : x_2)$, but over the point $(0 : 0 : 1)$, any point $(y_0 : y_1) \times (0 : 0 : 1)$ belongs to Z, so it holds that $\pi^{-1}(p) = \mathbb{P}^1 \times p$. For a self-explanatory reason, the map π (or sometimes also the variety Z) is called the *blow-up* of p. ☆

EXAMPLE 5.37 *(Blow-up of a complete intersection)* Consider the projective space $\mathbb{P}^1 \times \mathbb{P}^n$ with bihomogeneous coordinates so that points are shaped like $(x_0 : x_1) \times (y_0, \ldots, y_n)$, and let Z be a hypersurface of bidegree $(1, d)$; this means that Z is defined by an equation of the form

$$x_0 f(y_0, \ldots, y_n) - x_1 g(y_0, \ldots, y_n) = 0,$$

where f, g are homogeneous forms of degree d in the y_i. The fibres of the second projection $p\colon Z \to \mathbb{P}^n$ are of two types:

i) $p^{-1}(a) = \mathbb{P}^1 \times \{a\}$ if $a \in Z(f, g)$;

ii) $p^{-1}(a)$ is a point, for $a \notin Z(f, g)$.

Indeed, if f and g vanish at he point a equation (5.37) is fulfilled for all $(x_0 : x_1)$, and if one of them does not, it can be inverted, and one of the x_i's is determined by the other. ☆

EXAMPLE 5.38 *(Double covers of $\mathbb{P}^2$)* Again, we equip $\mathbb{P}^1 \times \mathbb{P}^2$ with coordinates so that points are on the form $(x_1 : x_2) \times (y_0 : y_1 : y_2)$.

Consider the variety $X = Z_+(F) \subset \mathbb{P}^1 \times \mathbb{P}^2$ defined by the bihomogeneous polynomial

$$F = x_0^2 y_0 + x_0 x_1 y_1 + x_1^2 y_2$$

of bidegree $(2, 1)$. Note that X admits two morphisms, $p\colon X \to \mathbb{P}^1$ and $q\colon X \to \mathbb{P}^2$ coming from the two projections. To analyze the geometry of X, we again consider the fibres of these two morphisms.

THE FIBERS OF p: For points $(a : b) \in \mathbb{P}^1$, we find

$$p^{-1}((a : b)) = Z_+(a^2 y_0 + aby_1 + b^2 y_2),$$

which may be identified with the line $a^2y_0 + aby_1 + b^2y_2 = 0$ in the projective plane $\mathbb{P}^2 = (a : b) \times \mathbb{P}^2$. All fibers of p are therefore isomorphic to $\mathbb{P}^1$.

THE FIBERS OF q: Over each point $(a : b : c) \in P^2$ the fibre of q will be

$$q^{-1}((a : b : c)) = Z_+(ax_0^2 + bx_0x_1 + cx_1^2) \subset \mathbb{P}^1 \times (a : b : c).$$

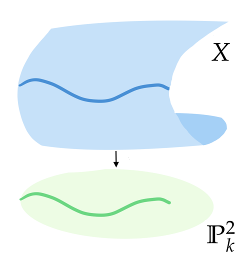


Figure 5.1: A double cover of $\mathbb{P}^2$

The quadratic equation $ax_0^2 + bx_0x_1 + cx_1^2 = 0$ has two solutions whenever $b^2 - 4ac \neq 0$, so in this case the fibre $q^{-1}((a : b : c))$ consists of two points. In the case $b^2 = 4ac$, there is a repeated root, and the fibre $q^{-1}((a : b : c))$ is a singleton. We conclude that the projection $q\colon X \to \mathbb{P}^2$ is what one calls a 'double cover'; it is two-to-one over the open set $X - q^{-1}(C)$ in $\mathbb{P}^2$, where C is the conic $C = Z_+(y_1^2 - 4y_0y_2)$, but over C is just is one-to-one. We say it 'ramifies' over C.

☆

Exercises

5.10 Let $\pi\colon \mathbb{P}^1 \times \mathbb{P}^1 \dashrightarrow \mathbb{P}^2$ be the rational map defiened by the assignment $(y_0 : y_1) \times (x_0 : x_1) \mapsto (x_0y_1 : x_1y_0 : x_0y_1)$.

a) Show that π is defined everywhere except at $p = (1 : 0) \times (1 : 0)$.

b) Show that the two fibres A_1 and A_2 through p are collapsed to the points $p_1 = (0 : 1 : 0)$ and $p_2 = (1 : 0 : 0)$ and that π restricts to an isomorphism of $\mathbb{P}^1 \times \mathbb{P}^1 \backslash A_0 \cup A_1$ with $\mathbb{P}^2 \backslash L$ where L is the line through p_1 and p_2.

c) Show that a linear section of $\mathbb{P}^1 \times \mathbb{P}^1$ maps to a curve whose closure in $\mathbb{P}^2$ either is a conic through the points p_1 and p_2, or a line through one of them.

d) Show that an irreducible hypersurface of bidegree $(2, 2)$ in $\mathbb{P}^1 \times \mathbb{P}^1$ is mapped to a curve whose closure is an irreducible quartic curve with a double point in both p_1 and p_2

5.11 Show that the coordinate ring $A(C(S_{n,m}))$ of the cone over the Segre variety $S_{n,m}$ is not a UFD. Give explicit examples of height one primes that are not principal.

5.12 Show that under the Segre map the fibres of the two projections from $\mathbb{P}^1 \times \mathbb{P}^1$ onto $\mathbb{P}^1$ embed as lines in $\mathbb{P}^3$. Show that if Z is an effective divisor in $\mathbb{P}^1 \times \mathbb{P}^1$ of bidegree $(0, n)$ or $(n, 0)$, then Z is a union of lines.

5.13 Let $\Phi\colon \mathbb{P}^1 \to \mathbb{P}^1$ be given as $(f_0(x_0, x_1) : f_1(x_0, x_1))$ where the f_i's are two homogeneous polynomials of the same degree without common zeros. Show that graph $\Gamma_\Phi \subseteq \mathbb{P}^1 \times \mathbb{P}^1$ is the zero locus of the bihomogeneous polynomial $y_1 f_0(x_0, x_1) - y_0 f_1(x_0, x_2)$.

5.14 Show that an irreducible divisor of degree $(1, 2)$ in $\mathbb{P}^1 \times \mathbb{P}^1$ under the Segre embedding either is embedded as a twisted cubic or as a plane cubic having a singular point.

5.15 Show that the rational normal curve C_d in $\mathbb{P}^d$ is the intersection of the Segre variety $S_{1,d-1}$ in $\mathbb{P}^{2d-1}$ with an approriate linear subspace of dimension d.

5.16 Consider the space curve C with a parametrization $\Phi\colon \mathbb{P}^1 \to \mathbb{P}^3$ given as $\Phi(x_0 : x_1) = (x_0^4 : x_0^3x_1 : x_0x_1^3 : x_1^4)$. Show that C lies on the quadratic surface $S_{1,1}$

and is of bidegree $(1,4)$. Determine the bihomogeneous polynomial F defining it.

★

5.5 The Veronese embeddings

This is a class morphisms of $\mathbb{P}^n$ into a larger projective space $\mathbb{P}^N$ given by all the monomials of a given degree d in $n+1$ variables. There are $\binom{n+d}{d}$ such monomials, so the number N is given as $N = \binom{n+d}{d} - 1$. These Veronese embeddings[7] deserve their name; they are *closed embeddings*. They depend on two natural numbers n and d, and the corresponding embedding will be denoted by $\Phi_{n,d}$ or most often just by Φ with n and d tacitly understood.

[7] They also go under the monstrosity of a name, the *d-uple embeddings*.

Giuseppe Veronese (1854–1917) Italian mathematician

The Veronese surface
Veronese-flaten

EXAMPLE 5.39 We already met some morphisms of this kind. The parameterizations of the rational normal curves are of this shape with $n = 1$. They are morphisms Φ from $\mathbb{P}^1$ into $\mathbb{P}^d$ which are expressed as

$$\Phi(x_0 : x_1) = (x_0^d : x_0^{d-1}x_1 : \ldots : x_0x_1^{d-1} : x_1^d)$$

in the homogeneous coordinates $(x_0 : x_1)$ of $\mathbb{P}^1$. Conics in $\mathbb{P}^2$ and the twisted cubic for instance, are prominent members of this clan. ☆

EXAMPLE 5.40 The *Veronese surface* is another example of the sort with $n = 2$ and $d = 2$. In that case the embedding of $\mathbb{P}^2$ into $\mathbb{P}^5$ is given as

$$\Phi(x_0 : x_1 : x_2) = (x_0^2 : x_1^2 : x_2^2 : x_0x_1 : x_0x_2 : x_1x_2)$$

in terms of the homogeneous coordinates $(x_0 : x_1 : x_2)$ on $\mathbb{P}^2$. Notice that the maps in both these examples are morphisms according to the little lemma (Lemma 4.48 on page 89). ☆

The definition

5.41 To fix the notation let $\mathcal{I}$ denote the set of sequence $I = (\alpha_0, \ldots, \alpha_n)$ of non-negative integers such that $\sum_i \alpha_i = d$; there are $N+1$ of them. The sequences from $\mathcal{I}$ will serve as indices for the monomials of degree d; that is, when I runs through $\mathcal{I}$, the polynomials $M_I = x_0^{\alpha_0} \ldots x_n^{\alpha_n}$ will run through the monomials of degree d in the x_i's. We let (m_I) for $I \in \mathcal{I}$, in some order, be homogeneous coordinates on the projective space $\mathbb{P}^N$.

One may think about the affine space $\mathbb{A}^N$ as the vector space consisting of homogeneous forms of degree d in the $n+1$ variables $x_0, \ldots, x_n$, and consequently $\mathbb{P}^N$ will be the space of classes $[F]$ of such forms modulo scaling. Using the coordinates m_I a form F thus is expressed as $F = \sum_{I \in \mathcal{I}} m_I(F) M_I$.

5.42 The *Veronese embedding* with parameters n and d is then the mapping $\Phi_{n,d}\colon \mathbb{P}^n \to \mathbb{P}^N$ that sends the point $[x] = (x_0 : \ldots : x_n)$ to the point in $\mathbb{P}^N$ whose homogeneous coordinates are given as $m_I(\Phi([x])) = M_I(x)$; that is,

Veronese embeddings
Veronese-embeddingene

$$m_I(\Phi_{n,d}([x])) = x_0^{\alpha_0} \ldots x_n^{\alpha_n},$$

when $I = (\alpha_0, \ldots, \alpha_n)$. The monomials M_I are homogeneous of the same degree d, and they do not vanish simultaneously anywhere. The mapping $\Phi_{n,d}$ is therefore a morphism as follows from Lemma 4.48 on page 89. But as alluded to above, much more is true:

PROPOSITION 5.43 *The Veronese map $\Phi_{n,d}$ is a closed embedding of $\mathbb{P}^n$ into $\mathbb{P}^N$ where $N = \binom{n+d}{d} - 1$.*

The image of $\Phi_{n,d}$ is called a *Veronese variety* and will be denoted by $V_{n,d}$. More generally every subvariety of $\mathbb{P}^N$ projectively equivalent with $V_{n,d}$ is also called a *Veronese variety*.

Veronese varieties
Veronese-varieteter

PROOF: To simplify the notation we let Φ stand for $\Phi_{n,d}$.

We shall use a particular subset of the standard open subsets $D_+(m_I)$ that considerably simplifies the argument. They cover the image of Φ, but Lemma 5.2 requires they cover the closure, so we shall assume the image is closed. This is justified by the general result that images of projective varieties always are closed (Theorem 10.18 on page 197). It is possible to use the full covering of standard opens, like we did in the case of the rational normal curves, but it leads into combinatorial intricacies which do not make us much wiser; and one might even consider it false teaching not to follow the broad road.

There are three salient points in the proof.

Firstly, the basic open subset $D_i = D_+(x_i)$ of $\mathbb{P}^n$ where the i-th coordinate x_i does not vanish, maps into one of the distinguished open subsets of $\mathbb{P}^N$, namely the one corresponding to the pure power monomial x_i^d. To make the notation simpler we let m_i denote the corresponding homogeneous coordinate[8] on $\mathbb{P}^N$; so that Φ maps D_i into $D_+(m_i)$. The first of these distinguished open subsets D_i is isomorphic to $\mathbb{A}^n$, with the fractions $x_j x_i^{-1}$ for $j \neq i$ as coordinates, and the second to $\mathbb{A}^N$ with $m_I m_i^{-1}$ as coordinates.

[8] That is, m_i equals the coordinate m_I with I being the sequence $I = (0, 0, \ldots, d, \ldots, 0)$ that has a d in slot i and zeros everywhere else.

Secondly, although the $n+1$ basic open subsets $D_+(m_i)$ do not cover the entire $\mathbb{P}^N$, they cover the image $\Phi(\mathbb{P}^n)$. Hence it suffices (Lemma 5.2 on page 96) to see that for each index i the restriction $\Phi|_{D_i}$ has a closed image and is an isomorphism onto its image.

The third salient point is that the restrictions $\Phi|_{D_i}\colon \mathbb{A}^n \to \mathbb{A}^N$ have sections that are linear projections. Once this is establishes, we are through in view of Lemma 5.4 above. To exhibit a section, we introduce the n monomials $M_{ij} = x_j x_i^{d-1}$ where $j \neq i$ and denote the corresponding homogenous coordinates by m_{ij}. Then $m_{ij}(\Phi(x)) m_i(\Phi(x))^{-1} = x_j x_i^{-1}$, from which ensues that the projection onto the affine space $\mathbb{A}^n$ corresponding to the coordinates $m_{ij} m_i^{-1}$ with $j \neq i$ is a section of the map $\Phi|_{D_i}$, and we are through. ❏

Two corollaries

5.44 Even though the Veronese varieties are specific varieties interesting enough on their own, they are also of a general theoretical interest. As an illustration we offer two important results, both easy corollaries of the embedding theorem.

Corollary 5.45 *Let $f(x_0, \ldots, x_n)$ be a non-zero homogeneous polynomial. Then the distinguished open subset $D_+(f)$ of $\mathbb{P}^n$ is affine.*

We already know this for linear polynomials, the news is it is true for polynomials of any degree.

Proof: Let d be the degree of f. The crucial point is that the locus $Z_+(f)$ in $\mathbb{P}^n$ becomes a *linear section* of the Veronese variety $V_{n,d}$. It will be equal to $Z_+(L) \cap V_{n,d}$ where L is the linear expression in the m_I's sharing the coefficients of the expression of f in terms of the M_I's; that is, if $f = \sum_{I \in \mathcal{I}} a_I M_I$ one has $L = \sum_{I \in \mathcal{I}} a_I m_I$. Indeed, it then holds true that $f(x_0, \ldots, x_n) = L(\Phi_{n,d}([x]))$, at least up to non-zero scaling.

The distinguished open set $D_+(f)$ is the complement of $Z_+(f)$, and hence equals the intersection of $V_{n,d}$ with the complement of $Z_+(L)$. The latter is the distinguished open subset $D_+(L)$ of $\mathbb{P}^N$ and is isomorphic to $\mathbb{A}^N$. Consequently $\Phi(D_+(f))$, which is isomorphic to $D_+(f)$, is closed in $\mathbb{A}^N$, and therefore it is affine. ❏

5.46 The next corollary is a forerunner of a series of results about intersections of closed subvarieties of projective space. They rely on the notion of dimension and will be treated when that is in place, but the present simple version catches in many ways the spirit of the general case — if a set of homogeneous equations has infinitely many simultaneous solutions, we may add one equation and still have a common solution.

Corollary 5.47 *Assume that $X \subseteq \mathbb{P}^n$ is a subvariety which is not a point, and that $f(x_0, \ldots, x_n)$ is a homogeneous polynomial. Then $Z_+(f) \cap X$ is not empty.*

Proof: Assume that $X \cap Z_+(f) = \emptyset$. Then $X \subseteq D_+(f)$; but $D_+(f)$ is affine and therefore X being closed in $D_+(f)$ is affine. So X is both affine and projective. Corollary 4.46 on page 88 applies, and X is a point. ❏

Equations of the Veronese varieties

There is a very natural collection of quadratic equations, which we shall shortly present, that cut out the Veronese varieties. However, what is not at all obvious is that these quadrics generate the prime ideal $I(V_{n,d})$. There seems be no good modern reference in introductory texts including all details. The result is mentioned in several texts, but usually the proof suffers the fate of being swept under the rug. Although it requires a certain amount of notational complexity and some combinatoric struggle, we shall offer a detailed account. It illustrates one of the subtleties one meets in projective geometry – the difference between knowing enough equation to cut out a variety V and knowing the ideal $I(V)$ of all polynomials vanishing along V.

5.48 Recall that the homogeneous coordinates m_I of $\mathbb{P}^N$ are indexed by sequences $(\alpha_0, \ldots, \alpha_n)$ of non-negative integers whose sum equals d. To the coordinate m_I corresponds the monomial $M_I = x_0^{\alpha_0} \ldots x_n^{\alpha_n}$ in the x_i's, which equals the restriction of m_I to the Veronese variety $V_{n,m}$; that is, one has $m_I|_{V_{n,m}} = M_I$.

Obviously the relation $M_I M_J = M_K M_L$ holds true whenever the four indexing sequences I, J, K and L satisfy the relation $I + J = K + L$; indeed, the exponents of each x_i on the two sides of the equality then add up to the same number. This means that the quadratic polynomials $m_I m_J - m_K m_L$ vanish along $V_{n,m}$, and these quadratic polynomials are the ones alluded to at the top of this section. But not only do they belong to the ideal $I(V_{n,m})$, they also generate $I(V_{n,m})$, and the present section is entirely devoted to a proof this.

Proposition 5.49 *Let $\mathfrak{a}$ be the ideal in the polynomial ring $k[m_I | I \in \mathcal{I}]$ generated by the quadric polynomials $m_I m_J - m_K m_L$ for $I + J = K + L$. Then $\mathfrak{a}$ is a prime ideal, and it holds true that $\mathfrak{a}$ is the homogeneous ideal of the Veronese variety $V_{n,d}$; that is, $\mathfrak{a} = I(V_{n,d})$.*

The proof follows the same broad lines we already saw when studying the Segre varieties, but the details are certainly more involved. As for Segre varities, the proof consists of showing that every monomial $m_{I_1} \ldots m_{I_r}$ in the m_I's can, modulo the ideal $\mathfrak{a}$, be brought on a normal form such that different normal forms correspond to different monomials in the x_i's. From this ensues that the map $k[m_I | I \in \mathcal{I}]/\mathfrak{a} \to k[x_0, \ldots, x_n]$ is injective so that $\mathfrak{a}$ is prime; indeed, the classes of the normal forms are mapped injectively to linearly independent elements in $k[x_0, \ldots, x_n]$, and hence they form a linear basis for $k[m_I | I \in \mathcal{I}]/\mathfrak{a}$.

5.50 We begin with introducing some *ad hoc* terminology. Say that a sequence S from $\mathcal{I}$ is *standard of type s, t* if it is shaped like

$$S = (\gamma_0, 0, \ldots, 0, \gamma_s, \ldots, \gamma_t, 0, \ldots, 0);$$

that is, the non-zero components apart from the first are found in the region where indices lie between s and t, a condition that obviously is restrictive if and only if $0 < s$ or $t < n$, and of course is meaningful only when $s \leqslant t$. More over we also request that $\gamma_i < d$ for all i. The corresponding monomial m_S is as well said to be *standard of type* s, t.

Recall that we use the notation m_i for the coordinate indexed by the sequence $(0, \ldots, 0, d, 0 \ldots, 0)$ whith the d located in slot i. It corresponds to the monomial x_i^d.

5.51 The fulcrum of the proof is the following technical lemma which establishes the normal form. It states that any monomial in the m_I's is congruent modulo $\mathfrak{a}$ to a product of powers of the m_i's and a certain number of monomials of standard type s_i, s_{i+1} with the sequence $[s_i, s_{i+1})$ of integral intervals forming a partition of $\{1, \ldots, n\}$.

LEMMA 5.52 *Let $m = m_{I_1} \cdot \ldots \cdot m_{I_c}$ be a monomial in the m_I's of degree c. There exists a strictly increasing sequence $s_0, \ldots, s_a$ of integers between 1 and n with $s_0 = 1$ and $s_a = n$ such that modulo $\mathfrak{a}$ the equality*

$$m = m_{s_{a-1}, s_a} \cdot \ldots \cdot m_{s_0, s_1} \cdot p$$

holds, where each $m_{s_i, s_{i+1}}$ is standard of type s_i, s_{i+1}, and where p is a product $p = m_1^{b_1} \cdot \ldots \cdot m_n^{b_n}$ with $\sum_i b_i = c - a$.

PROOF: The proof goes by induction on the number of variables n and the degee c of the monomial. By induction on c, we may assume that any product $m_{I_1} \cdot \ldots \cdot m_{I_{c-1}}$ is congruent to a monomial of the requested form, and thus may be replace by

$$m_{s_{a-1}, s_a} \ldots m_{s_1, 1} \cdot p,$$

and we are to analyse what happens when a new factor m_I is introduced; that is, we must bring $m = m_I \cdot m_{s_{a-1}, s_a} \ldots m_{s_1, 1} \cdot p$ on the required form.

Assume that $I = (\alpha_0, \ldots, \alpha_n)$ and let $S = (\gamma_0, 0 \ldots, 0, \gamma_s, \ldots, \gamma_n)$ be the standard sequence such that $m_S = m_{s_{a-1}, s_a}$ where we have let $s_{a-1} = s$ to ease the notational stress. We also denote by δ_i the i-th component of S (so it equals zero when $0 < i < s$ and γ_i else).

We want to play an exchange game with the product $m_I m_{s_{a-1}, s_a} = m_I m_S$ and replace it by $m_K m_L$ for juicily chosen K and L. There will be two cases.

We first assume that $\alpha_i + \delta_i < d$ for all i. Let t be the smallest integer so that $\sum_{0 < i \leqslant t} (\alpha_i + \delta_i) > d$. Certainly there is one, since $\sum_i (\alpha_i + \delta_i) = 2d$ and $\alpha_0 + \delta_0 < d$.

Consider the two sequences

$$K = (\textstyle\sum_{0 \leqslant j < t} (\alpha_i + \delta_i) - d, 0, \ldots, 0, \alpha_t + \delta_t, \ldots, \alpha_n + \delta_n)$$

and

$$L = (d - \textstyle\sum_{0 < j < t} (\alpha_j + \delta_j), \alpha_1 + \delta_1, \ldots, \alpha_{t-1} + \delta_{t-1}, 0, \ldots, 0)$$

Since $a_i + \delta_i < d$ for all i and by the choice of t, it is a matter of trivial computations to verify that both sequences K and L belong to $\mathcal{I}$ and that $I + S = K + L$.

We also contend that $s \leqslant t$; indeed, this ensues from $\delta_i = 0$ for $0 < i < s$ and $\sum_{0<i\leqslant s} \alpha_i \leqslant d$.

It follows that $m_I m_S$ has the same class as $m_K m_L$ modulo $\mathfrak{a}$, and our mononial m takes the form

$$m = m_K m_L \cdot \ldots \cdot m_{s_1,s_0}.$$

The crucial observation is now that the monomial m_L and all the monomials $m_{s_i,s_{i+1}}$ correspond to sequences in $\mathcal{I}$ whose indices vanish for $i \geqslant s$. Since $s < n$, induction (on n) applies, and the product $m_L m_S \ldots m_{s_1,s_0}$ can be brought on normal form involving factors $m_{s_i,s_{i+1}}$ with $s_{i+1} < t$. The second crucial observation is that m_K is standard of type t, n and thereby we are through.

The second case arrives when one of the sums $\alpha_i + \delta_i$ exceeds d, say this happens for $i = \nu$. Then

$$J = (\alpha_0 + \delta_0, \ldots, \alpha_\nu + \delta_\nu - d, \ldots, \alpha_n + \delta_n)$$

will be a new sequence in $\mathcal{I}$, and it holds true $m_I m_S = m_J m_\nu$. Absorbing the factor m_ν in the product p brings us to a case with smaller c for which induction applies. ❏

PROOF OF PROPOSITION 5.49: To finish the proof we need merely establish the counterpart of the normal form for monomials in the x_i's. That is, we must show – with the stress on unambiguous – that any monomial $M = x_0^{c_0}, \ldots, x_n^{c_n}$ of degree a multiple of d is expressed in an umambiguous way as a product

$$M = M_{S_1} \cdot \ldots \cdot M_{S_r} \cdot x_1^{b_1 d} \cdot \ldots \cdot x_n^{b_n d}$$

where the S_i's are standard sequences forming a partition of $\{1, \ldots, n\}$ and the b_i's are non-negative integers. By long division one finds non-negative integers b_i and γ_i such that $\gamma_i < d$ and $c_i = b_i d + \gamma_i$. They are of course uniquely defind by the monomial M.

To proceed, let t be the largest integer so that $\sum_{0<i\leqslant t} \gamma_i < d$; it exists since $\gamma_1 < d$. This choice makes the sequence $T = (d - \sum_{0<i\leqslant t} \gamma_i, \gamma_1, \ldots, \gamma_t, 0, \ldots, 0)$ a standard sequence of type $1, t$ and clearly the corresponding monomial M_T is a factor of M. Induction then applies to MM_T^{-1}, and we are through. ❏

5.6 Conics and the Veronese surface

The Veronese surface deserves special attention. Among the many interesting properties it enjoys is that it is intimately related to plane conics; or more precisely to the degenerate ones; that is, it parameterizes the double lines. A feature of quadrics of any number of variables is that the ones of rank one—that is, those being squares of linear forms—are parameterized by the Veronese varieties $V_{n,2} = \Phi_{n,2}(P^n)$. However, we'll confine ourselves to the plane case leaving the general case as an exercise for the zealous students. Quadratic forms have a tendency to behave weirdly when the characteristic of the ground field is two, so all we do in this section holds only when the characteristic of k is different from two.

5.53 The plane conics are the loci in $\mathbb{P}^2$ where a quadratic form in in three variables vanishes. These forms constitute a vector space of dimension six, which we babtized S_3 in Example 2.63 on page 43. Proportional forms having the same zero locus, the conics are in a naturally one-to-one correspondence with the projective space $\mathbb{P}^5 = \mathbb{P}(S_3)$ of lines in S_3.

Every quadratic form does not define a conic in the strict sense of the word, there are degenerate ones. The union of two lines for instance, arise as the zero locus of a product of two distinct linear forms, and when the two lines become equal, we get the *double lines*; they correspond to squares of a linear forms. This means that inside $\mathbb{P}(S_3)$ we find two particular subsets; one is in a one-to-one correspondence with the set of pairs of distinct lines and the other with the set of double lines. Since a linear form is determined up to scale by its square, the latter is in a one-to-one correspondence with the set of lines. Now, the set of lines constitute a $\mathbb{P}^2$, so that the double lines form a $\mathbb{P}^2$ lying canonically within $\mathbb{P}(S_3)$; and of course, that $\mathbb{P}^2$ is a Veronese surface. And the other one? Well, it is the projection of a Segre variety $S_{2,2}$!

5.54 To fix the notation, we let V be the vector space of linear forms in the variables T_0,T_1 and T_2. It has coordinates t_0, t_1 and t_2 so that any linear form $L \in V$ is expressed as

$$L = t_0T_0 + t_1T_1 + t_2T_2.$$

The space S_3 of quadratic forms in the variables T_i has a basis $\{T_iT_j\}$ with $i \leqslant j$ and $0 \leqslant i, j \leqslant 2$. When revealing that the square locus equals the Veronese surface, it will be convenient to use another, but very similar, basis; the two differ merely by the factor 2 of the cross terms. The new basis $\{\mathcal{B}\}$ is

$$\mathcal{B} = \{T_0^2, T_2^2, T_2^2, 2T_0T_1, 2T_0T_2, 2T_1T_2\}. \tag{5.7}$$

The corresponding homogeneous coordinates in $\mathbb{P}(S_3)$ will be denoted t_{ij} with the usual constraint on the indices so that any form P from S_3 are written as $P = \sum_i t_{ii}T_i^2 + \sum_{i<j} 2t_{ij}T_iT_j$, summing over apt indices i and j. This is in concordance with the notation

$$P(T) = TA(t)T^t$$

from Example 2.63 on page 43; the coefficient matrix $A(t) = (t_{ij})$ is a *symmetric* 3×3-matrix and T^t stands for the transpose of the row-vector $T = (T_0, T_1, T_2)$.

5.55 When a linear form L is multiplied by a scalar α, the square L^2 gets multiplied by α^2. This insures that the assignment $[L] \mapsto [L^2]$ yields a well defined map between the projective spaces $\mathbb{P}(V)$ and $\mathbb{P}(S_3)$; that is, a map $\psi\colon \mathbb{P}(V) \to \mathbb{P}(S_3)$. And the point is that ψ is projectively equivalent to the Veronese map; *i.e.* with an appropriate choice of bases it equals the map defined in Paragraph 5.39 on page 113.

This is achieve simply by the multinomial formula; it entails the identity

$$L^2 = (t_0T_0 + t_1T_1 + t_2T_2)^2 = \sum_{i,j} t_it_jT_iT_j = \sum_i t_i^2T_i^2 + \sum_{i<j} t_it_j(2T_iT_j) \qquad (5.8)$$

where the sums extend over indices $0 \leqslant i, j \leqslant 2$ complying to indicated constraints. Expressed in the basis $\mathcal{B}$, the map ψ may therefore be described as

$$\psi(t_0 : t_1 : t_2) = (t_0^2 : t_1^2 : t_2^2 : t_0t_1 : t_0t_2 : t_1t_2),$$

and we recognise this as the Veronese map from 5.39. We have thus established:

Proposition 5.56 *Assume that the characteristic of the ground field k is different from two. Then the subvariety of the space $\mathbb{P}(S_3)$ of plane conics consisting of the double lines, is a Veronese surface $V_{2,2}$.*

5.57 The assumption that the characteristic be different from two is essential. In characteristic two all the cross terms in (5.8) vanish, and the image of the squaring map becomes the *linear* subspace generated by the squares T_0^2, T_1^2, T_2^2. The squaring map, though being bijective and having a closed image, is not an embedding; it does not induce an isomorphism between the sheaf of regular functions.

5.58 As above any quadratic form P has a coefficient matrix and may be expresses as the matrix product

$$P(x) = TA(t)T^t$$

with $A(t) = (t_{ij})$ is a *symmetric* 3×3-matrix. In Example 2.5 we saw that the form P is a square of a linear form L precisely when the matrix $A(t)$ is of rank one; in other words, precisely when all the 2×2-minors of the matrix $A(t)$ vanish, and in fact these minors generated the prime ideal of the locus of squares, *i.e.* the homogeneous prime ideal of the Veronese surface:

$$A(t) = \begin{pmatrix} t_{00} & t_{01} & t_{02} \\ t_{10} & t_{11} & t_{12} \\ t_{20} & t_{21} & t_{22} \end{pmatrix}$$

Proposition 5.59 *The homogenoeus ideal of $I(V_{2,2})$ in $k[t_{ij}]$ is generated by the minors of the symmetric matrix $A(t)$.*

Proof: The general result in Proposition 5.49 tells us that the homogeneous ideal $I(V_{2,2})$ is generated by quadrics, and a simple dimension count yields that there are six independent ones; indeed, restricting quadric polynomials in the t_{ij}'s to the Veronese surface yields the exact sequence

$$0 \longrightarrow I_2 \longrightarrow k[t_{ij}]_2 \xrightarrow{\rho} k[t_i]_4 \longrightarrow 0,$$

where as usual subscripts denote homogeneous parts of a given degree, and where the map ρ substitutes t_it_j for t_{ij} (hence the degree doubles). The sequence yields the equality[9] $\dim_k I_2 = \binom{5+2}{2} - \binom{2+4}{2} = 21 - 15 = 6$. Now, being a symmetric matrix, $A(t)$ has precisely six minors (one for each choice of a row

[9] Recall that the space of forms of degree d in $n+1$ variables is of dimension $\binom{n+d}{d}$.

and a column, but interchanging the row and the column gives the same minor), and it is not difficult to verify that they are linearly independent (you are asked to do that in Exercise 5.17 below). ❑

The real points of the Veronese surface be can be realized as a projection into $\mathbb{R}^3$, at least if one allows the surface to have self intersections. The surface depicted in the margin is parametrized by three out of the six quadratic terms in the parametrization (2.2) above, and is the image of the unit sphere in $\mathbb{R}^3$ under the map $(x, y, z) \mapsto (xy, xz, yz)$. This specific real surface is often called the *Steiner surface* after the Swiss mathematician Jacob Steiner who was the first to describe it. It also goes under the name of the *Roman surface* since Steiner visited Rome when he discovered it.

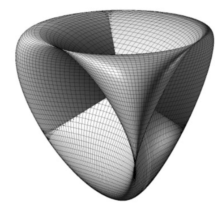


Figure 5.2: The Roman surface – a projection of a real Veronese surface

Exercises

5.17 Verify that the minors of $A(t)$ are linearly independent.

5.18 *(The degenerate conics.)* This exercise is about the locus D of degenerate conics in the space $\mathbb{P}(S_3)$ of plane conics , *i.e.* those being either the union of two lines or a double line. (The characteristic of k is different from two.)

a) Prove that D is a cubic hypersurface. Hint: The equation is $\det A(t) = 0$.
b) Show that the map $\eta \colon \mathbb{P}(V) \times \mathbb{P}(V) \to \mathbb{P}(S_3)$ that sends a pair (L, M) of linear form to the product LM, is a well defined morphism that is a two-to-one map onto D.
c) Prove that the restriction of η to the diagonal in $\mathbb{P}(V) \times \mathbb{P}(V)$ is the Veronese map $\Phi_{2,2}$.
d) Prove that D is a projection of the Segre variety $S_{2,2}$ in $\mathbb{P}^8$.
e) Prove that D is the secant variety of $V_{2,2}$; that is, if a line meets the Veronese surface $V_{2,2}$ in two points, it is entirely contained in D.

5.19 *(A generalization.)* Most of the results in the present section can be generalized to forms in $n+1$ variables and the Veronese varieties $V_{n,2}$. Let V the space of linear forms in $n+1$ variables. Interpret $\mathbb{P}^N$, with $N = \binom{n+2}{2}$, as the space of quadratic forms in $n+1$ variable and consider the map $\psi \colon \mathbb{P}(V) \to \mathbb{P}^{N-1}$ that sends $[L]$ to $[L^2]$.

a) Prove that ψ is a well defined morphism which is projectively equivalent to a Veronese map. What restrictions on the characteristic of k is there?
b) Prove that the image of ψ is equal to the locus of rank-one quadrics.

5.20 *(Another generalization.)* Some of what we have done not only holds for squares of linear forms, but for higher powers as well. Let d and n be natural numbers and put $N_d = \binom{n+d}{d}$, and denote by V the space of linear forms of degree d in $n+1$ variables. Prove the map $\mathbb{P}(V) \to \mathbb{P}^{N_d - 1}$ that sends $[L]$ to $[L^d]$ is a well defined morphism which is projectively equivalent to a Veronese map under certain restrictions on the characteristic of the ground field k.

★

5.7 Geometry of the spaces of polynomials

Let V_d be the vector space of polynomials in the variable x of degree at most d. It has as basis the powers $1, x, \ldots, x^d$. We let $\mathbb{P}^d$ denote the projective space associated to V_d. Thus, up to a non-zero scalar multiple, a polynomial

$$p(x) = a_0 + a_1 x + \ldots + a_d x^d$$

may be identified with the point $(a_0 : a_1 : \cdots : a_d) \in \mathbb{P}^d$.

Inside $\mathbb{P}^d$ there are interesting closed subsets corresponding to polynomials with repeated roots. More precisely, for any partition[10] λ of d, we have the subset

$$V_\lambda \subseteq \mathbb{P}^d$$

defined as the closure of the set of polynomials with r roots $x_1, \ldots, x_r$ with multiplicities $\lambda_1, \ldots, \lambda_r$. Here and now we announce the following result, but at the present stage of the theory we lack tools to give a complete proof, so we only observe the high geometric points.

[10] That is, λ is a tuple of positive integers $(\lambda_1, \ldots, \lambda_r)$ so that $\sum \lambda_i = d$.

Proposition 5.60 *For each partition $\lambda = (\lambda_1, \ldots, \lambda_r)$ of d, the set V_λ is a projective variety of dimension*

$$\dim V_\lambda = r.$$

Proof: We first need to show that V_λ is irreducible. If we view $\mathbb{P}^1$ as a space of non-zero polynomials of degree at most one, we have a "multiplication map"

$$\phi : \mathbb{P}^1 \times \cdots \times \mathbb{P}^1 \to \mathbb{P}^d \tag{5.9}$$

that sends $(u_1 : v_1) \times \cdots \times (u_r : v_r)$ to the product

$$(u_1 x + v_1)^{\lambda_1} \cdot \cdots \cdot (u_r x + v_r)^{\lambda_r}.$$

By expanding the product we see that ϕ is defined by homogeneous polynomials (or if you want, scaling each $u_i x + v_1$ by s_i, results in the product being scaled s byr $s_1^{\lambda_1} \cdot \ldots \cdot s_r^{\lambda_r}$). Moreover, the product is non-zero since each factor is, so ϕ is a morphism of projective varieties.

We shall ultimately prove the important result that images of projective varieties under morphisms are closed (Theorem 10.21 on page 199). The image of ϕ therefore is closed, and so by definition V_λ is precisely the image of ϕ. Since the product on the left-hand side of (5.9) is irreducible, so is V_λ.

The product on the left-hand side has dimension r, which shows the inequality $\dim V_\lambda \leqslant r$. In fact, over an algebraically closed field, any polynomial can be factored uniquely into linear forms (up to order), and this shows that ϕ has finite fibres, so again by a future theorem (Theorem 10.8 on page 193) we may conclude that $\dim V_\lambda = r$. ❑

Note that $V_{(1,\cdots,1)}$ equals the entire ambient space $\mathbb{P}^d$, and in that case the linear factors can freely be permuted so that the cardinality of the fibre is d! Among the remaining V_λ's, there is one which is maximal, namely $V_{(2,1,\ldots,1)}$ which corresponds to polynomial with a repeated root. It is of dimension $d-1$, hence defined by one polynomial. From the theory of polynomials of one variable, we already know that this locus is defined by the *discriminant* Δ, which is a polynomial in the coefficients $a_0, \ldots, a_d$. A consequence of what we just proved is that Δ is irreducible.

The discriminant
diskriminanten

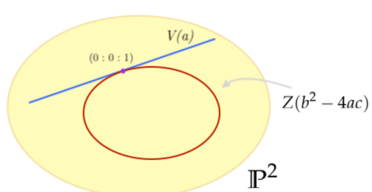
We also have $V_(d)$, which corresponds to the smallest stratum where all the roots coincide. This is a curve in $\mathbb{P}^d$, and up to a linear change of coordinates $V_(d)$ is in fact exactly the rational normal curve ov degree d!

Degree 2

Quadratic polynomials $ax^2 + bx + c$ up to scaling are parameterized by a $\mathbb{P}^2$ with homogeneous coordinates $(a : b : c)$. The space of quadratics with a repeated root is given by

$$V_2 = Z(b^2 - 4ac) \subset \mathbb{P}^2,$$

which is a plane conic curve.



Degree 3

For cubic polynomials $ax^3 + bx^2 + cx + d$, which are parameterized by points is projective space $\mathbb{P}^3$, we have two interesting loci $V_{(2,1)}$ and $V_{(3)}$. The variety $V_{(2,1)}$ equals the hypersurface $Z_+(\Delta)$ where Δ is the discriminant

$$\Delta = b^2c^2 - 4\,a\,c^3 - 4\,b^3 d + 18\,a\,b\,c\,d - 27\,a^2 d^2,$$

where as the variety $V_{(3)}$ is given by the ideal

$$\mathfrak{a} = (c^2 - 3\,b\,d, b\,c - 9\,a\,d, b^2 - 3\,a\,c).$$

Up to a linear change of coordinates this is the twisted cubic curve. This is pictured in Figure 5.3.

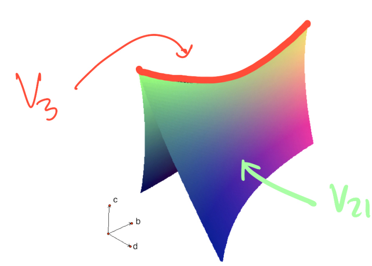


Figure 5.3: The discriminant hypersurface $V_{(2,1)}$ (green) and $V_{(3)}$ (red)

Degree 4

For quartic polynomials there are 5 partitions: (1,1,1,1), (2,1,1), (2,2), (3,1) and (4). The inclusions of the corresponding varieties are pictured in the following diagram:

$$\begin{array}{ccccc} & & & V_{(3,1)} & \\ & & \swarrow & & \nwarrow \\ V_{(1,1,1,1)} = \mathbb{P}^4 & \longleftarrow & V_{(2,1,1)} & & V_{(4)} \\ & & \nwarrow & & \swarrow \\ & & & V_{(2,2)} & \end{array}$$

Here $V_{(2,1,1)} = Z_+(\Delta)$ is the discriminant hypersurface

$$\begin{aligned} \Delta = \; & 256\, a^3e^3 - 192\, a^2b\, d\, e^2 - 128\, a^2c^2e^2 + 144\, a^2c\, d^2e - 27\, a^2d^4 + 144\, a\, b^2c\, e^2 \\ & -6\, a\, b^2d^2e - 80\, a\, b\, c^2d\, e + 18\, a\, b\, c\, d^3 + 16\, a\, c^4e - 4\, a\, c^3d^2 - 27\, b^4e^2 \\ & +18\, b^3c\, d\, e - 4\, b^3d^3 - 4\, b^2c^3e + b^2c^2d^2. \end{aligned}$$

The variety $V_{(3,1)}$ is an irreducible surface and parameterizes quartics with one triple root. It is defined by one quadric and one cubic polynomial:

$$\mathfrak{a} = (12\, a\, e - 3\, b\, d + c^2, 27\, a\, d^2 + 27\, b^2e - 27\, b\, c\, d + 8\, c^3),$$

and $V_{(2,2)}$ is also a surface. It is defined by the ideal

$$\begin{aligned} \mathfrak{a} \; = \; & (8\, b\, e^2 - 4\, c\, d\, e + d^3, 16\, a\, e^2 + 2\, b\, d\, e - 4\, c^2e + c\, d^2, 8\, a\, d\, e - 4\, b\, c\, e + b\, d^2, a\, d^2 - b^2e, \\ & 8\, a\, b\, e - 4\, a\, c\, d + b^2d, 16\, a^2e + 2\, a\, b\, d - 4\, a\, c^2 + b^2c, 8\, a^2d - 4\, a\, b\, c + b^3). \end{aligned}$$

This is a surface of degree 4 in $\mathbb{P}^4$, which in fact, is a projection of the Veronese surface in $\mathbb{P}^5$.

Finally, $V_{(4)}$ corresponds to quartics with a quadruple root and is defined by the ideal:

$$\mathfrak{a} = (8\, c\, e - 3\, d^2, 6\, b\, e - c\, d, 9\, b\, d - 4\, c^2, 36\, a\, e - c^2, 6\, a\, d - b\, c, 8\, a\, c - 3\, b^2),$$

and as before, this is exactly the rational normal curve, up to a linear coordinate change.

These varieties are usually highly singular. For instance, $V_{(2,1,1)}$ is singular along both $V_{(2,2)}$ and $V_{(3,1)}$. The surface $V_{(2,2)}$ is smooth, whereas $V_{(3,1)}$ is singular along the curve $V_{(4)}$.

5.8 Appendix: Bihomogeneous polynomials

We shall work with two sets of variables $x_0, \ldots, x_n$ and $y_0, \ldots, y_m$, and use the shorthand notation $f(x, y)$ for polynomials $f(x_0, \ldots, x_n, y_0, \ldots, y_m)$.

5.61 A polynomial $f(x, y)$ is said to be *bihomogeneous* of bidegree (a, b) if is homogeneous of degree a when considered a polynomial in the x_i's and of degree b when considered a polynomial in the y_j's. Another way of phrasing this, is to say that

Bihomogeneous polynomials
bihomogene polynomer

$$f(sx, ty) = s^a t^b f(x, y)$$

for all scalars s and t. Independently scaling the x_i's and the y_i's by a scalars s and t gives an action of the multiplicative group $k^* \times k^*$ of pairs of non-zero scalars on the polynomial ring $k[x_0, \ldots, x_n, y_0, \ldots, y_m]$. A pair (s, t) sends a polynomial f to the polynomial $f^{(s,t)}(x, y) = f(sx, ty)$. For a polynomial, being bihomogeneous is then equivalent to being an eigenvector for this action.

Recollecting all monomial terms of the same bidegree one sees that any polynomial may be expressed as the sum of bihomogeneous components (one for each bidegree, but most of them will be zero) and that these components are unique.

5.62 An ideal $\mathfrak{a}$ in the polynomial ring $k[x_0, \ldots, x_n, y_0, \ldots, y_m]$ is called *bihomogeneous* if all the bihomogeneous components of a polynomial f belong to $\mathfrak{a}$ when f does. Adapting the proof from the homogeneous case one proves the following characterisations of bihomogeneous ideals.

Bihomogeneous ideals
bihomogene idealer

Proposition 5.63 *Let $\mathfrak{a}$ be an ideal in $k[x_0, \ldots, x_n, y_0, \ldots, y_m]$. The following three statements are equivalent.*

i) The ideal $\mathfrak{a}$ is bihomogeneous;

ii) The ideal $\mathfrak{a}$ can be generated by bihomogeneous polynomials;

iii) For all $(s, t) \in k^* \times k^*$ it holds true that $\mathfrak{a}^{(s,t)} = \mathfrak{a}$; that is, $\mathfrak{a}$ is invariant under independent scaling of the sets of variables.

Proof: The first assertion implies the second because $\mathfrak{a}$ will be generated by all the bihomogeneous components of members of a set of generators, and the third ensues directly from the second. The only thing requiring some work is therefore that the first assertion follows from the last.

To see that, we shall apply induction on the number of bihomogeneous components of a polynomial $f \in \mathfrak{a}$; if there is just one, the statement is a tautology. So assume f given and decompose f as $f = \sum_{a,b} f_{a,b}$ in its bihomogeneous components. Furthermore, let (a_0, b_0) be the smallest among the bidegrees of the components of f in the lexicographical order. We have

$$f(sx, ty) = s^{a_0} t^{b_0} f_{a_0,b_0} + \sum_{a,b} f_{a,b}$$

where the sum extends over bidegrees different from (a_0, b_0). And this gives

$$f(sx, ty) - s^{a_0}t^{b_0}f(x, y) = \sum_{a,b}(s^a t^b - s^{a_0}t^{b_0})f_{a,b}$$

Now, observe that the left side belongs to $\mathfrak{a}$ since both $f(sx, ty)$ and $f(x, y)$ lie there. The sum to the right has less terms than the original decomposition of f so that induction applies, and moreover, all the coefficients $s^a t^b - s^{a_0}t^{b_0}$ are nonzero for a generic choice of s and t (the ground field is algebraically closed, hence infinite); indeed, $a_0 \leqslant a$ and in case of equality it holds that $b_0 < b$. We may then conclude, by induction, that the $f_{a,b}$'s all belong to $\mathfrak{a}$. ❏

5.64 An immediate corollary is that the radical of a bihomogeneous ideal $\mathfrak{a}$ is bihomogeneous. Indeed, the radical $\sqrt{\mathfrak{a}}$ equals the intersection of all the prime ideals containing $\mathfrak{a}$ and as $\mathfrak{a}^{(s,t)} = \mathfrak{a}$ for all s, t it holds true that $\mathfrak{p}^{(s,t)}$ contains $\mathfrak{a}$ whenever $\mathfrak{p}$ does. The set of primes containing $\mathfrak{a}$ is therefore invariant under scaling, and hence the intersection is as well.

Exercises

5.21 Let $\mathfrak{a}$ be a bihomogeneous ideal.

a) Prove that the set of associated primes and the set of isolated primary components of $\mathfrak{a}$ are invariant under the scaling action of $k^* \times k^*$ HINT: Use uniqueness.

b) Prove that the associated prime ideals and the isolated components of $\mathfrak{a}$ are bihomogenenous. HINT: Prove there are no subgroups of finite index in $k^* \times k^*$.

c) Prove that for an embedded associated prime $\mathfrak{p}$ there are $\mathfrak{p}$-primary components of $\mathfrak{a}$ that are bihomogenenous. HINT: Prove that if $\mathfrak{q}$ is any $\mathfrak{p}$-primary component, then the intersection $\bigcap_{s,t} \mathfrak{q}^{(s,t)}$ is one as well.

5.22 This exercise is a generalisation of appendix. Let G be a group acting on the Noetherian ring R. Let $\mathfrak{a}$ be an ideal that is invariant under the action; that is, $\mathfrak{a}^g = \mathfrak{a}$ for alle $g \in G$ (where $\mathfrak{a}^g = \{ g(x) \mid x \in \mathfrak{a} \}$). Assume that G has no subgroups of finite index.

a) Prove that if $\mathfrak{a}$ is invariant under G, then the same holds for the radical $\sqrt{\mathfrak{a}}$.

b) Prove that all the primes associated to $\mathfrak{a}$ are invariant under G.

c) Prove that all the isolated primary components of $\mathfrak{a}$ are invariant.

d) Prove that for each embedded associated prime $\mathfrak{p}$, the ideal $\mathfrak{a}$ has at least one invariant $\mathfrak{p}$-primary component.

e) Prove that $k^* \times \ldots \times k^*$ has no subgroup of finite index (regardless of the characteristic of k, but k must be algebraically closed). Prove that the same holds for any connected Lie group. HINT: In both cases the maps $g \mapsto g^n$ are surjective.

★

Chapter 6

Birational maps and blowing up

Topics in Chapter 6: *Rational maps – maximal set of definition – function fields – birational maps and birational equivalence – normalization – Cremona group – blowing up – flops*

Two varieties X and Y over the field k are said to be *birationally equivalent* if they have isomorphic non-empty open subsets, that is, one may find open dense subsets $U \subseteq X$ and $V \subseteq Y$ and an isomorphism $U \simeq V$. As open dense subsets have the same function field as the surrounding variety, any two birationally equivalent varieties have isomorphic function fields (as algebras over the ground field). In this chapter we shall see that the converse also holds. The theory of varieties up to birational equivalence is thus basically equivalent to the theory of fields finitely generated over the ground field.

Birationally equivalent varieties
Birasjonalt ekvivalente varieteter

The problem of classifying varieties up to birational equivalence is known as *birational geometry*. This is a much coarser classification than classification up to isomorphism, and hence it is *a priori* an easier task (but still, challenging enough). However, for non-singular projective curves, as we later shall see, the two are equivalent. Two such curves are isomorphic if and only if they are birationally equivalent – that is, if and only if their function fields are isomorphic over k.

Shigefumi Mori (1951–)
Japanese Mathematician

Already for projective non-singular surfaces, the situation is completely different. There are infinitely many non-isomorphic surface in the same birational class (see example 6.17 below for a simple example of two), and they can form a very complicated hierarchy. For varieties of higher dimension, the picture is even more complicated, but the so-called *Minimal Model Program* has evolved during the last 40 years, and shed much light on the situation.

Another important question is whether there are non-singular varieties[1] in every birational class; or phrased differently, whether every variety X is birationally equivalent to a non-singular one — or has a *non-singular model* as one also says. For curves this easy: a curve X is non-singular if and only if all its local rings are integrally closed in the function field $k(X)$, and by a normalization procedure, one achieves a non-singular model of X. For surfaces it is substantially more complicated, but it was proven for surfaces by several people, let us mention Zariski who proved it form surface over algebraically

[1] We have not yet defined this important class of varieties (we shall do so in Chapter 8); informally one may say that they play the role of manifolds in algebraic geometry.

closed fields of characteristic zero and by Shreeram Abhyankar for surfaces over algeraically closed fields of positive characteristic. In higher dimension, we have *resolution of singularities* when the ground field is of characteristic zero, as proved in the 1960s by Heisuke Hironaka. However, it is a long-standing open problem whether such a procedure exists in positive characteristic.

Heisuke Hironaka (1931–)
Japanese Mathematician

6.1 *Rational and birational maps*

6.1 Just like we spoke about rational functions on a variety being function defined and regular on a non-empty open subset, one may speak about *rational maps* from a variety X to another Y. Strictly speaking, these are pairs consisting of an open subset $U \subseteq X$ and a morphism $\phi\colon U \to Y$. Commonly a rational map is indicate by a broken arrow like $\phi\colon X \dashrightarrow Y$.

Rational maps
rasjonale avbildninger

EXAMPLE 6.2 Let $U \subseteq \mathbb{P}^2$ be the open subset $U = D_+(x_0) \cap D_+(x_1) \cap D_+(x_2)$ where all three homogeneous coordinates are non-zero. In $D_+(x_0)$ one has affine coordinates $t_1 = x_1/x_0$ and $t_2 = x_2/x_0$, and in U they are both non-vanishing. The assignment $(t_1, t_2) \mapsto (1/t_1, 1/t_2)$ therefore gives a morphism $U \to U$, and hence a rational map $\sigma\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$. ☆

EXAMPLE 6.3 Consider the projective line $\mathbb{P}^1$ with homogeneous coordinates $(x_0 : x_1)$. For each $r \in \mathbb{N}$ the map $x_1/x_0 \mapsto x_0^r/x_1^r$ is a morphism from $D_+(x_0)$ to $D_+(x_1)$, hence gives a rational map $\mathbb{P}^1 \dashrightarrow \mathbb{P}^1$. ☆

6.4 If V is another open subset of X containing U, an *extension* of ϕ to V is a morphism $\psi\colon V \to Y$ such that $\psi|_U = \phi$; it is common usage to say that ϕ is defined on V. An open subset $U \subseteq X$ is called a *maximal subset of definition* for ϕ if ϕ is defined on U and cannot be extended to any strictly larger open subset. The next proposition tells us that every rational map has a unique maximal set of definition:

Extension of morphisms
utvidelse av morfier

Maximal subsets of definition
maksimale definisjons mengder

PROPOSITION 6.5 *Let X and Y be two varieties, and $\phi\colon X \dashrightarrow Y$ a rational map. Then ϕ has a unique maximal set of definition U_ϕ.*

PROOF: Since X is a Noetherian topological space, any non-empty collection of open subsets has a maximal element. Hence maximal sets of definition exists, and merely the unicity statement requires some work.

Let $U \subseteq X$ be an open subset where ϕ is defined. Assume that V_1 and V_2 are open subsets of X containing U and that both are maximal sets of definition for ϕ. Let the two extensions be ϕ_1 and ϕ_2. Both restrict to morphisms on the intersection $V_1 \cap V_2$, and the salient point is that these two restrictions coincide. Indeed, both ϕ_1 and ϕ_2 restrict to ϕ on U, and because Y is a variety the Hausdorff axiom holds for Y. Consequently, the subset of $V_1 \cap V_2$ where ϕ_1 and ϕ_2 coincide, is closed; and since they coincide on U, which is dense in $V_1 \cap V_2$, they coincide along the entire intersection $V_1 \cap V_2$. This means that ϕ_1

and ϕ_2 can be patched together to give a map defined on $V_1 \cup V_2$, which is a morphism (being a morphism is a local property). My maximality, it follows that $V_1 = V_2$. ❑

Thus every rational map has a canonical set U_ϕ where it is defined. And in the sequel whenever we speak about a rational function without specifying the set of definition, it will tacitly be understood that we consider it a function on the maximal set of definition U_ϕ. Of course, one still is free to consider ϕ on any smaller open set (but one cannot do so secretly).

EXAMPLE 6.6 The rational map from Example 6.2 extends to open set $U_\sigma = \mathbb{P}^2 \backslash \{(0:0:1),(0:1:0),(1:0:0)\}$. Away from the three points, the assignment $(x_0 : x_1 : x_2) \mapsto (x_1x_2 : x_0x_2 : x_0x_1)$ is meaningful and defines a morphism. For points in the open set $D_+(x_0) \cap D_+(x_1) \cap D_+(x_2)$ we may scale by $1/x_0$ and $1/x_1x_2$ resepectively, and the assignment becomes $(1 : x_1/x_0 : x_2/x_0) \mapsto (1 : x_0/x_1 : x_0/x_2)$, and we recover the map in Example 6.2. The set U_σ is the maximal open set where σ is defined. ☆

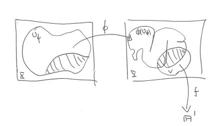
EXERCISE 6.1 Give an example to show that the proposition does not hold when X merely is a prevarietry. HINT: Take a new look at 'the bad guy'. (Example 3.60 on page 63). ★

Functoriality

6.7 In general rational maps can not be composed. If the image of one map lies entirely within the complement of the maximal set of definition of the other one, there is no way of composing them. To avoid such a calamity, one introduces the concept of a dominating map. One says that a rational map ϕ from X to Y is *dominating* if $\phi(U_\phi)$ is a dense set in Y (be aware that the image is not necessarily open, but though it can be dense).

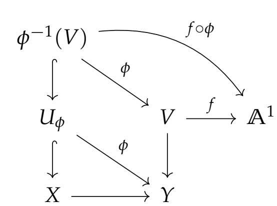
Dominating rational maps
dominerende rasjonale avbildninger

A dominant rational map $\phi\colon X \dashrightarrow Y$ may be composed with every rational map $\psi\colon Y \dashrightarrow Z$. As $\phi(U_\phi)$ is dense, the intersection $V_\psi \cap \phi(U_\phi)$ is not empty, and so $\phi^{-1}(V_\psi)$ is non-empty as well. But being open, $\phi^{-1}(V_\psi)$ is also dense and meets U_ϕ in the open and dense subset $W = U_\phi \cap \phi^{-1}(V_\psi)$. And on W the composition $\psi \circ \phi|_W$ is meaningful. Thus varieties and dominant rational maps form a category Rat_k whose objects are the varieties over k and whose morphisms are the dominant rational maps[2].



[2] By convention, we consider them to be morphisms on their maximal set of definition

6.8 Observe that a rational maps from X to $\mathbb{A}^1$ is just the same as a rational function, and so the set of rational maps $X \dashrightarrow \mathbb{A}^1$ is noting else than the rational function field $k(X)$. If now $\phi\colon X \dashrightarrow Y$ is dominating, taking $Z = \mathbb{A}^1$ in the above, we see that composition gives a map $\phi^*\colon KY) \to k(X)$ between the two function fields. It is homomorphism simply because addition and multiplication of functions are defined pointwise, so we obtain a contravariant functor from Rat_k to the category of finitely generated field extensions of k. An important

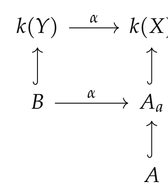


property is that the construction of ϕ^* is reversible:

Theorem 6.9 (Main Theorem of Rational maps) *Given two varieties X and Y and a k-algebra homorphism $\alpha\colon k(Y) \to k(X)$. Then there exists a (unique) dominant rational map $\phi\colon X \dashrightarrow Y$ such that $\phi^* = \alpha$.*

Proof: We begin by choosing open and affine sets, one in each of the varieties X and Y — call them U and V — with $U \subseteq X$ and $V \subseteq Y$. They have coordinate rings $A = \mathcal{O}_X(U)$ and $B = \mathcal{O}_Y(V)$, and the function fields $k(X)$ and $k(Y)$ are the fraction fields of A and B respectively. As U and V were randomly chosen, there is no reason for the homomorphism α to send B into A, but replacing B by an appropriate localization, we may arrange the situation for that to be true.

The k-algebra B is finitely generated over k and has generators $b_1, \ldots, b_s$. The images $\alpha(b_i)$ are of the form $\alpha(b_i) = a_i a^{-1}$ with the a_i's and a all belonging to A (the field $k(X)$ is the fraction field of A and a a common denominator for the $\alpha(b_i)$'s). But then α sends B into the localized ring A_a.



Translating this little piece of algebra into geometry will finish the proof. The localization A_a is the coordinate ring of the distinguished affine open subset U_a of U, and by the main theorem about morphisms between affine varieties, there is a morphism $\phi\colon U_a \to V$ with ϕ^* equal to $\alpha|_V$. Hence ϕ represents a rational and dominating map with the requested property that $\alpha = \phi^*$ ❑

Corollary 6.10 *Two varieties X and Y are birationally equivalent if and only if the function fields are isomorphic extensions of k.*

Formulated in a slightly more general way, we have:

Corollary 6.11 *The category Rat_k is equivalent to the category of finitely generated field extensions of k.*

Example 6.12 The map σ from Examples 6.2 is again a good example It can be composed with itself, and clearly $\sigma^2 = \mathrm{id}_U$, so that σ is birational. This shows that the open set W in the Pragraph 6.7 is not necessarily the maximal open set where the composition is defined.

One easily checks that a line through two of the coordinate points is collapsed to coordinate point by σ (for instance, the line $x_0 = 0$ is map to $(1:0:0)$), so the composition is *a priori* not defined in the whole maximal set of definition U_σ, eventhough it *a fortiori* can be extended to the entire $\mathbb{P}^2$. ☆

Birational maps

6.13 A *birational map* is a rational map which has a rational inverse. To give *Birational maps* *birasjonale avbildninger* a birational map between two varieties X to Y is to give open sets $U \subseteq X$ and $V \subseteq Y$ and an isomorphism $\phi\colon U \to V$, and when such a map is extant, one says that X and Y are *birationally equivalent*. Be aware that the open set U might be smaller than the maximal set of definition U_ϕ as in Example 6.14 below.

Birationally equivalent varieties
birasjonalt ekvivalente varieteter

In the vernacular of category theory one would express this by saying they are isomorphic in the category Rat_k. The main theorem (Theorem 6.9 above) tells us that X and Y are birationally equivalent if and only if their functions fields $k(X)$ and $k(Y)$ are isomorphic as k-algebras.

Examples

6.14 *(A quadratic transform)* It is pretty obvious that the map σ from Problem 6.6 on page 135 and which we also met in Examples 6.2 and 6.6 above, is birational — it is even its own inverse. Recall that it is given as $\sigma(x_0 : x_1 : x_2) = (x_0^{-1} : x_1^{-1} : x_2^{-1})$. It is regular on the open set $D_+(x_0x_1x_2)$ where no coordinate vanishes, and it maps $D_+(x_0x_1x_2)$ into itself. Clearly σ^2 equals the identity on $D_+(x_0x_1x_2)$. The map σ is called a *quadratic transform*.

Quadratic transforms
kvadratisk transformasjon

It is worth while understanding the map σ better. Multiplying all components by $x_0x_1x_2$ we obtain the expression $\sigma(x_0 : x_1 : x_2) = (x_1x_2 : x_0x_2 : x_0x_1)$, which reveals that σ is defined away from the three 'coordinate points' $e_2 = (0;0;1)$, $e_1 = (0;1;0)$ and $e_0 = (1;0;0)$, since if two coordinates do not vanish neither does their product. It also reveals that each of the three lines $L_0 = V(x_0)$, $L_1 = V(x_1)$ and $L_{x_2} = V(2)$ are collapsed to the corresponding coordinate point. For instance, points on L_0 are of the form $(0 : x_1 : x_2)$ and σ maps them all to the point $(x_1x_2 : 0 : 0) = (1 : 0 : 0) = e_0$ (the equality being valid whenever $x_1x_2 \neq 0$).

The map σ can not be extended beyond $\mathbb{P}^2 \backslash \{e_0, e_1, e_2\}$. Indeed, through each of the coordinate points pass two of the lines that are collapsed, and the two lines are mapped to *different* points by σ. Therefore, by continuity, we are left no chance of defining σ at the coordinate points. So $\mathbb{P}^2 \backslash \{e_0, e_1, e_2\}$ is the maximal set of definition for σ.

6.15 *(Birational automorphisms)* Note that the set of birational self-maps $\phi : X \dashrightarrow X$ form a *group* under composition. We denote this group by $\mathrm{Bir}(X)$. This is an important invariant of X; one can use it to prove that two varieties are not birational to each other. The case $X = \mathbb{P}^n$ is particularly important; $\mathrm{Bir}\,\mathbb{P}^n$ is called the n-th *Cremona groups* named after the Italian mathematician Luigi Cremona. We denote this group by $Cr_n(k)$. In view of Theorem 6.9 on the facing page this group is nothing but the Galois group of $k(x_1, \ldots, x_n)$ over k.


Antonio Luigi Gaudenzio Giuseppe Cremona (1830–1903) Italian mathematician

Automorphisms $\phi : X \to X$ are of course birational; hence we get a subgroup

$$\mathrm{Aut}(X) \subset \mathrm{Bir}(X)$$

Cremona groups
Cremona-grupper

In particular, the *projective linear group* $\mathrm{PGL}(k, n+1) = \mathrm{GL}(n+1, k)/k^*$ acts on

$\mathbb{P}^n$ by automorphisms; and $Cr_n(l)$ contains it as a subgroup. We note however, that $Cr_n(k)$ is in general much bigger than $\mathrm{Aut}(\mathbb{P}^n)$.

Except for $n = 1$ and $n = 2$ nothing much is known about these groups. For $n = 1$ it is just the group $\mathrm{PGL}(2, k)$ — every automorphism $\mathbb{P}^1$ is linear. For $n = 2$ there is a famous theorem of Max Noether's (the father of Emmy Noether) that any birational automorphism of $\mathbb{P}^2$ is a composition of quadratic transforms and linear automorphisms; in other words, the Cremona group $Cr_2(k)$ is generated by $\mathrm{PGL}(3, k)$ and the map σ.

Max Noether (1844–1921)
German mathematician

For $n \geqslant 3$, the group $Cr_n(k)$ is enormous. In particular, it is a theorem by Hudson and Pan that $Cr_n(k)$ would require an uncountable set of generators. In particular, $Cr_3(k)$ is so huge that it admits no simple description as in the cases $n = 1$ and $n = 2$.

Nevertheless, the fact that $\mathbb{P}^n$ admits so many birational automorphisms is actually a quite useful thing: it makes the group $\mathrm{Bir}(X)$ a useful invariant for proving that a given variety is irrational. The main example of this is due to Iskovskikh and Manin (1970), who showed that a smooth hypersurface $X \subset \mathbb{P}^4$ of degree 4 has finite birational automorphism group. In particular, such an X cannot be rational. This is a highly non-trivial result, because the hypersurface X really behaves like a rational variety in many respects. For instance, it is *unirational*, in the sense that there is a dominant rational map $\mathbb{P}^3 \dashrightarrow X$.

6.16 *(Maps from $\mathbb{P}^1$)* Any rational map $\mathbb{P}^1 \to \mathbb{P}^n$ is defined everywhere[3]; in other words, the maximal set of definition U_ϕ of ϕ is equal to the entire projective line $\mathbb{P}^1$. Chose homogeneous coordinates $(x_0 : \ldots : x_n)$ on $\mathbb{P}^n$ and let $D_+(x_i)$ be one of the distinguished open sets that meet the image of U_ϕ under ϕ. The variety $D_+(x_i)$ is an affine n-space with coordinates $x_j x_i^{-1}$ with $0 \leqslant j \leqslant n$ and $j \neq i$.

[3] This example is a forerunner for the general theorem asserting that any rational map from a regular curve into a projective space is in fact regular everywhere.

The inverse image $V = \phi^{-1}(D_+(x_i) \cap \phi(U_\pi))$ is an open set, and the n component functions of $\phi|_V$ are rational functions on $\mathbb{P}^1$ regular on V. They may be brought on the form f_j/f_i with $0 \leqslant j \leqslant n$ and $j \neq i$, where the polynomial f_i is a common denominator of the f_j's and does not vanish on V. At points in V the relation $x_j x_i^{-1} = f_j f_i^{-1}$ holds.

The idea is to use the f_k's (now including f_i) as the homogenous components of a morphism Φ of $\mathbb{P}^1$ into $\mathbb{P}^n$ and define the extension Φ by the assignment $\Phi(x) = (f_0(x) : \ldots : f_n(x))$. It might be that the $n+1$ polynomials f_k have a common factor, but it can be discarded, and we may assume that the f_k's are without common zeros. Then it is easily checked (remember Paragraph 4.49 on page 90) that Φ is a morphism that extends ϕ.

6.17 *(The quadratic surface)* In this example we use homogeneous coordinates $(x : y : z : w)$ on the projective space $\mathbb{P}^3$. The quadric $Q = Z_+(xz - yw) \subseteq \mathbb{P}^3$ is birationally equivalent to the projective plane $\mathbb{P}^2$, but the two are not isomorphic. This is one of the simplest example of two non-isomorphic projective and non-singular surfaces being birationally equivalent.

To begin with, the two are not isomorphic. They are not even homeomorphic since any two curves[4] in $\mathbb{P}^2$ intersect, but on the quadric there are families of disjoint lines (in fact, there are two such). For example the two disjoint lines $x = y = 0$ and $x + z = y + w = 0$ both lie on Q.

Next, we exhibit a birational map $\phi\colon Q \dashrightarrow \mathbb{P}^2$. It will be defined on the open set $U = D_+(w) \cap Q$. In $D_+(w) \simeq \mathbb{A}^3$, where we by mildly abusing the language, use coordinates named x, y and z, the equation of Q becomes, $y = xz$. It is almost obvious that the projection $\mathbb{A}^3 \to \mathbb{A}^2$ sending (x, y, z) to (x, z) induces an isomorphism from $Q \cap D_+(w)$ to $\mathbb{A}^2$, but it is a rewarding exercise for the students to check all details. HINT: The inverse map is given as $(x, z) \mapsto (x, xz, z)$.

[4] A *curve* is a closed subset of Krull dimension one; notice that this is a purely topological notion.

☆

EXERCISE 6.2 In this exercise the previous example is elaborated. Show that the projection $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with centre $p = (0 : 0 : 0 : 1)$ is well defined on $Q \backslash \{p\}$ and when restricted to $D_+(w) \cap Q$, becomes the isomorphism in the example. Show that the plane $Z_+(y)$ meets Q along two lines passing by p, and that these lines under the projection are collapsed to two different points in $\mathbb{P}^2$. HINT: The projection is given as $(x : y : z : w) \mapsto (x : y : z)$. ★

Exercises

6.3 Let $\phi\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ be the rational map that sends $(x_0 : x_1 : x_2)$ to $(x_2^2 : x_0x_1 : x_0x_2)$. Determine largest set of definition. Show that ϕ is birational , and determine what curves are collapsed.

6.4 Let $\phi\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ be the rational map that sends $(x_0 : x_1 : x_2)$ to $(x_2^2 - x_0x_1 : x_1^2 : x_1x_2)$. Determine the set largest set where ϕ is defined. Show that ϕ is birational, and determine what curves are collapsed.

6.5 Let $\phi\colon \mathbb{A}^2 \to \mathbb{A}^4$ be the map defined by

$$\phi(x, y) = (x, xy, y(y - 1), y^2(y - 1)).$$

a) Show that $\phi(0, 0) = \phi(0, 1) = (0, 0, 0, 0)$, and that ϕ is injective on $U = \mathbb{A}^2 \backslash \{(0, 0), (0, 1)\}$.
b) Show that $\phi|_U$ is an isomorphism between U and its image. HINT: $\phi|_U$ takes values in $V = \mathbb{A}^4 \backslash Z(x)$ and the map $V \to \mathbb{A}^2$ sending (u, v, w, t) to (u, v) is a left section for $\phi|_U$.
c) Show that the image of ϕ is given by the polynomials $ut - vw$, $w^3 - t(t - w)$ and $u^2w - v(v - u)$.

6.6 Let $\sigma\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ be the map defined as $\sigma(x_1 : x_2 : x_3) = (x_2x_3 : x_1x_3 : x_1x_2)$. Show that σ is defined off the three points $p_1 = (0 : 0 : 1)$, $p_2 = (0 : 1 : 0)$ and $p_3 = (1 : 0 : 0)$ and that the line $Z(x_i)$ is mapped to the point p_i. Show that

$U_\sigma = \mathbb{P}^2 \backslash \{p_1, p_2, p_3\}$ is the maximal set of definition of σ. Show that $\sigma^2 = \mathrm{id}_{\mathbb{P}^2}$ and that on the set $U = D_+(x_1) \cap D_+(x_2) \cap D_+(x_3)$ σ restricts to the σ defined in Example 6.2 on page 130.

6.7 Consider the map $(u : v) \mapsto (u^2 v^{-2} : u^3 v^{-3} : 1)$. Prove that it is morphism on $\mathbb{P}^1 \backslash \{(1:0)\}$. Prove that it can be extended to a morphism $\mathbb{P}^1 \to \mathbb{P}^2$.

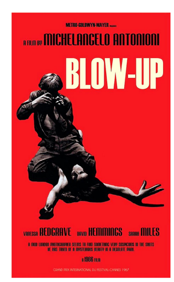
6.8 Let $\phi\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ be the rational map given by $\phi(x : y : z) = (y^2 : xy : x^2)$. Show that ϕ is a morphism away from the point $p = (0 : 0 : 1)$, and that image is the conic C parametrized by $(u^2 : -uv : v^2)$. Show that line $ax + by$ passing by p is mapped to the point $(a^2 : -ab : b^2)$.

Show that $\psi(x : y : z) = (x^2 : xy : xz : yz : y^2)$ is a rational map $\mathbb{P}^2 \dashrightarrow \mathbb{P}^4$ defined in $U = \mathbb{P}^2 \backslash (0 : 0 : 1)$. Show that the image is contained in the closed set given by the equations $x_0 x_4 - x_1^2 = x_1 x_3 - x_4 x_2 = x_1 x_2 - x_0 x_3 = 0$. Describe the image.

6.2 Blowing up

There is in some sense an 'atomic' way of birationally modifying a variety, and in an ideal situation one might hope for expressing a given birational map as a sequence of these smallest modifications. This is a subtle and difficult issue, far from out agenda; here we shall only give a few examples of these 'blowing ups'.

6.18 From a variety Y, containing a closed subset Z, the *blow-up* of Y along Z is constructed as a new variety, X, together with a birational morphism $\pi\colon X \to Y$, which is an isomorphism over the open set $Y \backslash Z$. And the inverse image of Z will be a 'divisor'; that is, it will be of codimension one, whatever the codimension of Z might be. So $\pi^{-1}(Z)$ is in most cases 'much bigger' than Z, and the name 'blow-up' is deserved.



A poster for the movie 'Blow-Up' (1966)

6.19 The simplest example is the case of a point in $\mathbb{P}^2$, which is quite illustrative and, at least for blow-ups of regular points on surfaces, even prototypical. After having chosen coordinates we may assume that the point is $(0 : 0 : 1)$, and the homogeneous coordinates are $(x_0 : x_1 : x_2)$. The blown up plane $\mathrm{Bl}_p \mathbb{P}^2$ will be the closed subvariety of the product $\mathbb{P}^2 \times \mathbb{P}^1$ given by the bihomogeneous[5] equation

$$u_1 x_0 - u_0 x_1 = 0, \tag{6.1}$$

where $(u_0 : u_1)$ are homogeneous coordinates on $\mathbb{P}^1$. The blow-up morphism $\pi\colon \mathrm{Bl}_p \mathbb{P}^2 \to \mathbb{P}^2$ is just the restriction of the first projection to $\mathrm{Bl}_p \mathbb{P}^2$,

[5] We shall see more to bihomogeneous polynomials later. Note that scaling either the x_i's or the u_i's result in scaling the equation, so the zeros are well defined. That the zero locus is closed is a general fact, but in the present case follows from the discussion right after the proof

Proposition 6.20 *The projection $\pi\colon \operatorname{Bl}_p \mathbb{P}^2 \to \mathbb{P}^2$ is birational. The fibre $E = \pi^{-1}(p)$ is equal to $\{p\} \times \mathbb{P}^1$ and π induces an isomorphism between $\operatorname{Bl}_p \mathbb{P}^2 \backslash E$ and $\mathbb{P}^2 \backslash \{p\}$.*

The fibre E is sometimes called *the exceptional locus* or the *exceptional divisor*.

The exceptional locus
det eksepsjonelle lokuset

The exceptional divisor
den eksepsjonelle divisoren

Proof: That the set $E = \{p\} \times \mathbb{P}^1$ lies in the fibre over p comes for free: points in E are shaped like $(u_0 : u_1) \times (0 : 0 : 1)$, and equation (6.1) is evidently satisfied in such points.

To prove the statement that π is an isomorphism off E, consider the rational map $\psi\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^1$ that sends $(x_0 : x_1 : x_2)$ to $(x_0 : x_1)$ is a well defined morphism in $\mathbb{P}^2 \backslash \{p\}$ and hence we obtain a morphism ϕ into $\mathbb{P}^2 \times \mathbb{P}^1$ by sending x to $(x, \psi(x))$ (which by the way is just the graph of ψ). In coordinates

$$\phi(x_0 : x_1 : x_2) = (x_0 : x_1) \times (x_0 : x_1 : x_2),$$

so obviously equation (6.1) is satisfied along the image of ϕ, and consequently ϕ takes values in $\operatorname{Bl}_p \mathbb{P}^2$. Moreover, it clearly holds that $\pi \circ \phi = \operatorname{id}_{\mathbb{P}^2 \backslash \{p\}}$. It remains to see that the image of ϕ is equal to the complement of the exceptional fibre.

But if $q = (u_0 : u_1) \times (x_0 : x_1 : x_2)$ is a point on the product where (6.1) holds , it holds true that $u_1 x_0 = u_0 x_1$, so if, for instance, $x_0 \neq 0$, one finds

$$\begin{aligned} q = (u_0 : u_1) \times (x_0 : x_1 : x_2) = (x_0 u_0 : x_0 u_1) \times (x_0 : x_1 : x_2) = \\ = (x_0 u_0 : u_0 x_1) \times (x_0 : x_1 : x_2) = (x_0 : x_1) \times (x_0 : x_1 : x_2), \end{aligned}$$

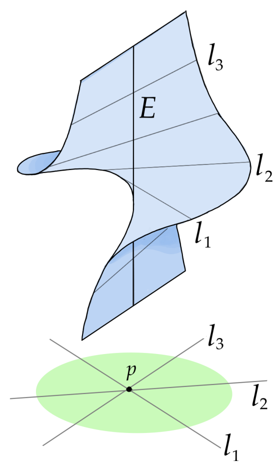
since $x_0 \neq 0$ implies that $u_0 \neq 0$. Thus q lies in in the image of ϕ. ❏

The blown up plane has a covering of four open affine sets U_i all isomorphic to $\mathbb{A}^2$. The basic open sets $D_+(x_0)$ and $D_+(x_1)$ of $\mathbb{P}^2$ lie in the part where the blow-up morphism is birational, and thus they give isomorphic open sets in the blow-up.

The set $D_+(x_2)$, however, is much more interesting. In what follows we shall identify $D_+(x_2)$ with the affine plane $\mathbb{A}^2$ using the coordinates $x = x_0 x_2^{-1}$ and $y = x_1 x_2^{-1}$, so the points are shaped like $(x : y : 1)$. The inverse image $\operatorname{Bl}_p \mathbb{A}^2$ of $D_+(x_2)$ (which is the blow-up of $\mathbb{A}^2$) contains the exceptonal divisor E and is therefore no more affine; one needs two open affines U_0 and U_1 to cover it. These are inherited from the basic open cover of $E = (0 : 0 : 1) \times \mathbb{P}^1$ in that $U_i = \operatorname{Bl}_p(\mathbb{A}^2) \cap D_+(u_i)$, and both are naturally isomorphic to $\mathbb{A}^2$: in $\mathbb{P}^1$ it holds, for instance, that $D_+(u_0) = \mathbb{A}^1$ with coordinate $u_1 u_0^{-1}$, and so $\mathbb{A}^2 \times D_+(u_0) \simeq \mathbb{A}^3$ with affine coordinates (x, y, t) where $t = u_1 u_0^{-1}$. The equation (6.1) then reads $y = tx$, and we infer that U_0 is nothing but the image of the map $(x, t) \mapsto (x, tx, t)$, and this has the projection onto the first and last coordinate as inverse, so $U_0 \simeq \mathbb{A}^2$ with (x, t) as coordinates. Similarly, one finds that in U_1 it holds that $x = t^{-1} y = sy$ with $s = t^{-1}$, so that $U_1 \simeq \mathbb{A}^2$ with coordinates (y, s), and the maps is given as $(y, s) \mapsto (sy, y)$.

6.21 Notice that restricting the blow-up morphism to U_0 and using coordinates on U_0 as above, we obtain the map $\mathbb{A}^2 \to \mathbb{A}^2$ given as $(x,t) \mapsto (x,tx)$, which we have met on an earlier occasion (Example 2.59 on page 40).

It is also noteworthy that the points of the exceptional fibre E correspond bijectively to lines through p (frequently illustrated by drawing the blow-up as a winding staircase): equations shaped like $u_0x_1 - u_1x_0 = 0$, where u_0 and u_1 are scalars not both zero, are precisely the equations for lines trough p, and simultaneous scaling u_0 and u_1 gives the same line.



General blow-ups of affine varieties

Let $X \subset \mathbb{A}^n$ be an affine variety, and let $Z = Z(f_1,\ldots,f_r)$ be a closed subset. Imitating the construction of the previous paragraph, we form the *blow-up of* $\mathbb{A}^n$ *along* Z as follows. We consider the open set $U = X - Z$ and the morphism

Blow-up of a closed subset
oppblåsning av en lukket undermengde

$$\begin{aligned} \phi\colon U &\to \mathbb{P}^{r-1} \\ x &\mapsto (f_1(x) : \ldots : f_r(x)), \end{aligned}$$

which is well defined since at least one of the f_i's does not vanish in a point outside Z. Let Γ_ϕ denote the graph of ϕ, *i.e.*

$$\Gamma_\phi = \{(x,\phi(x)) | x \in U\} \subset U \times \mathbb{P}^{r-1}.$$

We define the *blow-up* $\mathrm{Bl}_Z X$ as the closure of Γ_f inside $X \times \mathbb{P}^{r-1}$. By construction, $\mathrm{Bl}_Z X$ admits a morphism $\pi\colon \mathrm{Bl}_Z X \to X$, given by the first projection, and we call this the *blow-up morphism*. Note that π is an isomorphism when restricted to $\pi^{-1}(U)$ via the first projection, simply because Γ_ϕ is a graph. We call $E = \mathrm{Bl}_Z X - \pi^{-1}(U)$ the *exceptional locus* or the *exceptional divisor*. The morphism π is typically not an isomorphism along E: in the previous example π mapped E to a point.

The blow-up morphism
oppblåsningsmorfien

The exceptional locus
det eksepsjonelle lokus

The exceptional divisor
den eksepsjonelle divisoren

For a subvariety $Y \subset X$, which is not contained in Z, we can restrict the map ϕ to the intersection $U \cap Y$ and close up the image of $U \cap Y$ in $\mathrm{Bl}_Z X$, and in this way we obtain a subvariety $\widetilde{Y} \subset \mathrm{Bl}_Z X \subset X \times \mathbb{P}^{r-1}$. The subvariety $\widetilde{Y}$ is called the *strict transform* of Y via π, and $\widetilde{Y}$ is the blow-up of Y along $Y \cap Z$. The preimage $\pi^{-1}(Y)$ is called the *total transform*: we have $\pi^{-1}(Y) = \widetilde{Y} \cup e$, where $e \subset Y$ does not dominate Y via π.

The strict transform
den strenge transformen
The total transform
den totale transformen

In many cases, the equations of $\mathrm{Bl}_Z X$ may be described as follows. Note that the graph $\Gamma \subset U \times \mathbb{P}^{n-1}$ consists of the pairs $(x_1, \ldots, x_n) \times (y_1 : \ldots : y_n)$ so that $(y_1, \ldots, y_n)$ is proportional to $(f_1(x), \ldots, f_n(x))$. In other words, in those points the rank of the matrix

$$M = \begin{pmatrix} f_1(x) & f_2(x) & \ldots & f_r(x) \\ y_1 & y_2 & \ldots & y_r \end{pmatrix} \tag{6.2}$$

is equal to one. The matrix is meaningful for all points in $X \times \mathbb{P}^{r-1}$, so it follows that the 2×2-minors of M defines a closed subset $W \subset X \times \mathbb{P}^{r-1}$ containing the blow-up $\mathrm{Bl}_Z X$. In general, if the functions $f_1, \ldots, f_r$ are arbitrary, the inclusion $\mathrm{Bl}_Z X \subseteq W$ can be strict. (See problem xxx). However, if the f_i's form a *regular sequence*[6], then we have equality:

[6] We have not yet spoken about regular sequences, but take a look at appendix 11.5 to Chapter 11

Proposition 6.22 *Let $X \subset \mathbb{A}^n$ be an affine variety and let $f_1, \ldots, f_r$ denote non-zero elements in $A(X)$. Then if the $f_1, \ldots, f_r$ form a* regular sequence, *then the blow-up $\mathrm{Bl}_Z X$ of X in $Z = Z(f_1, \ldots, f_r)$ is defined by the 2×2-minors of the matrix* (6.2).

Example 6.23 Let us consider the blow-up of $\mathbb{A}^n$ at the origin $p = (0, \ldots, 0)$. Let $x_1, \ldots, x_n$ denote affine coordinates on $\mathbb{A}^n$ and $y_1, \ldots, y_n$ homogeneous coordinates on $\mathbb{P}^{n-1}$ and let $U = \mathbb{A}^n - \{p\}$. The morphism $\phi \colon U \to \mathbb{P}^{n-1}$ is simply the 'quotient space' morphism used in the construction of $\mathbb{P}^{n-1}$:

$$\begin{aligned} \phi : U &\to \mathbb{P}^{n-1} \\ (x_1, \ldots, x_n) &\mapsto (x_1 : \ldots : x_n). \end{aligned}$$

Let Γ denote the graph of this morphism; that is, $\Gamma \subset \mathbb{A}^n \times \mathbb{P}^{n-1}$ consists of the pairs $(x_1, \ldots, x_n) \times (y_1 : \ldots : y_n)$ so that $(y_1, \ldots, y_n)$ is proportional to $(x_1, \ldots, x_n)$. In other words, the rank of the matrix

$$M = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ y_1 & y_2 & \ldots & y_n \end{pmatrix}$$

is equal to one. Let $W \subset \mathbb{A}^n \times \mathbb{P}^{n-1}$ denote the subset defined by these minors (scaling either row, of course scales the minors). We claim that in fact $\mathrm{Bl}_Z X = W$ in this case. To prove this, we need only check that W is irreducible and of dimension n: since we know this to be true for $\mathrm{Bl}_Z X$, which equals the closure of Γ, the inclusion $\mathrm{Bl}_Z X \subset W$ will be an equality.

Consider the affine open set $U_1 = p_2^{-1}(D_+(y_1))$ in $\mathbb{A}^n \times \mathbb{P}^{n-1}$, where p_2 denotes the second projection. We know that $U_1 \simeq \mathbb{A}^n \times \mathbb{A}^{n-1} \simeq \mathbb{A}^{2n-1}$, where in the second factor, we may set $y_1 = 1$. One easily checks that $W \cap U_1$ is defined by the equations $x_i = y_i x_1$ for $i = 2, \ldots, n$; indeed, modulo these equations $x_i y_j - x_j y_i = x_1 y_i y_j - x_1 y_j y_i = 0$. Eliminating the x_i's with $i \neq 1$, we see that $W \cap U_1 \simeq \mathbb{A}^n$. The same happens for all the other charts $W \cap U_i$. The $W \cap U_i$'s form a cover of $\mathrm{Bl}_Z X$ and each intersection $W \cap U_i \cap U_j$ is non-empty. Hence $\mathrm{Bl}_Z X$ is irreducible of dimension n, as we wanted to show. ☆

Example 6.24 *(Blow-up of the twisted cubic)* A good example of an ideal which is not generated by regular sequence is the ideal of the affine cone over the twisted cubic, that is, $S = Z(I)$ where $I = (q_0, q_1, q_2)$ where

$$q_0 = xz - y^2, \quad q_1 = xw - yz, \quad q_2 = yw - z^2$$

While this is not a regular sequence, we can still consider the rational map $\phi : \mathbb{A}^4 \to \mathbb{P}^2$ given by $p \mapsto (q_0(p) : q_1(p) : q_2(p))$. It is clear that the graph of ϕ is contained in the ideal J generated by the 2×2-minors of

$$\begin{pmatrix} u_0 & u_1 & u_2 \\ xz - y^2 & xw - y * z & yw - z^2 \end{pmatrix}$$

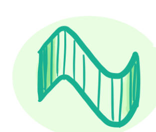
However that ideal is not prime; a primary decomposition of J is given by $a \cap I$, where

$$\mathfrak{a} = (wu_0 - zu_1 + yu_2, zu_0 - yu_1 + xu_2)$$

(This is no surprise; $Z(J)$ should have codimension 2, but $Z(f_0, f_1, f_2)$ is already of codimension 2 and contained in $Z(J)$). Thus the graph of ϕ is in fact defined by the prime ideal $\mathfrak{a}$. Writing the equations in $\mathfrak{a}$ in the following form reveals a bit more about the blow-up morphism $\pi : X \to \mathbb{A}^4$.

$$\begin{pmatrix} x & y & z \\ y & z & w \end{pmatrix} \cdot \begin{pmatrix} u_2 \\ -u_1 \\ u_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{6.3}$$

For a point $(x, y, z, w) \in \mathbb{A}^4 - S$ so that the matrix $M = \left(\begin{smallmatrix} x & y & z \\ y & z & w \end{smallmatrix}\right)$ has rank 2, the equation (6.3) has a unique solution $(u_0 : u_1 : u_2)$ (two linear equations in $\mathbb{P}^2$ define a point), so $\pi^{-1}(x, y, z, w)$ is a point. If $(x, y, z, w) \in S$, M has rank 1, and then the two equations reduce to one, so that the fiber $\pi^{-1}(x, y, z, w)$ is a $\mathbb{P}^1 \subset \mathbb{P}^2$.


X


$\mathbb{P}^3$

The equations (6.3) also define the blow-up of $\mathbb{P}^3$ along the twisted cubic. ☆

Exercise 6.9 Let f_1, f_2, f_3 be x^2, xy, y^2 respectively in $\mathbb{A}^2$. Show that the blow-up $\mathrm{Bl}_Z \mathbb{A}^2$ is in fact isomorphic to the blow-up of $\mathbb{A}^2$ at the origin $p = Z(x, y)$ (which lives in $\mathbb{A}^2 \times \mathbb{P}^1$). ★

Exercise 6.10 Show that $\mathbb{P}^1 \times \mathbb{P}^1$ is birational with $\mathbb{P}^2$. Show that $\mathrm{PGL}(2, k) \times \mathrm{PGL}(2, k)$ acts by automorphisms on $\mathbb{P}^1 \times \mathbb{P}^1$. Conclude that there is a subgroup $\mathrm{PGL}(2, k) \times \mathrm{PGL}(2, k) \subseteq \mathrm{Bir}\,\mathbb{P}^2$. different from $\mathrm{PGL}(3, k)$. ★

Exercise 6.11 Argue as in Example 6.24 to study the blow-up of $\mathbb{P}^2 \times \mathbb{P}^2$ along the diagonal Δ (which is given by the minors of a 2×3-matrix). ★

Chapter 7

More on dimension

Topics in Chapter 7: *Krull dimension and dimension of spaces – Finite maps, Lying–Over and Going–Up – Noether's Normalization Lemma – Transcendence degree and dimension – The dimension of $\mathbb{A}^n$ – Maximal chains in varieties – Varieties are catenary – The dimension of a product*

In Chapter 2, we defined the dimension of a topological space as the suprenum of the length r of chains

$$X_0 \subset X_1 \subset \ldots \subset X_r = X$$

of closed and irreducible subsets of X. This definition strongly resembles the Krull dimension of a ring, and indeed, we showed that $\dim X = \dim A(X)$ whenever X is an affine variety. In this chapter, we will use this link to algebra to give more refined results of dimensions of varieties, and in particular how it behaves with respect to morphisms. We will use the algebraic properties of finitely generated k-algebras, together with the tools of Cohen-Seidenberg (Lying over, Going-Up, Going-Down) as our main tools.

The simplest instance of this concerns *finite morphisms*, which are essentially morphisms with a finite number of preimages. Given a finite map $f : X \to Y$ it is intuitive that the dimension of X should not be larger than that of Y. While this is not true for general topological spaces (even for inclusions of open sets!), we shall see that it holds when X and Y are varieties.

A related notion is that of *dominant morphisms*, which are morphisms $f : X \to Y$ so that the image $f(X)$ is dense in Y. In this case, it is intuitive that the dimension of X should not be smaller than that of Y, and indeed this holds in the case of varieties.

Together these two results say that $\dim X = \dim Y$ for any finite and dominating morphism $f : X \to Y$. This is a highly useful result for computing dimensions. For instance it allows us to compute the dimension of a variety which is parameterized by a rational map $f : \mathbb{A}^n \dashrightarrow X$.

In the other direction, a number of results about dimension can be deduced from looking at finite and dominating morphisms $\pi : X \to \mathbb{A}^d$. It is a consequence of the Noether Normalization Lemma, proved by Emmy Noether, that such morphisms always exist for affine varieties $X \subset \mathbb{A}^n$. In fact, π is induced by a generic projection $\pi : \mathbb{A}^n \dashrightarrow \mathbb{A}^d$.

For a variety X there is another very good candidate for the dimension (which in many texts serves as the definition of the dimension), namely the transcendence degree of the function field $k<(X)$ over the ground field. Coupled with the Lying–Over and the Going–Up Theorems of Irvin Cohen and Abraham Seidenberg the Normalization Lemma leads to the basic result that the two ways of defining the dimension coincide.

We shall formulate and prove the Normalization Lemma in the geometric context we work; that is, over an algebraically closed field. However it remains true over any field, and over infinite fields the poof is *mutatis mutandis* the same as then one we give, but a slight twist is needed when the ground field is finite.

7.1 Some basic properties of polynomial maps

Dominant maps

7.1 Morphisms between varieties whose image is *dense* in the target, are called *dominant*. So for instance, any surjective morphism is dominant. Also, the inclusion of an open set $\iota\colon U \to X$ in an irreducible space is a dominant morphism. Dominant morphisms are somehow easier to handle than general morphisms between varieties, and often proofs can be reduced to the case of dominant maps.

Dominant morphisms
dominerende avbildninger

Note that a morphism $\phi\colon X \to Y$ is dominant if and only if any dense open U meets the image $\phi(X)$. This makes it clear that the composition of two dominant morphisms is dominant: indeed, if $\psi\colon Z \to X$ is the second dominant morphism and $U \subseteq X$ is dense and open, the image $\psi(Z)$ meets $\phi^{-1}(U)$, which is just another way of saying that $\phi(\psi(Z))$ meets U.

7.2 Suppose that X and Y are varieties and that $\phi\colon X \to Y$ is a dominant morphism. For any regular function f on Y that does not vanish identically, we may find an open dense set in U in Y where f does not have any zeros. Since by assumption the image $\phi(X)$ is dense in Y, the intersection $U \cap \phi(X)$ is non-empty, and it follows that $f \circ \phi$ does not vanish identically on $\phi^{-1}(U)$. In other words, the composition map $\phi^*\colon A(Y) \to A(X)$ is *injective*. This leads to

LEMMA 7.3 *A morphism $\phi\colon X \to Y$ between affine varieties is dominant if and only if the corresponding homomorphism $\phi^*\colon A(Y) \to A(X)$ is injective.*

PROOF: Half the proof is already done. For the remaining part, suppose that the image $\phi(X)$ is not dense. Its closure $Z = \overline{\phi(X)}$ in Y is then a proper closed subset, and $I(Z)$ is a non-zero ideal. Any function $f \in I(Z)$ vanishes along $\phi(X)$, and hence $\phi^*(f) = f \circ \phi = 0$, and ϕ^* is not injective. ❑

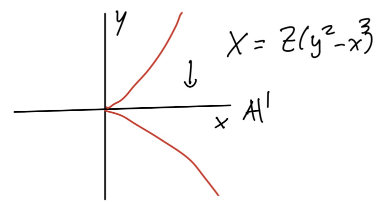
Finite maps

7.4 A polynomial map $\phi\colon X \to Y$ between two closed algebraic sets X and Y is said to be *finite* if the composition map $\phi^*\colon A(Y) \to A(X)$ makes $A(X)$ into a finitely generated $A(Y)$-module.

Finite polynomial maps
endelige avbildninger

More generally, a morphism $\phi\colon X \to Y$ between two varieties is said to be *finite* if inverse image of affine open subsets, are affine and every point $y \in Y$ has an affine open neighbourhood U such that the restriction $\phi|_{\phi^{-1}(U)}$ is a finite polynomial map. A morphism having the first property, that inverse image of open affines are open affines, is referred to as being *affine*. One easily checks that the composition of two finite morphisms (or polynomial maps) is finite.

Finite morphisms
endelige morfismer



EXAMPLE 7.5 Let $X = Z(y^2 - x^3) \subset \mathbb{A}^2$ denote the cuspidal cubic and let $\phi : X \to \mathbb{A}^1$ denote the projection on to the x-axis. Then $\phi^* : A(\mathbb{A}^1) \to A(X)$ is given by the inclusion $k[x] \subset k[x,y]/(y^2 - x^3)$, and

$$k[x,y]/(y^2 - x^3) = k[x] \oplus k[x]y$$

as a $k[x]$-module. Hence ϕ is finite.

Note that 'most' fibers $\phi^{-1}(q)$ consists of two points (for all q in the open set $D(x) \subset \mathbb{A}^2$). On the other hand $A(X)$ has rank 2 as a $k[x]$-module. We shall see later that this is a general phenomenon. ☆

7.6 In the previous paragraph we essentially gave two definitions of a finite morphism, and we must see that they coincide, which in fact, is not entirely trivial:

PROPOSITION 7.7 *If $\phi : X \to Y$ is a map between affine varieties such that $\phi^*\colon A(Y) \to A(X)$ makes $A(X)$ a finite $A(Y)$-module, then $V = \phi^{-1}(U)$ is affine for all affine $U \subseteq Y$ and moreover $A(V)$ is a finite $A(U)$-module.*

PROOF: The claim is true for distinguished open subsets $D(f)$ of Y, since clearly $\phi^{-1}(D(f)) = D(f \circ \phi)$. But if $f_1, \dots, f_r$ are regular functions on U that generate the unit ideal in $A(U)$, the $\phi \circ f_i$'s generate the unit ideal in $\mathcal{O}_X(\phi^{-1}(U))$ as well, and $\phi^{-1}(U)$ is affine according to the criterion in Proposition 3.54 on page 61. Restrictions of generators for $A(X)$ over $A(Y)$ obviously form generators for $A(\phi^{-1}(D(f)))$ over $A(D(f))$ for each distinguished open subset $D(f)$, and thus the second claim follows from the lemma below. ❑

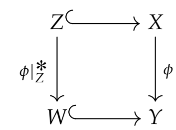
LEMMA 7.8 *Let A be a ring and M an A-module. Let $f_1, \dots, f_r$ be elements in A that generate the unit ideal, and let $m_1, .., m_t$ be elements that generate M_{f_i} for all i. Then the m_i's generated M.*

PROOF: Write $1 = \sum_i a_i f_i$. Let $\alpha\colon A^t \to M$ be the map that sends e_i to m_i. For each $x \in M$ there is an N so that all $f_i^N \cdot x \in \operatorname{im} \alpha$. Then $x = (\sum_i a_i f_i)^{rN} x$, and developing the N-power, each term contains a factor of shape f_i^N, and so $x \in \operatorname{im} \alpha$. ❑

7.9 Finite morphisms have the virtue of being closed, and hence they are surjective when they are dominating. Equally important, their fibres are finite. Moreover, as alluded to above, two affine varieties which are related by a dominant finite morphism, have the same dimension.

7.10 Here comes the basic property of finite polynomial maps. It is frequently referred to as Going–Up although Lying–Over would be the proper name in the Cohen–Seidenberg nomenclature.

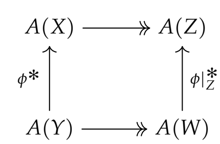
Proposition 7.11 (Lying–Over) *Let $\phi\colon X \to Y$ be a finite morphism between varieties . Then ϕ closed. If it is dominating, it is surjective.*



Proof: As Y is covered by affine opens and and inverse images of affine opens are affine opens, it suffices to prove the proposition when X and Y are affine.

We begin with proving that ϕ is surjective when it is dominating. So assume there is a y in Y not belonging to the image of ϕ. Then by Lemma 2.58 above, it holds true that $\mathfrak{m}_y A(X) = A(X)$. Now $A(X)$ being finite as an $A(Y)$-module, it follows from Nakayama's lemma that $A(X)$ is killed by an element of the shape $1 + a$ with $a \in \mathfrak{m}_y$. The assumption that ϕ be dominant ensures that ϕ^* is injective, and since $0 = (1 + a) \cdot 1 = \phi^*(1 + a)$, it follows that $a = -1$. This is absurd because a belongs to $\mathfrak{m}_y$ which is a proper ideal.

To see that ϕ is a closed map, let $Z \subseteq X$ be closed, and decompose Z into its irreducible components $Z = Z_1 \cup \ldots \cup Z_r$. Then the image $\phi(Z)$ satisfies $\phi(Z) = \phi(Z_1) \cup \ldots \cup \phi(Z_r)$, and it suffices to show that each $\phi(Z_i)$ is closed. That is, we may assume that Z is irreducible. Define W to be the closure of $\phi(Z)$, and observe that the restriction $\phi|_Z\colon Z \to W$ is a dominating and finite map (any generating set for $A(X)$ over $A(Y)$ reduces to one for $A(Z)$ over $A(W)$). Hence, by the first part of the proof, it is surjective. In other words, $\phi(Z) = W$, and $\phi(Z)$ is closed. ❑



7.12 Not only are finite maps surjective, but any closed irreducible subset of the target is dominated by a closed irreducible subset of the source

Proposition 7.13 (Going–Up) *Let $\phi\colon X \to Y$ be a finite and dominating morphism between two varieties and let $Z \subseteq Y$ be a closed and irreducible subset. Then there exists a closed and irreducible subset $W \subseteq Y$ such that $\phi(W) = Z$.*

Proof: Consider $\phi^{-1}(Z)$ which is non-empty since ϕ is surjective (proposition 7.11) and let $W_1, \ldots, W_r$ be its components. Again by Proposition 7.11 the images $\phi(W_i)$ are closed and of course, their union equals Z. Since Z is assumed to be irreducible, it follows that for at least one index i it holds that $\phi(W_i) = Z$. ❑

Lemma 7.14 *Let $\phi\colon X \to Y$ be a dominating finite morphism between varieties, and suppose that $Z \subset X$ is a proper and closed subset. Then $\phi(Z)$ is a proper subset of Y.*

PROOF: If $\phi(Z) = Y$, then pick an affine $V \subset Y$. Then $\pi^{-1}V$ is affine, and $\phi(Z \cap \phi^{-1}V) = V$, so we may reduce to the case that both X and Y are affine.

Assume that $\phi(Z) = Y$ and let $f \in I(Z)$. Since ϕ^* makes $A(X)$ a finitely generated module over $A(Y)$, it makes $A(X)$ integral over$A(Y)$, anf there is a relation

$$f^r + \phi^*(a_{r-1})f^{r-1} + \ldots + \phi^*(a_1)f + \phi^*(a_0) = 0$$

where the a_i's are elements of $A(Y)$, and we may assume that r is the least integer for which there is such a relation.

Since $f(x) = 0$ for all $x \in Z$, the relation implies that $\phi^*(a_0) = a_0 \circ \phi$ vanishes along Z. But since $\phi(Z)$ is equal to Y, the composition map $\phi|_Z^*$ is injective, and hence $a_0 = 0$. Since $A(X)$ is an integral domain, this contradicts the assumption that r was chosen minimal. Hence we conclude that $f = 0$, and $Z = X$. ❑

7.15 We do not yet know that varieties are of finite dimension, so some care must be taken to include the case of (the *a posteriori* non-exsitent) infinite dimensional varieties, and we resort to Noetherian induction.

PROPOSITION 7.16 (GOING–UP II) *Let $\phi\colon X \to Y$ be a finite and dominating morphism between varieties. Then* $\dim X = \dim Y$.

PROOF: To begin with we take any chain

$$W_0 \subset W_1 \subset \ldots \subset W_r \tag{7.1}$$

in X and push it down to Y with the help of ϕ. Each $\phi(W_i)$ is irreducible and closed in Y after Lying–Over, and Lemma 7.14 ensures that strict inclusions are preserved. Hence

$$\phi(W_0) \subset \phi(W_1) \subset \ldots \subset \phi(W_r)$$

is a chain of closed irreducible subsets of Y of length r. Taking the supremum of lengths of chains as (7.1), gives $\dim X \leqslant \dim Y$. To establish the reverse inequality, we shall lift chains in Y to chains in X by recursively climbing down[1] a given chain. Let a chain

$$Z_0 \subset Z_1 \subset \ldots \subset Z_r \tag{7.2}$$

in Y be given, and suppose we have found a chain

$$W_\nu \subset W_{\nu+1} \subset \ldots \subset W_r$$

in W with $\phi(W_i) = Z_i$. The restriction $\phi_\nu = \phi|_{W_\nu}$ is a finite map from W_ν to Z_ν and after Going–Up (Proposition 7.13) there is a closed irreducible subset of $W_{\nu-1}$ of X_ν such that $\phi_\nu(W_{\nu-1}) = Z_{\nu-1}$. In this way every chain (7.2) can be lifted to a chain of the same length, and we conclude that $\dim Y \leqslant \dim X$ ❑

[1] It may sound paradoxical that one uses Going–Up to climb down, but it comes from the transition between ideals and subvarieties reversing inclusions.

Exercises

7.1 We shall come back to a closer analysis of the fibres of finite polynomial maps, but for the moment we content ourselves with this exercise. Let $\phi\colon X \to Y$ be a finite morphism (or polynomial map). Show that all fibres of ϕ are finite. HINT: Pick a point y in Y and argue that the ring $A(X)/\mathfrak{m}_y A(X)$ is a finite dimensional vector space over k hence has only finitely many maximal ideals.

7.2 Show that the composition of two composable finite morphisms (or polynomial maps) is finite. Show that the composition of two composable dominant morphisms is dominant.

7.3 Let $\phi\colon X \to Y$ and $\psi\colon Y \to Z$ be two morphisms between affine varieties. Show that if the composition $\psi \circ \phi$ is finite (respectively dominant), then ϕ is finite (respectively dominant). Show by examples that ψ need not be finite (respectively dominant) even if $\psi \circ \phi$ is.

★

7.2 *Noether's Normalization Lemma*

The Normalization Lemma

We now turn to one of the most famous results of Emmy Noether's, her so-called Normalization Lemma. We shall stated in our context of varieties which means for algebras over an algebraically closed field. The proof however, works *mutatis mutandis* for any domain of finite type over any infinite field, and in fact, this general version will be useful for us at a later occasion. There is also a slight twist to the proof below making it valid over finite fields as well (which we shall not need).

7.17 The proof of the Normalization Lemma is an inductive argument, and the basic ingredient is the induction step as formulated in the following lemma:

LEMMA 7.18 *Let $X \subseteq \mathbb{A}^m$ be an affine variety whose fraction field $k(X)$ has transcendence degree at most $m-1$; then there is a linear projection $\pi\colon \mathbb{A}^m \to \mathbb{A}^{m-1}$ so that $\pi|_X\colon X \to \mathbb{A}^{m-1}$ is a finite morphism.*

PROOF: Let $A(X) = k[T_1, \ldots, T_m]/I(X)$ be the coordinate ring of X and denote by t_i the image of T_i in $A(X)$. Since the transcendence degree of $A(X)$ over k is less than m, the m elements $t_1, \ldots, t_m$ can not be algebraically independent and must satisfy an equation

$$f(t_1, \ldots, t_m) = 0,$$

where f is a polynomial with coefficients in k. Let d be the degree of f and let f_d be the homogeneous component of degree d. Put $u_i = t_i - a_i t_1$ for $i \geqslant 2$ where the a_i's are scalars to be chosen. This gives[2]

$$0 = f(t_1, \ldots, t_m) = f_d(1, a_2, \ldots, a_m) t_1^d + Q(u_2, \ldots, u_m)$$

where Q is a polynomial whose terms all are of degree less than d in t_1. Now, since the ground field is infinite, a generic choice of the scalars a_i implies that $f_d(1, a_2, \ldots, a_m) \neq 0$ indeed, the complement of $Z(f_d(1, t_2, \ldots, t_m))$ in $\mathbb{A}^{m-1}$ is even dense (see exercise for the case that k is merely assumed to be infinite 7.4 below). Hence the element t_1 is integral over $k[u_2, \ldots, u_m]$ and by consequence, $A(X)$ is a finite module over the algebra $k[u_2, \ldots, u_m]$. The projection $\mathbb{A}^m \to \mathbb{A}^{m-1}$ sending $(t_1, \ldots, t_m)$ to $(u_2, \ldots, u_m)$ does the trick. ❑

[2] Recall that for any polynomial $p(x)$ it holds true that $p(x+y) = p(x) + yq(x,y)$ where q is a polynomial of total degree less than the degree of f.

EXERCISE 7.4 Let k be an infinite field and $f(t_1, \ldots, t_n)$ a non-zero polynomial with coefficients from k. Show that $f(a_1, \ldots, a_n) \neq 0$ for infinitely many choices of a_i from k. HINT: Use induction on n and expand f as $f(t_1, \ldots, t_n) = \sum_i g_i(t_1, \ldots, t_{n-i}) t_n^i$. ★

7.19 By induction on m one obtains the full version of the Normalization Lemma:

THEOREM 7.20 (NOETHER'S NORMALIZATION LEMMA) *Assume that $X \subseteq \mathbb{A}^m$ is a closed subvariety and that the function field $k(X)$ is of transcendence degree n over k. Then there is a linear projection $\pi\colon \mathbb{A}^m \to \mathbb{A}^n$ such that the projection $\pi|_X\colon X \to \mathbb{A}^n$ is a finite map.*

PROOF: We keep the notation from the lemma with the coordinate ring of X being $A(X) = k[T_1, \ldots, T_m]/I(X)$ and the t_i's being the images of the T_i's in $A(X)$, and proceed by induction on m. If $m \leqslant n$, the elements $t_1, \ldots, t_m$ must be algebraically independent since they generate the field $k(X)$ over k. But any non-zero polynomial in $I(X)$ would give a dependence relation among them, so we infer that $I(X) = 0$, and hence that $X = \mathbb{A}^m$.

Suppose then that $m > n$. By Lemma 7.18 above, there is a finite projection $\phi\colon X \to \mathbb{A}^{m-1}$. The image $\phi(X)$ is closed by Proposition 7.11 on page 146 and of the same transcendence degree as X since $k(X)$ is a finite extension of $k(\phi(X))$. Applying the induction hypothesis to $\phi(X)$, we may find a finite projection $\pi\colon \phi(X) \to \mathbb{A}^n$. The composition $\pi \circ \phi$ is then a finite map $X \to \mathbb{A}^n$. ❑

7.21 An important corollary of the Normalization lemma is that the dimension of a variety coincides with the transcendence degree of its rational function field over the ground field.

THEOREM 7.22 *Let X by any variety. Then* $\dim X = \operatorname{trdeg}_k k(X)$.

In particular, the theorem states that the affine n-space $\mathbb{A}^n$ is of dimension n. Indeed, the function field $K(\mathbb{A}^n)$ of the affine space is the field of rational function $k(x_1, \ldots, x_n)$ in n variables which is of transcendence degree n over k. We may also infer that the dimension of X is finite, which is not *a priori* clear. However, the field $k(X)$ is a finitely generated extension of the ground field k and we know *a priori* that the transcendence degree $\operatorname{trdeg}_k k(X)$ is finite.

PROOF: There are two parts of the proof; the case of $\mathbb{A}^n$ and the general case, and the latter is easily reduced to the former by way of the Normalization Lemma and the Going–Up Theorem. Indeed, replacing X by some open dense and affine subset that has the same dimension as X (which we may find according to Exercise **??** on page ??), we may assume X to be affine. By the Normalization Lemma there is a finite map $X \to \mathbb{A}^n$ with $n = \operatorname{trdeg}_k k(X)$ and consequently $\dim X = \dim \mathbb{A}^n = n$ in view of the Going–Up Theorem.

The case of $\mathbb{A}^n$ is done by induction on n; obviously it holds that $\mathbb{A}^1$ is one dimensional (the ring $k[t]$ is a PID). So assume that $n > 1$ and let $Z \subset \mathbb{A}^n$ be a maximal proper and closed subvariety sitting on top of a chain of maximal length. Then $\dim Z = \dim \mathbb{A}^n - 1$ and $\operatorname{trdeg}_k k(Z) \leqslant n - 1$ because $I(Z) \neq 0$. Noether's Normalization Lemma gives us a finite and dominating morphism $Z \to \mathbb{A}^m$, where $m = \operatorname{trdeg}_k k(Z)$. By induction it holds true that $\dim \mathbb{A}^m = m$ and thus $\dim Z = \operatorname{trdeg}_k k(Z)$. This yields

$$\dim Z = \dim \mathbb{A}^n - 1 = \operatorname{trdeg}_k k(Z) \leqslant n - 1,$$

and therefore $\dim \mathbb{A}^n \leqslant n$. The other inequality is trivial; there is an obvious ascending chain of linear subspaces of length n in $\mathbb{A}^n$. ❑

7.23 As promised in Pargraph 2.44 we now can give the following:

COROLLARY 7.24 *Any dense open subsets of a variety has the same dimension as the surrounding variety.*

PROOF: The variety and the open subset have the same function field. ❑

Recall that two varieties X and Y are said to be *birational* if there are open subsets $U \subset X$ and $V \subset Y$ related by an isomorphism $U \to V$. We saw earlier that this was equivalent to saying that $k(X) \simeq k(Y)$. In particular, from the proposition above, we deduce

COROLLARY 7.25 *Birational varieties have the same dimension.*

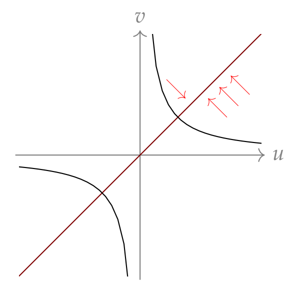
EXAMPLE 7.26 A nice illustration of the perturbation process on which the proof of the Normalization Lemma is based, is the classical hyperbola X with equation $uv = 1$ in the affine plane $\mathbb{A}^2$. The coordinate ring of X equals $k[u, v]/(uv - 1)$ which may be identified with the extension $k[u, 1/u]$ of $k[u]$, the hyperbola being the graph of the function $1/u$. The inclusion $k[u] \subseteq k[u, 1/u]$ corresponds to the projection of X onto the u-axis.

This projection map is not finite although its non-empty fibres all consist of one point; indeed, any relation $1/u^n = \sum_{i<n} f_i(u)/u^i$ with $f_i \in k[u]$ would result in the relation $1 = \sum_{i<n} f_i(u)u^{n-1}$ whose right side vanishes for $u = 0$.

However, perturbing u slightly, we obtain a subring over which $k[u, 1/u]$ is finite. For instance, the subring $k[u + 1/u]$ will do the job; indeed, $k[u, 1/u] = k[u, u + 1/u]$, and it is generated as an algebra by u over $k[u + 1/u]$. The integral dependence relation

$$u^2 - u(u + 1/u) + 1 = 0$$

holds true, and this shows that as a $k[u + 1/u]$-module $k[u, u + 1/u]$ is generated by the two elements 1 and u.



It is remarkable that almost *any* perturbation of u will work; that is, $k[u, 1/u]$ is finite over $k[au + b/u]$ as long as both the scalars a and b are non-zero. Geometrically, this corresponds to projecting X onto a line other than the two axes. ☆

Exercise 7.5 Show that $k[u, 1/u]$ is a finite module over $k[au + b/u]$ for any scalars a and b both being different from zero. ★

Maximal chains in varieties

Our second application of the Noether's Normalization Lemma is to establish that the Krull dimension of varieties behave decently in that all maximal chains have the same length. In particular they will be *catenary*. A ring is catenary if all saturated chains of prime ideals connecting two given prime ideals has the same length, and for a variety it means that all saturated chain connecting two given irreducible closed subsets have the same length.

Catenary rings
katenære ringer, fornorskningsfantomer kaller dem "kjedelige"

7.27 We begin with two lemmas to prepare the ground. The first asserts the highly expected fact that hypersurfaces in affine space $\mathbb{A}^n$ are of codimension one, a forerunner of the general Hauptidealsatz of Krull's. In particular they are maximal, closed irreducible and proper subsets.

Lemma 7.28 *The zero locus $X = V(f)$ in $\mathbb{A}^n$ of an irreducible polynomial f is of dimension $n - 1$.*

Proof: The coordinate ring $A(X)$ is given as $A(X) = k[T_1, \ldots, T_n]/(f)$ and the function field $k(X)$ is therefore generated by the t_i's; that is, we have $k(X) = k(t_1, \ldots, t_n)$ (with our usual convention in force that lower case letters denote the classes of the upper case versions), and the relation $f(t_1, \ldots, t_n) = 0$ holds among the t_i's.

At least one of the variables occurs in f, and we may as well suppose it is T_1. Thence $t_2, \ldots, t_n$ will be algebraically independent over k, and consequently we have $\operatorname{trdeg}_k k(X) = n - 1$. Indeed, any polynomial $g(T_2, \ldots, T_n)$ that satisfies $g(t_2, \ldots, t_n) = 0$, is a multiple of f and must therefore depend on T_1. ❑

Lemma 7.29 *Let X be a variety and Z a maximal proper, closed and irreducible subset. Then* $\dim Z = \dim X - 1$

Proof: By Noether's Normalization Lemma there is a finite surjective morphism $\pi\colon X \to \mathbb{A}^n$ with $n = \dim X$. The image $\pi(Z)$ is irreducible and closed (by Lying–Over), and we content that it is a maximal proper subset of this kind. Indeed, if the closed and irreducible subset W were lying strictly between $\pi(Z)$ and $\mathbb{A}^n$, our set Z would have been contained in one of the components of $\pi^{-1}(W)$, say W_0. By Lying–Over there is no inclusion relation between closed irreducible sets with the same image under a finite map; hence W_0 lies strictly between Z and X, which contradicts the hypothesis that Z is maximal. Hence $\pi(Z)$ is maximal in $\mathbb{A}^n$ and therefore of dimension $n-1$ by Lemma 7.28 above. We finish the proof by the Lying–Over Theorem which asserts that $\dim Z = \dim \pi(Z)$ since the restriction $\pi|_Z$ is finite. ❑

7.30 With the two previous lemmas up our sleeve the main theorem of the section is easy to prove.

> **Theorem 7.31** *All maximal chains of closed irreducible subvarieties in a variety are of the same length.*

Proof: The proof goes by induction on the dimension of X (which we have shown is finite) and the case $\dim X = 0$ is trivial. Let $Z_r \subset X$ be the largest member of a maximal chain

$$Z_0 \subset Z_1 \subset \ldots \subset Z_{r-1} \subset Z_r \subset X$$

of length r in X. The top member Z_r is a maximal proper closed subvariety of X, and therefore $\dim Z = \dim X - 1$ after Lemma 7.29 above. The induction hypothesis then takes effect and implies that $\dim Z_r = r - 1$, and we can conclude that $\dim X = r$. ❑

7.32 Localization of catenary rings are catenary so that rings of essentially finite type over k are catenary. In particular, such domains that are local will have maximal chains all of the same length. However, if the domain is not local, this not true any more. Problem 2.8 on page 36 described a semi-local domains with two maximal ideals, one of height one and the other of height two.

Codimension

If $Y \subset X$ is a non-empty irreducible closed subset of X, we define the *codimension* of Y in X, denoted $\operatorname{codim} Y$, as the supremum over the length r of any chain starting with Y: *codimension* *kodimensjon*

$$Y = X_0 \subsetneq X_1 \subseteq \cdots \subsetneq X_n \subset X$$

Proposition 7.33 *Let X be a variety and let $Y \subset X$ be closed subvariety.*

i) If X is affine, the codimension of Y in X equals the height of $I(Y)$ in $A(X)$:

$$\operatorname{codim} Y = htI(Y)$$

ii) $\dim X = \dim Y + \operatorname{codim} Y$

Proof: (i) follows by the Nullstellensatz: chains in X containing Y correspond exactly to chains of prime ideals in $A(X)$ containing $I(Y)$.

(ii) Both sides of the equation can be computed by going to an open subset of X, so we may assume X is affine. In that case, all maximal chains of prime ideals in $A(X)$ have the same length, so that a maximal chain containing the prime ideal $I(Y)$ also has length $\dim X$. ❑

The dimension of a product

The dimension of a product is what it should be, the sum of the dimensions of the factors. The Normalization Lemma gives an easy proof of this. It hinges on the fact that the product of two finite maps is finite, and the Normalization Lemma reduces the proof to the case of two affine spaces in which is trivial.

Proposition 7.34 *Let X and Y be two varieties. Then* $\dim X \times Y = \dim X + \dim Y$.

Lemma 7.35 *Let X, Y, Z and W be affine varieties. Let $\phi\colon X \to Y$ and $\psi\colon Z \to W$ be two finite morphisms. Then the morphism $\phi \times \psi\colon X \times Z \to Y \times W$ is finite.*

Proof: We first establish the lemma in the special case when $W = Z$ and $\psi = \mathrm{id}_Z$. In that case the map $(\phi \times \mathrm{id}_Z)^*\colon A(Y)\otimes A(Z) \to A(X)\otimes A(Z)$ is just $\phi^*\otimes \mathrm{id}_{A(Z)}$. If $a_1, \ldots, a_r$ are elements in $A(X)$ that generates $A(X)$ as an $A(Y)$-module, the elements $a_i\otimes 1$ generates $A(X)\otimes A(Z)$ as a module over $A(Y)\otimes A(Z)$, and we are through.

One reduces the general case to this special case by observing that $\phi \times \psi$ is equal to the composition

$$X \times Z \xrightarrow{\phi\times \mathrm{id}_Z} Y \times Z \xrightarrow{\mathrm{id}_Y\times\psi} Y \times W,$$

and using that the composition of two finite maps is finite. ❑

Proof of Proposition 7.34: By replacing X by a dense open affine subset U and Y by a dense open affine subset V, we may assume that X and Y affine; indeed, $U \times V$ is dense in $X \times Y$, and dense open subsets have the same dimension as the surrounding variety (Corollary 7.24 on page 150).

So assume that X and Y are affine. According to the Normalization Lemma there are finite and surjective maps $\phi\colon X \to \mathbb{A}^n$ and $\psi\colon Y \to \mathbb{A}^m$ with $n = \dim X$ and $m = \dim Y$. Then $\phi \times \psi\colon X \times Y \to \mathbb{A}^n \times \mathbb{A}^m \simeq \mathbb{A}^{n+m}$ is finite by lemma 7.35 above, and it is clearly surjective, hence $\dim X \times Y = n + m$ after Going–Up (Proposition 7.16 on page 147). ❑

7.3 *Krull's Principal ideal theorem*

Krull's Principal Ideal Theorem is another great German theorem, whose native name is *Krull's Hauptidealsatz*, but unlike the Nullstellensatz, it is mostly referred to by its English name in the Anglo-Saxon part of the world. The simplest version concerns the intersection of a hypersurface with a variety X in $\mathbb{A}^m$, and confirms the intuitive belief that the hypersurface cuts out a subvariety in X of dimension one less than $\dim X$. This statement must be taken with a grain of salt since the intersection could be empty, and of course, the variety X could be entirely contained in the hypersurface in which case the intersection equals X, and the dimension does not drop. If X is not irreducible, the situation is somehow more complicated. The different components of X can be of different dimensions and they may or may not meet the hypersurface.

A slight, but obvious generalization, is to consider a regular function f on a variety X and ask for the codimension of the zero set $Z(f)$, an of course, and the Hauptidealsatz still holds true. Another direction of generalisation is what Krull himself called the *Idealkättensatz*; that is, the case of several regular functions, say $f_1, \ldots, f_r$. As the dimension of the zero set can drop at most by one for each new f_i one introduces, it holds true that $Z(f_1, \ldots, f_r)$ is of codimension at most r.

The Principal Ideal theorem applies to general Noetherian ring which are enormously more delicate beings than the ones we meet in the world of varieties. We state it in its generality, but shall only need a more basic version for finitely generated.

Theorem 7.36 (The Principal Ideal Theorem) *Let A be Noetherian ring and let $f \in A$ be a non-zero element which is not a unit. Then the height of a minimal prime of the principal ideal (f) is a most one.*

Theorem 7.37 (Geometric Principal Ideal Theorem) *Assume that X is a variety and f a regular function on X that does not vanish identically. If $Z(f)$ is not empty, it holds true for every component Z of $Z(f)$ that* $\dim Z = \dim X - 1$.

Proof: We apply the above theorem to the case $A = A(X)$, which has no zero-divisors. If Z is a component of $Z(f)$, then $Z = Z(\mathfrak{p})$ for some prime ideal $\mathfrak{p}$ which contains f. Note that a prime ideal of height 0 must consist entirely of zero-divisors; thus $\mathfrak{p}$ must have height one. Hence $Z(f)$ has codimension 1. ❑

$$\begin{array}{ccc} B & \subset & L \\ \cup & & \cup \\ A & \subset & K \end{array}$$

7.38 The theorem asserts that irreducible hypersurfaces are of codimension one, but be aware that the converse is not true in general. There are plenty of varieties having irreducible subvarieties of codimension one that are not hypersurfaces. An example close at hand is the cone $X = Z(xy - zw)$ in $\mathbb{A}^4$ over a quadric in $\mathbb{P}^3$. In the coordinate ring $A = k[x,y,z,w]/(xy - zw)$ of X one has the primary decomposition

$$(x) = (x,z) \cap (x,w),$$

so that $Z(x) \cap X$ is the union of the two planes $Z(x,w)$ and $Z(x,z)$. One may prove that any hypersurface containing either of those planes can not be irreducible (see Exercise 7.6 below), the point being that A is not a UFD. From Kaplansky's criterion that a domain is a UFD if and only if "every prime (ideal) contains a prime (element)" one infers that in any variety whose coordinate ring is not a UFD, one may find irreducible subsets of codimension one that are not hypersurfaces. Indeed, in the coordinate ring there will be prime ideals of height one that are not principal.

7.39 There is no statement as clear and uniform as the one in Theorem 7.37 valid for genneral closed algebraic sets X (which may be reducible). It is not very difficult to exhibit examples of situations where the codimension of $Z(f)$ in X is equal to any prescribed number.

Exercises

7.6 Refering to the staging in Paragraph 7.38 above, let $X \subseteq \mathbb{A}^4$ be $X = Z(xy - zw)$. Show that for any hypersurface $Z(f)$ in $\mathbb{A}^4$ containing the plane $Z(x,w)$, the intersection $X \cap Z(f)$ is reducible. Clearly $Z(x,w)$ is a component of $Z(f) \cap X$, the point is to show that there are others.

7.7 Let $X \subseteq \mathbb{P}^n$ be a closed subvariety and assume that the cone $C(X)$ is a UFD. Show that any $Z \subseteq X$ of codimension one is of the shape $Z(f) \cap X$ for some hypersurface $Z(f)$ in $\mathbb{P}^n$.

7.8 Give examples of a closed algebraic set and a regular function f on a variety X such that the codimension of $Z(f)$ in X is equal to any prescribed number. HINT: Use disjoint unions.

★

The case of several functions vanishing

7.40 The Principal Ideal Theorem is about the dimension of loci where one constraint is imposed; that is, about intersections of a variety X with one hypersurface. However, it generalizes to intersections with a sequence of hypersurfaces, or in a slightly more general staging, to loci where several regular functions vanish. Since imposing that each one of the functions vanishes, increases the codimension with at most one, induction on the number of functions easily gives the following:

Theorem 7.41 *Suppose that X is a variety and that $f_1, \ldots, f_r$ are regular functions on X. Then every component Z of the zero locus $Z(f_1, \ldots, f_r)$ is of codimension at most r in X; that is* $\dim Z \geqslant \dim X - r$.

Proof: The proof goes by induction on r. Let W be a component of the locus $Z(f_1, \ldots, f_{r-1})$ that contains Z. By induction W is of codimension at most $r-1$ in X; that is, $\dim W \geqslant \dim X - r + 1$. Moreover, Z must be a component of of $W \cap Z(f_r)$, and therefore either f_r vanishes identically on W or $\dim W = \dim Z - 1$ by the Principal Ideal Theorem (Theorem 7.37 above). In the latter case obviously $\dim Z \geqslant \dim X - r$, and in the former, we find $Z = W$ and $\dim Z = \dim X - r + 1 \geqslant \dim X - r$. ❑

7.4 System of parameters and fibres of morphisms

The first two applications of the Hauptidealsatz we shall give are to the existence of so-called system of parameters, and to an estimate for the dimensions of fibres of dominant morphisms.

System of parameters

7.42 In commutative algebra one has the notion of a *system of parameters* in a Noetherian local ring A. If the Krull dimension of A is n and the maximal ideal $\mathfrak{m}$, such a system is a sequence of elements $f_1, \ldots, f_n$ of n elements in $\mathfrak{m}$ that generate an $\mathfrak{m}$-primary ideal; or expressed in symbols, such that $\sqrt{(f_1, \ldots, f_n)} = \mathfrak{m}$. An equivalent way of phrasing this, is to say that the quotient ring $A/(f_1, \ldots, f_n)$ is Artinian. Such system always exists due to the Hauptidealsatz. We are about to prove this for local rings of varieties.

System of parameters
parametersystemer

7.43 Translating the algebra into geometry we let X be an affine variety of dimension n and A the local ring $\mathcal{O}_{X,x}$ of X at a point $x \in X$. The elements $f_1, \ldots, f_n$ are regular functions on X that vanish at x, and requiring the f_i's to constitute a system of parameters in $\mathcal{O}_{X,x}$ is to ask that $\sqrt{(f_1, \ldots, f_n)\mathcal{O}_{X,x}} = \mathfrak{m}_x\mathcal{O}_{X,x}$.

By the Nullstellensatz this is equivalent to asking that in a neighbourhood of x the only common zero of the f_i's is the point x. The zero locus $Z(f_1, \ldots, f_n)$ thus has an irredundant decomposition into irreducibles shaped as $Z(f_1,\ldots,f_n) = \{x\} \cup Z_1 \cup \ldots \cup Z_r$, or in other words, x is an isolated point of the zero set $Z(f_1,\ldots,f_n)$.

7.44 A careful application of Krull's Principal Ideal Theorem yields that for affine varieties there are system of parameters near every point.

Proposition 7.45 *Let X be an affine variety of dimension n and $x \in X$ a point. Then there exists regular functions $f_1,\ldots,f_n$ on X such that x is an isolated point in $Z(f_1,\ldots,f_n)$.*

Proof: We will recursively construct a sequence of regular functions $f_1,\ldots,f_n$ on X all vanishing at x so that for ν with $1 \leqslant \nu \leqslant n$, every component of $X_\nu = Z(f_1,\ldots,f_\nu)$ that contains x, is of codimension ν . Clearly this suffices to establish the theorem; indeed, when $\nu = n$ that statement reads: All components of $Z(f_1,\ldots,f_n)$ containing x are of dimensions zero; hence there can only be one, and it must be equal to $\{x\}$.

Assume then that the ν first functions $f_1,\ldots,f_\nu$ are found. Let $Z_1,\ldots,Z_r$ be the components of $X_\nu = Z(f_1,\ldots,f_\nu)$ that contain x and let $\mathfrak{p}_1,\ldots,\mathfrak{p}_r$ denote the corresponding prime ideals in the coordinate ring $A(X)$. When $\nu < n$, every one of the Z_i's is of dimension at least one, and the $\mathfrak{p}_i$'s are strictly contained in $\mathfrak{m}_x$. Citing the prime avoidance lemma we infer that $\mathfrak{p}_1 \cup \ldots \cup \mathfrak{p}_\nu \subsetneq \mathfrak{m}$, and hence we may find functions $f_{\nu+1}$ vanishing at x, but which do not vanish identically along any of the Z_i's. It follows from the Principal Ideal Theorem that all components of $Z(f_{\nu+1}) \cap Z_i$ are of codimension one in Z_i; that is, they are all of codimension $\nu+1$ in X. ❑

Example 7.46 Consider the quadric cone $Q = Z(xy - z^2) \subset \mathbb{A}^3$. This has dimension 2 (by Krull), and x, y form a system of parameters. Indeed, we have $z^2 \in (x,y)$, so $\sqrt{(x,y)} = (x,y,z)$.

On the other hand x, z does not form a system of parameters: the ideal $\mathfrak{p} = (x,z)$ is prime in $A = k[x,y,z]/(xy - z^2)$ (since $A/\mathfrak{p} \simeq k[y]$), but being non-maximal, it is strictly contained in $\mathfrak{m} = (x,y,z)$. ☆

Example 7.47 At any singular point it is (almost by definition) impossible to find a system of parameters that generates the maximal ideal. We illustrate this by the example of a hypersurface. So let $X = Z(f)$ be a hypersurface in the affine space $\mathbb{A}^n$ and chose coordinates $x_1,\ldots,x_n$ in $\mathbb{A}^n$ so that the point p we are interested is the origin. That f is singular at p means that f vanishes at least to the second order there; in other words, f lies in the square $\mathfrak{n}^2 = (x_1,\ldots,x_n)^2$ of the ideal $\mathfrak{n} = (x_1,\ldots,x_n)$.

The maximal ideal $\mathfrak{m}$ of the local ring $\mathcal{O}_{X,p}$ equals the quotient $\mathfrak{m} = (x_1,\ldots,x_n)/(f) = \mathfrak{n}/(f)$, and in general there is an exact sequence

$$0 \longrightarrow (\mathfrak{n}^2 + (f))/\mathfrak{n}^2 \longrightarrow \mathfrak{n}/\mathfrak{n}^2 \longrightarrow \mathfrak{m}/\mathfrak{m}^2 \longrightarrow 0$$

which comes from the elementary isomorphism theorems for quotient modules. Now, when f belongs to the square $\mathfrak{n}^2$, the kernel to the left is zero. Hence $\mathfrak{n}/\mathfrak{n}^2 \simeq \mathfrak{m}/\mathfrak{m}^2$, and thus $\mathfrak{m}/\mathfrak{m}^2$ has dimension n.

If $f_1, \ldots, f_r$ are elements in $\mathcal{O}_{X,p}$ that generate the maximal ideal $\mathfrak{m}$, the classes of f_i's modulo $\mathfrak{m}^2$ generate the k-vector space $\mathfrak{m}/\mathfrak{m}^2$. But since f vanishes to the second order, $\dim_k \mathfrak{m}/\mathfrak{m}^2 = n$, and $\mathfrak{m}$ requires n generators, but $\dim X = n-1$, so one more generators than the dimension is always need. ☆

EXERCISE 7.9 Assume that $f_1, \ldots, f_n$ is a system of parameters at the point x of the affine variety X. Prove that for any $1 \leqslant \nu \leqslant n$ every component Z of $Z(f_1, \ldots, f_\nu)$ has codimension ν . Prove that any sequence of regular functions $f_1, \ldots, f_\nu$ that vanish at x and with all components Z of $Z(f_1, \ldots, f_v)$ satisfying $\operatorname{codim} Z = \nu$, can be completed to a system of parameters at x. ★

EXERCISE 7.10 Show that $x_0 - x_1, x_2, x_3$ is a system of parameters at the origin for $Z(x_0x_1 - x_2x_3)$. ★

EXERCISE 7.11 Gerenarlize the previous exercise in the following direction. Let $X \subseteq \mathbb{P}^n$ be a subvariety of dimension d. Assume that $L_1, \ldots, L_{d+1}$ are linear forms whose intersection is a subspace of codimension $d+1$ that does not meet X. Show that the L_i's form a system of parameters for the cone $C(X)$ at the origin. Finally, prove that such linear forms always can be found. ★

EXERCISE 7.12 Assume that X is a (irreducible) variety and that Y is a curve. Show that all components of all fibres of a dominant morphism $\phi\colon X \to Y$ are of codimension one in X. ★

Fibres of morphisms

7.48 According to Theorem 7.41 above, the fibre over the origin of a dominant map $\phi\colon X \to \mathbb{A}^r$ with the origin lying in the image, has all its components of dimension at least equal to $\dim X - r$. Indeed, the map has regular functions $f_1, \ldots, f_r$ as components, the fibre over the origin is just $Z(f_1, \ldots, f_r)$ and then 7.41 gives $\dim X - \dim Z \leqslant r$.

This observation can be generalised to fibres over any point of any variety, based on the fact that affine varieties unconditionally possess systems of parameters at all points, and yields the estimate in Proposition 7.49 below. Strict inequality commonly occur, but merely for special points belonging to a "small" subset of the target Y. In xxx we shall give a more precise statement.

Proposition 7.49 *Let $\phi\colon X \to Y$ be a dominant morphism of varieties. For every point x in Y belonging to the image of ϕ, and for every component Z of the fibre $\phi^{-1}(x)$ it holds true that*

$$\dim Z \geqslant \dim X - \dim Y.$$

Proof: Replacing Y by a neighbourhood of x need is, we may assume that Y is affine. Let $r = \dim Y$. In Proposition 7.45 on page 157 we showed that affine varieties have a systems of parameters at every one of their points; hence there are regular functions $f_1, \ldots, f_r$ on Y such that x is isolated in $Z(f_1, \ldots, f_r)$. By further shrinking Y, we may assume that $\{x\} = Z(f_1, \ldots, f_r)$. The fibre is then described as $\phi^{-1}(x) = Z(f_1 \circ \phi, \ldots, f_r \circ \phi)$, and by Krull's Principal Ideal Theorem every one of its components have a codimension at most equal to r; that is,

$$\dim X - \dim Z \leqslant r = \dim Y,$$

and this gives the inequality in the proposition. ❑

Exercise 7.13 Let $a \neq 0$ be a scalar and let $f(x,y,z) = x^2 + y^2 - z^2$ and $g(x,y,z) = ax^2 + a^{-1}y^2 - z^2$ be the equations of two conics in $\mathbb{P}^2$. Choose coordinates $(u : v)$ on $\mathbb{P}^1$ and consider the subvariety $X \subseteq \mathbb{P}^1 \times \mathbb{P}^2$ given as $X = Z(uf - vg)$. Let $\phi\colon X \to \mathbb{P}^2$ be the rstriction to X of the projection $\mathbb{P}^1 \times \mathbb{P}^2 \to \mathbb{P}^2$. Describe all fibres of ϕ. ★

Exercise 7.14 *(The resolution of a determinantal variety.)* Let $M(x) = (x_{ij})$ be a generic 3×2-matrix. This means that the x_{ij}'s are variables so that M has coefficients in the polynomial ring $k[x_{11}, \ldots, x_{23}]$. The matrix $M(x)$ is shaped like

$$M(x) = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}.$$

When we endow the space of matrices $\mathbb{A}^6 \simeq \hom[k]k^2k^3$ with the coordinates x_{ij} one finds the determinantal variety $W \subseteq \mathbb{A}^6$—the locus of matrices of rank less than one; that is, W defined by the vanishing of the three 2×2-minors of M. See Example **??** where we proved that W is a subvariety.

Now introduce a copy of $\mathbb{P}^1$ with homogeneous coordinates $v = (t_1 : t_2)$. Inside the product $\mathbb{A}^6 \times \mathbb{P}^1$ consider the subvariety $\widetilde{W}$ of pairs $(M, [v])$ such that $M \cdot v = 0$; that is, it is the locus where $t_1 x_{1j} - t_2 x_{2j} = 0$ for $j = 1, 2, 3$.

a) Show that all fibres of the projection $p_2\colon \widetilde{W} \to \mathbb{P}^1$ are linear hyperplanes in $\mathbb{P}^5$.

b) For $i = 1, 2$ let $U_i = p_2^{-1}(D_+(t_i))$. Exhibit isomorphisms $U_i \simeq \mathbb{A}^3 \times D_+(t_i)$ compatible with the projection. This shows again that $\widetilde{W}$ is irreducible of dimension 4.

c) Show that the projection $p_1 \colon \widetilde{W} \to W$ is birational and describe all its fibres.

★

EXERCISE 7.15 Let E and F be two vector spaces of dimension m and n respectively with $n \leqslant m$. The space $\mathbb{M} = \hom[k]FE$ of linear maps is a vector space of dimension nm. The choice of ´ of bases in E and F induces coordinates x_{ij} on $\mathbb{M}$ so that

it may be identified with the affine space $\mathbb{A}^{nm}$ and elements M may be concidred to be $m \times n$-matrices. Use coordinates x_{ij} on $\mathbb{A}^{nm}$ such that $M(x) = (x_{ij})$. ★

7.5 *Applications to intersections*

7.50 The Principal Ideal Theorem has some important consequences for the intersections of subvarieties both in the affine spaces $\mathbb{A}^n$ and in the projective spaces $\mathbb{P}^n$. It allows upper estimates for the dimension of an intersection of two closed subvarieties expressed in terms of their dimensions.

A head on application of the Principal Ideal Theorem is futile since varieties in general are not complete intersections, but require several more equations than their codimensions indicate, however there is a a fabulous trick called the "Reduction to the diagonal" that paves the way.

7.51 In the projective case, it ensures the most striking result that the intersection is non-empty once a natural condition on the dimensions of the two is fulfilled, which may be concidered an exsistence theorem for simultaneous solutions of systems of homogeneous equations.

The intersection of two closed subvarieties of $\mathbb{P}^n$ will be non-empty once their dimensions comply to the following very natural condition. If X and Y designate the two subvarieties, then $X \cap Y \neq \emptyset$ once

$$\operatorname{codim} X + \operatorname{codim} Y \leqslant n. \tag{7.3}$$

One can even say more, for every irreducible component Z of the intersection $X \cap Y$, the following inequality holds

$$\operatorname{codim} Z \leqslant \operatorname{codim} X + \operatorname{codim} Y. \tag{7.4}$$

7.52 For intersections of subvarities of the affine space $\mathbb{A}^n$ a similar inequality holds true, but under the assumption that the intersection is non-empty; so in that case, no common point is garanteed.

EXERCISE 7.16 Give examples of projective varieties X and Y not satisfying he inequality (7.3) and having an empty intersection. HINT: Take two linear subvarieties $\mathbb{P}(V)$ and $\mathbb{P}(W)$ of $\mathbb{P}^n$ with $\dim \mathbb{P}(V) + \dim \mathbb{P}(W) < n$ (*e.g.* two skew lines in $\mathbb{P}^3$). ★

EXERCISE 7.17 Give examples of two closed subvarieties in affine space $\mathbb{A}^n$ satisfying 7.3 but having an empty intersection. ★

The affine case and reduction to the diagonal

7.53 The trick named "Reduction to the diagonal" is based the following observation. Let X and Y be two subvarieties of $\mathbb{A}^n$. The product $X \times Y$ lies of course as a closed subvariety of the affine space $\mathbb{A}^n \times \mathbb{A}^n = \mathbb{A}^{2n}$, and clearly the subset $X \cap Y$ is equal to the intersection $\Delta \cap X \times Y$, where Δ is the diagonal in $\mathbb{A}^n \times \mathbb{A}^n$. Moreover, it is not difficult to check that the two closed algebraic sets are isomorphic; either use their respective defining universal properties or resort to considering the defining ideals.

The salient point is that the diagonal is cut out by a set of very simple equations. If the coordinates on corresponding to the left factor in $\mathbb{A}^n \times \mathbb{A}^n = \mathbb{A}^{2n}$ are $\{x\}_i$ and those of the right factor $\{y_i\}$ the diagonal is given by the vanishing of the n functions $x_i - y_i$. Hence we can conclude by Krull's Principal Ideal Theorem that any (non-empty) component Z of $X \cap Y$ satisfies $\dim Z \geqslant \dim X \times Y - n$ but $\dim X \times Y = \dim X + \dim Y$ and we find

$$\dim Z \geqslant \dim X + \dim Y - n.$$

Summing up we formuate the result as a lemma

PROPOSITION 7.54 *Let X and Y be two subvarieties of $\mathbb{A}^n$ then any (non-empty) component of the intersection $X \cap Y$ satifies*

$$\operatorname{codim} Z \leqslant \operatorname{codim} X + \operatorname{codim} Y.$$

Of course, it might very well happen that $X \cap Y$ is empty, even for hypersurfaces. As well, the strict inequality might hold; for example it could happen that $X = Y$!

EXAMPLE 7.55 An inequality as in 7.54 does not hold for subvarieties of general varieties. For example, the two planes $Z_1 = Z(x, y)$ and $Z_2 = Z(z, w)$ in $\mathbb{A}^4$ intersect only in the origin, and they are both contained in the quadratic cone $X = Z(xz - yw)$ which is three-dimesional. Considered as subvarieties of X the two planes are of codimension one, but their intersection just being the origin, is of codimension three. ☆

Exercises

7.18 Let $X = Z(\mathfrak{p})$ and $Y = Z(\mathfrak{q})$ be two closed subsets of the algebraic set W and let ι_X and ι_Y denote the inclusion maps. Show that the intersection $X \cap Y = Z(\mathfrak{p} + \mathfrak{q})$ is characterised by the universal property that a pair of polynomial maps $\phi_X \colon Z \to W$ and $\phi_Y \colon Z \to W$ factors through $X \cap Y$ if and only if $\iota_X \circ \phi_X = \iota_Y \circ \phi$.

7.19 Give a categorical proof of the isomorphism between $X \cap Y$ and $\Delta \cap X \times X$; that is, a proof only relying on universal properties (and hence is valid in any category where the involved players exist).

7.20 Give a more mundane proof the isomorphism between $X \cap Y$ and $\Delta \cap X \times X$ using the ideals of the involved varieties.

7.21 Find examples of two irreducible quadratic curves in $\mathbb{A}^2$ with empty intersection. Do the same for two cubic curves.

★

The projective case

7.56 The proof of the intersection theorem for projective space applies the affine version to the affine cones over the involved varieties. We therefore begin with a few observations about them. The natural equality $C(X \cap Y) = C(X) \cap C(Y)$ is obvious, and if Z is a component of the intersection $X \cap Y$, the cone $C(Z)$ will be a component of $C(X \cap Y)$. Passing to cones increases the dimensions by one; that is, for any variety X it holds that $\dim C(X) = \dim X + 1$. Then of course, it holds true that $\operatorname{codim}_{\mathbb{P}^n} X = \operatorname{codim}_{\mathbb{A}^{n+1}} C(X)$; that is, the codimenion of X in $\mathbb{P}^n$ is the same as the codimension of its cone in $\mathbb{A}^{n+1}$. And thirdly, the most salient point is that intersection of cones always is non-empty; they meet at least in the origin.

7.57 The following theorem is one of the cornerstones in projective geometry. Whether two varieties intersect or not is as much a question of their size as of their relative position: If they are "large enough", they intersect.

Proposition 7.58 *Let X and Y be two projective varieties in the projective space $\mathbb{P}^n$. Assume that $\dim Y + \dim X \geqslant n$. Then the intersection $X \cap Y$ is non-empty, and any component Z of $X \cap Y$ satisfies*

$$\operatorname{codim} Z \leqslant \operatorname{codim} X + \operatorname{codim} Y.$$

Proof: Firstly, if $\dim X + \dim Y \geqslant n$ then $\dim C(X) + \dim C(Y) \geqslant n + 2$ and, as already noticed, the salient point is that the intersection $C(X) \cap C(Y)$ is always non-empty: The two cones both contain the origin! Moreover, the dimension of any component W of $C(X) \cap C(Y)$ satisfies $\dim W \geqslant \dim C(X) + \dim C(Y) - n - 1 = \dim X + \dim Y - n + 1 \geqslant 1$, and one deduces that the intersection $C(X) \cap C(Y)$ is not reduced to the origin, and hence is the cone over a non-empty subset in $\mathbb{P}^n$.

Since the cone over over a projective variety and the variety itself have the same codimension, respectively in $\mathbb{P}^n$ and $\mathbb{A}^{n+1}$, we deduce directly from Proposition 7.54 that

$$\operatorname{codim} Z \leqslant \operatorname{codim} X + \operatorname{codim} Y.$$

❑

Exercises

7.22 Given an example of a projective variety W of a given arbitrary dimension and two subvarieties X and Y of W with empty intersection, but which satisfy

$$\dim X + \dim Y = \dim W.$$

7.23 *(Join of two varities.)* There is projective analogue of the reduction to the diagonal trick. One can always consider the product inside $\mathbb{P}^n \times \mathbb{P}^n$ and subsequently foolow up with Segre embedding. However, the diagonal is not complete intersection in $\mathbb{P}^{n^2+2n}$. There is much better way with the product of the two projective varieties X and Y in $\mathbb{P}^n$ being replaced with their so-called *join*.

Imagin X and Y living in two different copies of the projective space $\mathbb{P}^n$, one endowed with homogeneous coordinates $(x_0 : \ldots : x_n)$ and the other one with $(y_0 : \ldots : y_n)$. Their ideals are respectively $\mathfrak{p} = (f_1, \ldots, f_r)$ and $\mathfrak{q} = (g_1, \ldots, g_s)$ where the f_i's are polynomials in the x's and the g_j's in the y's.

Consider the projective space $\mathbb{P}^{2n+1}$ equipped with homogeneous coordinates $(x_0 : \ldots : x_n : y_0 : \ldots : y_n)$, and let $\mathfrak{a}$ be the ideal in $k[x_0, \ldots, x_n, y_0, \ldots, y_m]$ generated by the f_j's and the g_j's . The *join* of X and Y is the variety $X * Y = Z_+(\mathfrak{a}) \subseteq \mathbb{P}^{2n+1}$.

The join of two varieties
***** av to variteter*

There are two (in this context) natural linear embeddings of $\mathbb{P}^n$ into $\mathbb{P}^{2n+1}$; we obtain one by requiering all the y_i's to vanish, and the other by demanding the other set of coordinates to vanish; that is, imposing that $x_i = 0$ for all i. This leads to natural embeddings of X and Y into $\mathbb{P}^{2n+1}$.

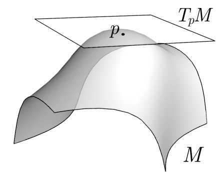
a) Prove that the cone $C(X * Y)$ in $\mathbb{A}^{n+m+2}$ over $X * Y$ equals the product $C(X) \times C(Y)$ of the cones over X and Y.
b) Prove that $\dim X * Y = \dim X + \dim Y + 1$.
c) Prove that the join $X * Y$ is the union of all lines connecting points in X to points in Y. (This explains the name *join*.)
d) Let $\Delta = Z_+(x_0 - y_0, \ldots, x_n - y_n)$. Show that Δ is a linear subspace of $\mathbb{P}^{2n+1}$ of dimension n. Prove that there is a natural isomorphism $\iota \colon \mathbb{P}^n \to \Delta$ such that $X * Y \cap \Delta$ corresponds to $X \cap Y$.
e) Describe the image of the diagonal under the Segre embedding $\mathbb{P}^n \times \mathbb{P}^n \to \mathbb{P}^{n^2+2n}$

Chapter 8

Non-singular varieties

Topics in Chapter 8: *Tangent spaces; singular points; the jacobian criterion*

In differential geometry, a manifold is a space M which is locally diffeomorphic to an open set in some $\mathbb{R}^n$. If $p \in M$ is a point of M, we can attach its *tangent space* of T_pM, which is a real vector space of dimension n. If M is embedded as a submanifold in some $\mathbb{R}^N$, we can think of tangent vectors v as vectors in the ambient space $\mathbb{R}^N$ that 'stick out' from p, and naively define T_pM as the vector space of vectors that tangentially pass through p. However, we do not like to think of manifolds as embedded in some $\mathbb{R}^N$, and then giving a precise definition of T_pM becomes a little bit subtle.



8.1 *Tangent spaces*

8.1 Let us first consider the most basic case, where $X = Z(f) \subset \mathbb{A}^n$ is a hypersurface in affine space and $p = (a_1, \ldots, a_n)$ is a point on X. Let L be a line through p, say parameterized as $(a_1 + tb_1, \ldots, a_n + tb_n)$ for $t \in k$. The set of intersection points $X \cap L$ is determined by the equation

$$g(t) = f(a_1 + tb_1, \ldots, a_n + tb_n) = 0.$$

Viewing $g(t)$ as a polynomial in $k[t]$, we may write it as $g(t) = t^e h(t)$ where $e \geqslant 0$ and h is a polynomial so that $h(0) \neq 0$. The integer e is called the *intersection multiplicity* of X and L at p. Note that $e > 0$ if and only if $p \in X \cap L$. (If $g(t)$ is the zero polynomial; *i.e.* if $L \subseteq X$, we define the intersection multiplicity to be $+\infty$.)

Intersection multiplicity
snittmultiplisitet

This anticipates the general notion of intersection multiplicities which will be introduced in Chapter 11 about Bézout's Theorem.

A line L is *tangent* to X at P if the intersection multiplicity of L at P is at least two. Note that this happens if and only if $g(t)$ has a multiple root at 0. This in turn happens if and only if $g'(0) = 0$; or in other words,

$$\sum_{i=1}^{n} b_i \frac{\partial f}{\partial x_i}(p) = 0.$$

It follows that L is tangent to X at p if and only if the vector $(b_1, \ldots, b_n)$ belongs to the *tangent space*

$$T_pX = \{\,(v_1, \ldots, v_n) \in k^n \mid \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot v_i = 0\,\}$$

8.2 More generally, if $X = Z(\mathfrak{a}) \subset \mathbb{A}^n$ is an affine variety, we define the tangent space of X at p as the subspace of vectors $v \in k^n$ satisfying the linear equations

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot v_i = 0 \quad \text{for all } f \in I.$$

The Jacobian matrix
Jacobi-matrisen

If $\mathfrak{a} = (f_1, \ldots, f_r)$, the tangent space is the null space of the *Jacobian matrix* which is defined by

$$J(f_\bullet) = \big(\frac{\partial f_i}{\partial x_j}(p)\big),$$

where $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant n$. Note that this is the Jacobian matrix as in calculus of several variables; the matrix with the gradient vectors ∇f_i as row vectors. Of course, the null space T_pX is naturally a k-vector space, and its dimension is given by

$$\dim T_pX = n - \operatorname{rank} J(f_\bullet). \tag{8.1}$$

An important consequence is that $\dim T_pX$ is an upper semicontinuous function of p, *i.e.* the sets $\{\, p \in X \mid \dim T_{X,p} \geqslant r\,\}$ are closed. Indeed, if we let I be the ideal generated by the $(n-r+1) \times (n-r+1)$-minors of the matrix $J(f_\bullet)$, we have

$$\{\, p \in X \mid \dim T_{X,p} \geqslant r\,\} = Z(\mathfrak{a} + I).$$

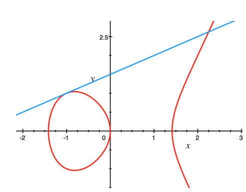
The affine tangent space
det affine tangentrommet

8.3 It is also convenient to introduce the *affine tangent space*, which is defined at a point $p = (a_1, \ldots, a_n)$ as the space

$$p + T_pX := \{\,(x, \ldots, x_n) \in \mathbb{A}^n \mid \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (x_i - a_i) = 0\,\}.$$

EXAMPLE 8.4 Let us consider the palne curve $C = Z(f) \subset \mathbb{A}^2$ which is given by the equation

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



For the point $p = (-1, 1) \in C$, the Jacobian matrix of the poliynomial f is the row-matrix

$$J = \begin{pmatrix} -3x^2 + 2 & 2y \end{pmatrix}\Big|_p = \begin{pmatrix} -1 & 2 \end{pmatrix},$$

and the affine tangent space to C at p has the equation

$$-(x+1) + 2(y-1) = 0.$$

☆

EXAMPLE 8.5 In this example we consider the space curve $C = Z(I) \subset \mathbb{A}^3$ which is the zero-locus of the ideal

$$\mathfrak{a} = (x + y + z^2 + xyz, x - 3z + x^2 + y^2).$$

For the point $p = (0, 0, 0) \in C$, one easily finds the Jacobian matrix which is

$$J = \begin{pmatrix} 1 + yz & 1 + xz & 2z + xy \\ 1 + 2x & 2y & -3 \end{pmatrix}\Bigg|_p = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -3 \end{pmatrix}$$

and so the affine tangent space at p is given by the two linear equations

$$x + y = x - 3z = 0.$$

Note that these equations exactly involve the linear terms of the generators of $\mathfrak{a}$ (cfr. Exercise 8.1 below) ☆

EXERCISE 8.1 Show that for $X = Z(I) \subset \mathbb{A}^n$ and $p = (0, \ldots, 0)$,

$$T_p X = Z(f^{(1)} \mid f \in \mathfrak{a})$$

where $f^{(1)}$ denotes the linear part of $f \in k[x_1, \ldots, x_n]$. ★

8.2 The Zariski tangent space

The above definition of the tangent space is easy to compute with, as it resembles the usual definition of a tangent space we know from undergraduate mathematics. However, it is not fully satisfactory, as it depends on the embedding of X into affine space $\mathbb{A}^n$. We will now describe a more intrinsic way to define the tangent space which makes sense for general varieties, and then show that this is equivalent to the previous definition.

Let X denote an algebraic variety over a field k, and let p be a point on X. Recall that we defined the *local ring* $\mathcal{O}_{X,p}$ of X at p as the ring rational functions which are regular at p. The local ring is an intrinsic invariant of X and p, and the ring which is best suited for studying the local geometry of X at p. We will therefore use this to defined the tangent space of X at p.

To motivate the definition, suppose first that $X \subset \mathbb{A}^n$ is an affine variety, and suppose for simplicity that $p = (0, \ldots, 0)$ is the origin (we may always arrange this by a linear change of coordinates). Write $\mathcal{M} = (x_1, \ldots, x_n) \subset k[x_1, \ldots, x_n]$. For each $f \in k[x_1, \ldots, x_n]$, we consider its *linearization at* p, given by

$$f^{(1)} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(0) x_i.$$

Note that the coordinates $x_1, \ldots, x_n$ give a basis for the dual space $(k^n)^*$. Hence we may view $f^{(1)}$ as a linear functional on k^n, and we get a k-linear map

$$\begin{aligned} d : \mathcal{M} &\rightarrow (k^n)^* \\ f &\mapsto f^{(1)} \end{aligned} \tag{8.2}$$

It is clear that d is surjective, since $d(x_i) = x_i$. If f lies in the kernel of d, then all the terms have order at least 2, so $f \in \mathcal{M}^2$. Hence d induces an isomorphism of k-vector spaces

$$\mathcal{M}/\mathcal{M}^2 \simeq (k^n)^*.$$

Returning to our variety X andthe tangent space $T_p(X)$, we take the dual of the inclusion $T_pX \subset k^n$, we obtain a surjection

$$(k^n)^* \to (T_pX)^*$$

obtained by restricting a linear functional on k^n to T_pX. Hence the composition

$$\theta : \mathcal{M}/\mathcal{M}^2 \to (k^n)^* \to (T_pX)^*$$

is also surjective.

Lemma 8.6 $\ker\theta = \mathcal{M}^2 + I(X)$.

Proof: Note that $f \in \ker\theta$ if and only if $f^{(1)}$ restricts to 0 on T_pX. This happens if and only if $f^{(1)} = g^{(1)}$ for some $g \in I(X)$ (since T_pX is the zero locus of $g^{(1)}$ for all $g \in I$). This happens if and only if $f - g \in \ker d = \mathcal{M}^2$, or equivalently, $f \in \mathcal{M}^2 + I(X)$. ❑

It follows that we have

$$(T_pX)^* \simeq \mathcal{M}/(\mathcal{M}^2 + I(X)) \simeq \frac{\mathcal{M}/I(X)}{(\mathcal{M}^2 + I(X))/I(X)} = \frac{\mathcal{M}/I(X)}{\mathcal{M}^2/I(X)} \simeq \mathfrak{m}/\mathfrak{m}^2. \quad (8.3)$$

where $\mathfrak{m} \subset \mathcal{O}_{X,p}$ is the maximal ideal. Taking duals, we now have:

Proposition 8.7 *For an affine variety $X \subset \mathbb{A}^n$, there is a natural isomorphism*

$$T_pX \simeq \operatorname{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k). \quad (8.4)$$

Using this result, we can give a formal definition of the tangent space for a general algebraic variety X. For a point $p \in X$, we define the *Zariski tangent space*, or simply *tangent space*, as the k-vector space

tangent space
tangentrommet

$$T_pX = \operatorname{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k).$$

This is functorial in (X, p) in the following sense. For a morphism of varieties $f : X \to Y$, the map between local rings $\mathcal{O}_{Y,f(p)} \to \mathcal{O}_{X,p}$ induces a k-linear map $\mathfrak{m}_{f(p)}/\mathfrak{m}^2_{f(p)} \to \mathfrak{m}_p/\mathfrak{m}^2_p$. If we dualize, we get an induced map on tangent spaces

$$df_p : T_pX \to T_{f(p)}Y.$$

Regular local rings

We recognize $\mathfrak{m}/\mathfrak{m}^2$ as the *Zariski cotangent space* from commutative algebra. This satisfies the following fundamental inequality:

Lemma 8.8 *If A is a Noetherian local ring with maximal ideal $\mathfrak{m}$ and residue field k, we have* $\dim_k \mathfrak{m}/\mathfrak{m}^2 \geqslant \dim A$

Proof: By Nakayama's lemma any basis $v_1, \dots, v_s$ of $\mathfrak{m}/\mathfrak{m}^2$ lifts to a set generators $x_1, \dots, x_s$ of the maximal ideal $\mathfrak{m}$. Krull's Hauptidealsatz implies that $ht\,\mathfrak{m} \leqslant s$, but $ht\,\mathfrak{m} = \dim A$. ❑

Way say that A is a *regular local ring* if the above inequality is an equality, or in other words, $\dim \mathfrak{m}/\mathfrak{m}^2 = \dim A$. Similarly, we say that X is *non-singular* at p if $\dim T_p X = \dim X$. Otherwise, we say that $p \in X$ is *singular*. We denote the set of singular points by $\operatorname{sing}(X)$.

regular local ring
regulær local ring
non-singular
ikke-singulær
singular point
singulært punkt

Proposition 8.9 *i) For a point $p \in X$, we have*

$$\dim T_p X \geqslant \dim X.$$

ii) p is a non-singular point on X if and only if the local ring $\mathcal{O}_{X,p}$ is a regular local ring.

iii) There is an open set $U \subseteq X$ such that $\dim T_p X = \dim X$ for all $p \in U$. In other words, the singular set $\operatorname{sing}(X)$ is closed.

Proof: (i) and (ii) follows directly from (8.4). For (iii), we may replace X by an open affine subset, and thus assume that $X \subset \mathbb{A}^n$. Then from (**??**), we see that $\operatorname{sing}(X) = \{\dim T_{X,p} \geqslant d+1\}$ is closed. ❑

From (**??**), we also have

Corollary 8.10 *An affine variety $X \subset \mathbb{A}^n$ of dimension d is non-singular at p if and only if $J(f_\bullet)$ has rank $n-d$.*

Examples

8.11 Consider the elliptic curve C in $\mathbb{A}^2$ given by the equation $y^2 = x(x^2-1)$, and let $P = (0,0)$. The local ring $\mathcal{O}_{C,P}$ is then regular; in fact, its maximal ideal is generated by the coordinate function y. The maximal ideal $\mathfrak{m}_P$ is generated by x and y, and in the local ring $\mathcal{O}_{C,P}$ the function $x-1$ is invertible. So it holds true that

$$x = y^2/(x^2-1). \tag{8.5}$$

8.12 On the other hand, the local ring $\mathcal{O}_{D,P}$ is not regular if D the *rational node* with equation $y^2 - x^2(x-1)$. To see this let $\mathfrak{n}_p$ be the ideal of $P = (0,0)$ in $\mathcal{O}_{\mathbb{A}^2,P}$ and $\mathfrak{m}_P$ that of P in $\mathcal{O}_{D,P}$. They are both generated by x and y, and restriction of functions induces a surjection $\theta\colon \mathfrak{n}_P/\mathfrak{n}_P^2 \to \mathfrak{m}_P/\mathfrak{m}_P^2$. We contend that this is an isomorphism. Therefore $\mathfrak{m}_P$ requires both generators x and y and is not principal. Indeed, the kernel of θ equals $(y^2 - x^2(x-1))\mathcal{O}_{\mathbb{A}^2,P} \cap \mathfrak{n}_P$ which is simply contained in the square $\mathfrak{n}_P$ since $y^2 - x^2(x-1)$ lies there. The situation is well illustrated with the short exact sequence

$$0 \longrightarrow (y^2 - x^2(x-1)) \longrightarrow \mathcal{O}_{\mathbb{A}^2,P} \longrightarrow \mathcal{O}_{C,P} \longrightarrow 0.$$

8.13 Consider again the cuspidal cubic $X = Z(f) \subset \mathbb{A}^3$, where $f = y^2 - x^3$. Let $p = (a,b) \in X$ be a point. The Jacobian of $f = y^2 - x^3$ is given by

$$J = \big(\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p)\big) = (-3a^2, 2b)$$

Thus the affine tangent space is given by

$$-3a^2(x-a) + 2b(y-b) = 0$$

in $\mathbb{A}^2$. Since X has dimension one, we see that X is singular at p if and only if $(a,b) = (0,0)$.

It is instructive to study the cotangent space $\mathfrak{m}/\mathfrak{m}^2$ at $p = (a,b)$. Write $\mathcal{M} = (x-a, y-b)$. Using the isomorphism (8.3), we have

$$\begin{aligned} \mathfrak{m}/\mathfrak{m}^2 = \mathcal{M}/(\mathcal{M}^2 + I(X)) &= \frac{(x-a, y-b)}{\big((x-a)^2, (x-a)(y-b), (y-b)^2, y^2 - x^3\big)} \\ &\simeq \frac{(x,y)}{\big(x^2, xy, y^2, (x+a)^3 - (y+b)^2\big)} = \frac{(x,y)}{\big(x^2, xy, y^2, 3a^2x + a^3 - 2by - b^2\big)} \end{aligned}$$

In the point $p = (0,0)$, this reduces to

$$\mathfrak{m}/\mathfrak{m}^2 \simeq (x,y)/(x^2, xy, y^2, y^2 - x^3) \simeq kx \oplus ky$$

Hence $\mathfrak{m}/\mathfrak{m}^2$ has dimension two, and so p is a singular point.

When $p \neq (0,0)$, this vector space has dimension one: It is spanned by x whenever $\operatorname{char} k \neq 2$ and y if $\operatorname{char} k = 2$.

8.14 Continuing the previous example, let us consider the parameterization of the cuspidal cubic $\phi : \mathbb{A}^1 \to \mathbb{A}^3$ given by $(x,y) \mapsto (t^2, t^3)$. This is induced by the ring map

$$k[x,y]/(y^2 - x^3) \to k[t] \tag{8.6}$$

sending $x \mapsto t^2, y \mapsto t^3$. If we let $p = 0$ be the origin in $\mathbb{A}^1$, so that $\phi(p) = (0,0)$, then the induced map $\phi^* : \mathcal{O}_{X,\phi(p)} \to \mathcal{O}_{\mathbb{A}^1,p}$ is just the localization of the map

(8.6) at the maximal ideal (x,y). Using the isomorphisms (8.3), we see that the induced map $\mathfrak{m}_{\phi(p)}/\mathfrak{m}^2_{\phi(p)} \to \mathfrak{m}_p/\mathfrak{m}_p^2$ is given by

$$(x,y)/(x,y)^2 + I(X) = (x,y)/(x^2, xy, y^2, y^2 - x^3) \to (t)/(t)^2.$$

And again, $x \mapsto t^2$ and $y \mapsto t^3$, which means that the differential $d\phi_p$ is the *zero map* at the point p.

☆

Exercises

8.2 Check that the map $d\phi_p$ in the previous example is an isomorphism for every point $p \in \mathbb{A}^1$ other than the origin.

8.3 Compute the singular points of the following curves in $\mathbb{A}^2$:

i) $x^2y + xy^2 - x^4 - y^4 = 0$

ii) $xy - x^6 - y^6 = 0$.

iii) $x^m - y^n = 0$ for m, n positive integers (beware that the ground field k may have positive characteristic!).

8.4 Compute the singular points of the *Steiner surface*

$$Z(x^2y^2 + y^2z^2 + z^2x^2 - xyz) \subset \mathbb{A}^3$$

8.5 For $X = Z(f_1, \ldots, f_r) \subset \mathbb{A}^n$ we define the *tangent bundle of* X has the set

$$T(X) = \{(x,v) \in X \times \mathbb{A}^n | \sum_{i=1}^n \frac{\partial f_j}{\partial x_i}(x) \cdot v_i = 0 \text{ for all } j\}$$

Show that $T(X)$ is an affine variety, and describe the morphism $p : T(X) \to X$ given by the first projection.

8.6 *(Tangent vectors and derivations.)* This exercise offers a different perspective on tangent vectors; elements of T_pX can be thought of as *derivations* of the local ring $\mathcal{O}_{X,p}$ into k. This perspective is similar to one appearing in multivariate calculus, where one talks about *directional derivatives* $\nabla_v f$. Formally, we define a *k-derivation* to be a k-linear map $D : \mathcal{O}_{X,p} \to k$ which is satisfies Leibniz' rule

Derivations
derivasjoner

$$D(fg) = f(p)D(g) + g(p)D(f).$$

We denote the set of such derivations by $\mathrm{Der}_k(\mathcal{O}_{X,p}, k)$.

a) Show that any derivation is zero on the constant functions $k \subset \mathcal{O}_{X,p}$;

b) Show that $\mathrm{Der}_k(\mathcal{O}_{X,p}, k)$ is a k-vector space;

c) Show that the map $f \mapsto f(p) + (f - f(p))$ defines a splitting of k-vector spaces$\mathcal{O}_{X,p} = k \oplus \mathfrak{m}$;

d) For an element $f \in \mathcal{O}_{X,p}$, we define its *differential* $df \in \mathfrak{m}/\mathfrak{m}^2$ as the projection of f onto the second factor, i.e., $df = f - f(p) \mod \mathfrak{m}^2$;

Differentials
differentialer

e) Let $X = \mathbb{A}^n$ and $p = (0, \ldots, 0)$. Show that the usual formula folds:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(0) \cdot dx_i;$$

f) Show that the map $d : \mathcal{O}_{X,p} \to \mathfrak{m}/\mathfrak{m}^2$ satisfies the Leibniz rule;

g) Deduce that any element $v \in T_pX = \mathrm{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k)$ determines a derivation $D_v : \mathcal{O}_{X,p} \to k$ by $D_v = v \circ d$:

h) Show that there is a natural isomorphism of k-vector spaces

$$\phi : T_pX \to \mathrm{Der}_k(\mathcal{O}_{X,p}, k).$$

In other words, we have a second intrinsic definition of the tangent space T_pX; it is the k-vector space of derivations $\mathcal{O}_{X,p} \to k$.

★

8.3 The Jacobian criterion in the projective case

Consider a hypersurface $X = Z_+(F) \subset \mathbb{P}^n$ given by a homogeneous polynomial F of degree d and let $p \in X$ be a point. For simplicity, we assume that $p = (1 : w_1 : \ldots : w_n) \in D_+(x_0)$. Then the affine tangent space of $X \cap D_+(x_0)$ in $D_+(x_0) \simeq \mathbb{A}^n$ is given by

$$p + T_pX = \{(z_1, \ldots, z_n) | \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (z_i - w_i) = 0\}$$

where $f(x_1, \ldots, x_n) = F(1, x_1, \ldots, x_n)$. We define the *projective tangent space* $\mathbb{T}_pX$ as the projective closure of this inside $\mathbb{P}^n$, that is, the hyperplane given by

$$\mathbb{T}_pX = \{(x_0 : \ldots : x_n) \in \mathbb{P}^n | \frac{\partial F}{\partial x_i}(1, w_1, \ldots, w_n) \cdot (x_i - w_i x_0) = 0\}$$

Since F is homogeneous of degree d, we may simplify this relation using Euler's formula

$$dF = \sum_j x_j \frac{\partial F_i}{\partial z_j}$$

To prove Euler's formula: Both sides are linear in F, so it suffices to establish the equality for monomials. Then if $F = x_1^{d_0} \cdot \ldots \cdot x_n^{d_n}$: Then $x_j \partial_j F = d_j F$, and summing over j concludes since $d = \sum_j d_j$.

Since $F(1, w_1, \ldots, w_n) = 0$, we get that

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(1, w_1, \ldots, w_n) \cdot (-w_i x_0) = \frac{\partial F}{\partial x_0}(1, w_1, \ldots, w_n) \cdot x_0$$

and so

$$\mathbb{T}_p(X) = \{(z_0 : \ldots : z_n) \in \mathbb{P}^n | \sum_{i=0}^{n} \frac{\partial F}{\partial x_i}(p) x_i = 0\}$$

More generally, for a projective variety $X \subset \mathbb{P}^n$), we define the projective tangent space $\mathbb{T}_p X$ as

$$\mathbb{T}_p(X) = \left\{ (x_0 : \ldots : x_n) \in \mathbb{P}^n \mid \sum_{i=0}^{n} \frac{\partial f_j}{\partial x_i}(p) x_i = 0 \text{ for all } f \in I(X) \right\}$$

The Jacobian criterion works well in $\mathbb{P}^n$ as well:

PROPOSITION 8.15 *Let $X = Z_+(F_1, \ldots, F_r) \subset \mathbb{P}^n$ be a closed algebraic set, and let*

$$J = \left(\frac{\partial F_i}{\partial z_j}(p) \right)$$

Then the the rank of J does not depend on the choice of representative for p. Moreover X is non-singular at p if and only if $\operatorname{rank} J = n - \dim X$.

EXAMPLE 8.16 Consider the Fermat hypersurface $X = Z(F) \subset \mathbb{P}^n$ given by

$$F = x_0^p + x_1^p + \ldots + x_n^p$$

Then the jacobian J is given by

$$J = (p x_0^{p-1}, p x_1^{p-1}, \ldots, p x_n^{p-1})$$

Thus if $\operatorname{char} k \neq p$, J has rank 1 at every point, and X is non-singular.

If $\operatorname{char} k = p$, then

$$F = (x_0 + x_1 + \ldots + x_n)^p$$

so $X = Z(x_0 + \ldots + x_n) \simeq \mathbb{P}^{n-1}$ is non-singular also here. ☆

The non-singular points are dense

The aim of this section is to prove that any variety X contains a dense open set $U \subset X$ consisting only of non-singular points, in other words, the set of singular points $\operatorname{sing}(X)$ is a closed proper subset. We know that the set of singular points $\operatorname{sing}(X)$ is a closed set in X, so what remains is to exclude that $\operatorname{sing}(X) = X$. For that purpose, we need only show that the non-singular locus is non-empty, i.e., that X contains a single non-singular point.

LEMMA 8.17 *Let $\phi : X \dashrightarrow Y$ denote a birational map. If X contains a non-singular point, then so does Y.*

PROOF: By definition, there exist non-empty open sets $V \subset X$ and $W \subset Y$ so that ϕ restricts to an isomorphism $\phi|_V : V \to W$. By assumption, there is an open set $U \subset X$ containing only non-singular points. Thus Y contains a non-singular point: any point in $\phi(V \cap U)$ will do. ❏

LEMMA 8.18 *Any variety is birational to a hypersurface $Y = Z(f) \subset \mathbb{A}^{n+1}$.*

PROOF:

❑

To conclude, we now need to prove that any hypersurface $Y = Z(f) \subset \mathbb{A}^{n+1}$ has a dense non-singular locus. Let $x_0, \ldots, x_n$ be affine coordinates on $\mathbb{A}^{n+1}$ and let $f \in k[x_0, \ldots, x_n]$ be an irreducible polynomial. Then $p \in X$ is a singular point if and only if $\frac{\partial f}{\partial x_i}(p) = 0$ for every $i = 0, \ldots, n$. Thus if $\operatorname{sing}(X) = X$, we find that each partial derivative $\frac{\partial f}{\partial x_i}$ lies in $I(X) = (f)$. But this is impossible unless $\frac{\partial f}{\partial x_i}$ is the zero polynomial, as $\frac{\partial f}{\partial x_i}$ has strictly smaller degree than f and f is assumed to be irreducible.

In characteristic 0 we are already done, since the condition that each $\frac{\partial f}{\partial x_i} \equiv 0$ implies that f is constant. If k has characteristic p: Looking at monomials in f, we find that the only way that $\frac{\partial f}{\partial x_i} \equiv 0$ for every i is that f is a polynomial in $x_0^p, \ldots, x_n^p$. But then, by the Binomial Theorem, we get that $f = g^p$ is a p-power of a polynomial $g \in k[x_1, \ldots, x_n]$, contradicting the assumption that f was irreducible. This completes the proof.

8.4 Normal varieties

For a variety X over k, the local rings $\mathcal{O}_{X,x}$ are integral domains, sitting in their fraction field $k(X)$. We say that X is *normal* at a point x, if the local ring $\mathcal{O}_{X,x}$ is *normal*, i.e., integrally closed in $k(X)$, and X is *normal* if it is normal everywhere.

normal varieties
normale varieteter

EXAMPLE 8.19 $\mathbb{A}^n$ and $\mathbb{P}^n$ are both normal. This follows because in each case $\mathcal{O}_{X,x}$ is the localization of a polynomial ring, which is normal. ☆

EXAMPLE 8.20 Consider the cuspidal curve $X = Z(y^2 - x^3)$. Then X is not normal at the origin $x = (0,0)$. Indeed, we have seen that $A(X) = k[x,y]/(y^2 - x^3) \simeq k[t^2, t^3]$ via the map sending $x \mapsto t^2$ and $y \mapsto t^3$. The local ring is therefore equal to $\mathcal{O}_{X,x} = k[t^2, t^3]_{(t)}$ with fraction field $k(t)$. This ring, however is not normal: the element $t \in k(t)$ satisfies the monic equation $W^2 - t^2 = 0$, but it does not lie in $\mathcal{O}_{X,x}$. ☆

EXAMPLE 8.21 *(Nodal cubic)* Let now $X = Z(y^2 - x^3 - x^2) \subset \mathbb{A}^2$ denote the nodal cubic, with coordinate ring $A = k[x,y]/(y^2 - x^3 - x^2)$ (where k now is a field whose characteristic is not two (if the characteristic is two, we are back in previous cuspidal case).

As above, the origin $(0,0)$ is a special point: a line $l \subset \mathbb{A}^2_k$ through the closed point $(0,0) \in X$ (with equation $y = tx$) will intersect X at $(0,0)$ and at one more point (with $x = t^2 - 1$), and this gives a parameterization of the curve, which is generically one-to-one.

We can turn this into an algebraic statement by introducing the parameter $t = yx^{-1}$ in the function field K of X, the equation $y^2 = x^3 - x^2$ then reduces to $t^2 = 1 + x$ after being divided by x^2. Moreover, the element t is integral, since it satisfies the monic equation $W^2 - x - 1 = 0$ (which has coefficients in A). Since $x = t^2 - 1$ and $y = x \cdot y/x = t^3 - t$, we see that

$$A = k[t^2 - 1, t^3 - t] \subseteq k[t] \subseteq K = k(t),$$

and since $k[t]$ is integrally closed, any element in K which is integral over A, can be written as a polynomial in t. So $\overline{A} = k[t]$ is the integral closure of A in $k(t)$. ☆

EXAMPLE 8.22 *(The quadratic cone)* Consider the quadric cone $X = Z(xy - z^2) \subset \mathbb{A}^3$ with coordinate ring $A = \mathbb{C}[x, y, z]/(xy - z^2)$. This variety is also singular at the origin, but we still claim that X is normal.

Let $B = \mathbb{C}[x, y]$, so that $A = B[z]/(z^2 - xy)$. Then $B \subset A$ is a ring extension making A into a finite B-module. We get an inclusion of fields $K(B) = \mathbb{C}(x, y) \subset K(A)$ obtained by adjoining the element z ($= \sqrt{xy}$). Write an element of $K(A)$ as $w = u + vz$ where $u, v \in K(B) = \mathbb{C}(x, y)$. If this is integral over A, it is also integral over B. In fact, w satisfies the minimal polynomial

$$T^2 - 2uT - (x^2 + y^2)v^2 = 0$$

If this is integral over B, we must have $2u \in \mathbb{C}[x, y]$ and hence $u \in \mathbb{C}[x, y]$. Moreover $u^2 - (x^2 + y^2)v^2 \in \mathbb{C}[x, y]$, so also $(x^2 + y^2)v^2 \in k[x, y]$. Note that $(x^2 + y^2) = (x - iy)(x + iy)$ is a product of coprime, and irreducible elements, so we must have also $v^2 \in k[x, y]$, and for the same reason $v \in k[x, y]$. Hence $u + vz \in B[z]$. ☆

EXERCISE 8.7 For the quadratic cone $X = Z(xz - y^2)$, prove that there is an isomorphism

$$\phi : A(X) \to \mathbb{C}[u^2, uv, v^2]$$

and use this to deduce that $A(X)$ is integrally closed in $k(X) = k(u, v)$. ★

Regular functions on normal varieties

PROPOSITION 8.23 *Let X be a normal affine variety of dimension $\geqslant 2$ and let $x \in X$ be a point. Then if f is a function which is regular at $X - x$, then f extends to a regular function on X.*

PROOF: Recall the 'algebraic Hartog's lemma' which says that for a notherian normal integral domain A, A is equal to the intersection of all its intersections at height 1 primes, i.e.,

$$A = \bigcap_{ht\, \mathfrak{p}=1} A_{\mathfrak{p}}$$

We apply this to $A = A(X)$. Let $\mathfrak{p}$ denote a prime ideal of height 1 and let $g \in \mathfrak{m}_x - \mathfrak{p}$ (which exists because $\dim X \geqslant 2$!). We get $A_g \subset A_{\mathfrak{p}}$. Note that $X - x = \bigcup_{h \in \mathfrak{m}_x} D(h)$, so if $f \in \mathcal{O}_X(X - x)$ is regular on $X - x$, it restricts to an element in $\mathcal{O}_X(D(g)) = A_g$, and hence $f \in A_{\mathfrak{p}}$. Hence $f \in A$ by the equality above, and so f is regular at p. ❑

Of course the above proposition fails when X has dimension 1: The regular function $\frac{1}{x}$ on $\mathbb{A}^1 - 0$ does not extend to a regular function on all of $\mathbb{A}^1$.

EXERCISE 8.8 From the result above, deduce a stronger statement: For a variety X which is normal at a point $x \in X$, any section $f \in \mathcal{O}_X(U - x)$ extends to a section of $\mathcal{O}_X(U)$ (where $U \subset X$ is an open set). ★

Normal varieties are non-singular in codimension 1

8.5 Desingularizations via blow-ups

DEFINITION *Given a variety X, a* desingularization *of X is a non-singular variety Y together with a birational morphism*

$$\pi : Y \to X$$

One can view the desingularization Y as a sort of "nice model" of X. X and Y share isomorphic open sets, but Y has the advantage of being non-singular, so it tends to be easier to compute its invariants. Note that the geometry of X and Y are closely related: $k(X) = k(Y)$, $Bir(X) = Bir(Y)$, $\mathcal{O}_Y(Y) = \mathcal{O}_X(X)$, etc. O

Let $C \subset \mathbb{A}^2$ be a curve passing through the point $p = (0,0)$. The blow-up $\mathrm{Bl}_p(\mathbb{A}^2)$ of $\mathbb{A}^2$ in p is the subset $\mathrm{Bl}_p(\mathbb{A}^2) \subset \mathbb{A}^2 \times \mathbb{P}^1$ defined by the equation

$$u_1 x - u_0 y = 0$$

where $(u_0 : u_1)$ are a homogeneous coordinates on $\mathbb{P}^1$, and $\mathrm{Bl}_p(\mathbb{A}^2)$ comes together with the blow-up morphism $\pi : X \to \mathbb{A}^2$. The exceptional divisor $E = \pi^{-1}(p)$ is defined by $x = y = 0$ in $\mathrm{Bl}_p(\mathbb{A}^2)$. The total transform of C is given by

$$\pi^{-1}(C) = E \cup \tilde{C}$$

and $\tilde{C}$ proper transform.

The main point here is that the morphism π restricts to a birational map $\pi : \tilde{C} \to C$. The curve $\tilde{C}$ typically has milder singularities than C; thus iterating the procedure eventually produces a non-singular model of C.

EXAMPLE 8.25 Consider the cuspidal cubic $C = Z(y^2 - x^3)$, which is singular at the origin $p = (0,0)$.

In the affine chart $U_0 = \mathbb{A}^2$, where we set $u_0 = 1$, we have setting $t = u_1$, affine coordinates (x, t) with the relation $y = tx$, and the total transform of C is given by the equation

$$y^2 - x^3 = 0,$$

which, when the relation $y = tx$ is taken in to account, takes the form

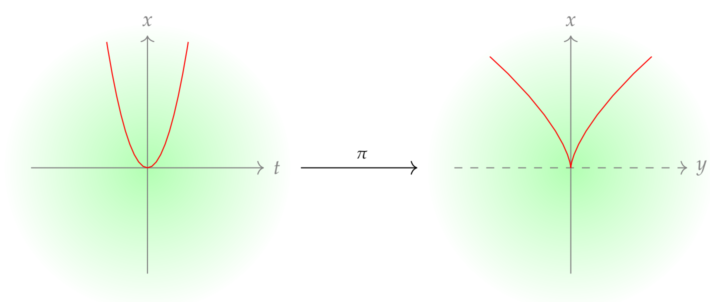
$$t^2x^2 - x^3 = x^2(t^2 - x) = 0.$$

It is clear that $\pi^{-1}(C)$ has two components; one given by $x = 0$, which is just the exceptional fibre $E \cap U_0$, and one with equation $t^2 - x = 0$, the proper transform. Notice that latter is non-singular (it is a parabola).

The other chart U_1, the coordinates are (y, s) with $x = sy$. Similar manipulations give

$$y^2 - x^3 = y^2 - s^3y^3 = y^2(1 - sy) = 0.$$

Once again there are two components; one given given by $y = 0$, which just the exceptional fibre $E \cap U_1$, and the proper transform $\tilde{C} \cap U_1$ given as the locus where $1 - sy = 0$. Note that also $\tilde{C}$ is non-singular in this chart (it is a hyperbola). It follows that $\tilde{C}$ is a non-singular curve, and that $\pi : \tilde{C} \to C$ is a desingularization of C.



☆

Example 8.26 Consider the nodal cubic $C = Z(y^2 - x^3 - x^2)$, which is singular at the origin $p = (0, 0)$. As in the previous example, we study the strict transform $\tilde{C}$ using the two charts U_0 and U_1. In U_0 the total transform of C is given by the equations

$$y^3 - x^3 - x^2 = 0, \quad tx - y = 0.$$

Eliminating y, we see that $\pi^{-1}(C) \cap U_0$ is isomorphic to the algebraic set given by the equalities

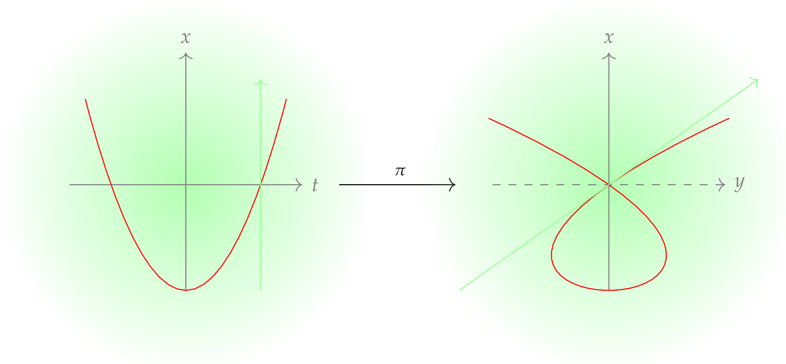
$$t^2x^2 - x^3 - x^2 = x^2(t^2 - x - 1) = 0.$$

There are two components: $Z(x)$, which defines $E \cap U_0$, and $Z(t^2 - x)$, which defines $\tilde{C} \cap U_0$. Note that the latter curve $\tilde{C} \cap U_0$ is non-singular.

In the second chart U_1, the total transform is given by

$$y^2 - s^3y^3 = y^2(1 - s^3y) = 0.$$

Here $\tilde{C} \cap U_1$, which is defined by $1 - s^3 y = 0$. Once again this is non-singular (it is just the graph of s^{-3}), and $\pi : \tilde{C} \to C$ is a desingularization of C. ☆



Example 8.27 *(Quadric cone)* Consider the quadratic cone $Q = Z(xy - z^2) \subset \mathbb{A}^3$. This has one singular point, namely $p = (0, 0, 0)$. We will produce a desingularization $\pi : \widetilde{Q} \to Q$ by studying the blow-up of Q and $\mathbb{A}^3$ at p.

The blow-up of $\mathbb{A}^3$ is given by $X \subset \mathbb{A}^3 \times \mathbb{P}^2$ defined by the 2×2-minors of

$$\begin{pmatrix} u_0 & u_1 & u_2 \\ x & y & z \end{pmatrix}$$

We want to consider the strict transform $\widetilde{Q}$ of Q. In the chart U_0, where $u_0 = 1$, the total transform is defined by the equations

$$\begin{aligned} y &= u_1 x & z &= u_2 x \\ u_1 z &= y u_2 & xy - z^2 &= 0, \end{aligned}$$

or after eliminating y and z, it is defined by the single equation

$$xy - z^2 = u_1 x^2 - u_2^3 x^2 = x^2(u_1 - u_2^2).$$

Once again this locus has two components: $Z(x)$, which equals the exceptional divisor $E \cap U_0$; and $Z(u_1 - u_2^2)$, which equals the strict transform $\widetilde{Q} \cap U_0$. Note that the latter component is non-singular (it is isomorphic to $\mathbb{A}^2$ in fact).

In the open affine U_1, where $u_1 = 1$ and which has coordinates u_0, u_2, y, similar manipulations show that the total transform is given by

$$y^2(u_0 - u_2^2) = 0$$

which has again $\widetilde{Q} \cap U_1$ non-singular. Finally, in U_2, with coordinates u_0, u_1, z, the total transform is given by

$$z^2(u_0 u_1 - 1) = 0$$

which shows that $\widetilde{Q} \cap U_2$ is non-singular. Hence $\widetilde{Q}$ is non-singular everywhere, and we have found the desired desingularization. ☆

EXERCISE 8.9 In the previous example, we can also try to blow-up the line $L = Z(x, y)$. Show that also this produces non-singular model of Q. ★

EXAMPLE 8.28 *(The Atiyah flop)* Let x, y, z, w be coordinates on $\mathbb{A}^4$, and consider the quadratic cone

$$Q = Z(xw - yz) \subset \mathbb{A}^4$$

Note that this is the affine cone over the Segre quadric $Z_+(x_0x_3 - x_1x_2)$ in $\mathbb{P}^3$. The variety Q is singular at the point $p = (0, 0, 0, 0)$. The blow-up of $\mathbb{A}^4$ at p is the subvariety $\widetilde{\mathbb{A}^4} \subset \mathbb{A}^4 \times \mathbb{P}^3$ given by the vanishing of the 2×2-minors of the matrix

$$\begin{pmatrix} u_0 & u_1 & u_2 & u_3 \\ x & y & z & w \end{pmatrix}$$

We claim that the blow-up X of Q at p is non-singular. In the affine chart $U_0 = \pi^{-1}(D_+(u_0) = \mathbb{A}^4 \times D_+(u_0)$, we have $y = xu_1$, $z = xu_2$, $w = xu_3$. Eliminating these, shows that the total transform $U_0 \cap \pi^{-1}(Q)$ is given by

$$x^2(u_3 - u_1u_2) = 0$$

in $\mathbb{A}^4$ (with coordinates x, u_1, u_2, u_3). Thus $X \cap U_0$ is isomorphic to $u_3 - u_1u_2 = 0$ in $\mathbb{A}^3$, which is non-singular. The same happens for the other affine charts, so X is non-singular.

There are other ways of desingularizing Q. One can blow up each of the planes $P_1 = Z(x, y)$ and $P_2 = Z(x, z)$.

The blow-up of $\mathbb{A}^4$ along P_1 is given by the subvariety of $\mathbb{A}^4 \times \mathbb{P}^1$ defined by the equation

$$u_0y - u_1x = 0$$

Intersecting this with $\pi^{-1}(Q)$ gives the total transform

$$u_0y - u_1x = xy - zw = 0$$

This algebraic set is not irreducible: it decomposes as $E \cup X$, where X is the strict transform given by

$$u_0y - u_1x = u_0w - u_1z = xy - zw = 0$$

☆

Chapter 9

Curves

9.1 *Curves*

A *curve* is a variety X of dimension one. The aim of this Chapter is to establish the basic facts of the birational geometry of curves.

curve
kurve

Our main interest is that of non-singular curves. We recall from Chapter 8 that a variety is non-singular if the local ring $\mathcal{O}_{X,P}$ is regular. For one-dimensional Noetherian rings, as the ring $\mathcal{O}_{X,P}$ is, being regular is equivalent to being *normal*, that is, integrally closed in its field of fractions. This is one of the reasons why the theory for curves is substantially easier than for general varieties – one obtains a non-singular model just by normalizing X. Another reason is the fact rational maps from a non-singular curve into a projective space is defined everywhere, that is to say, it is a regular map. This result, which is a sort of advanced form of l'Hôpital's rule, has a counterpart in theory the general, similar to a theorem in complex analysis called Hartog's theorem, which asserts that rational maps from normal varieties are defined of closed subset of codimension at least two.

This implies that birational maps between projective non-singular curves are isomorphisms, and consequently there is up to isomorphism only one non-singular and projective curve in a birational class. In a nutshell, there is no distinction between biregular and birational geometry of curves.

Another consequence of the extension property of rational maps between curves is that every non-singular curve is isomorphic to an open set of a non-singular *projective* curve. In particular, any finitely generated field of transcendence degree one over an algebraically closed field k is the function field of a projective and non-singular curve.

Discrete valuation rings

9.1 A Noetherian local ring A with $\dim A = n$ and with maximal ideal $\mathfrak{m}$ is said to be *regular* if the maximal ideal can be generated by n elements; that is, by as many elements as the dimension indicates. Nakayama's lemma tells us that the

Regular local rings
regulære lokale ringer

minimal number of generators of $\mathfrak{m}$ equals the so called embedding dimension $\dim_{A/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2$ of A, so A is regular precisely when the Krull dimension and the embedding dimension coincide. A general ring A is *regular* if all the local rings $A_\mathfrak{p}$ are regular.

9.2 When it comes to one-dimensional rings, which is our main concern in this section, A is regular if and only if $\mathfrak{m}$ is principal. This has many equivalent formulations, and we list the few we shall need.

Proposition 9.3 *Let A be a Noetherian local domain with maximal ideal $\mathfrak{m}$ of dimension one. Then the following are equivalent*

i) The maximal ideal $\mathfrak{m}$ is principal;

ii) A is a PID and all ideals are powers of $\mathfrak{m}$;

iii) A is integrally closed.

iv) A is regular, i.e., $\dim_k \mathfrak{m}/\mathfrak{m}^2 = 1$.

Proof: $i) \Rightarrow ii)$. Let x a generator for the maximal ideal $\mathfrak{m}$ and let $\mathfrak{a} \subseteq A$ be a non-zero ideal. Let n be the largest integer such that $\mathfrak{a} \subseteq \mathfrak{m}^n$. Krull's intersection theorem asserts that $\bigcap_i \mathfrak{m}^i = 0$, and the ideal $\mathfrak{a}$ is therefore not contained in all powers of $\mathfrak{m}$ and such an n exists. Since $\mathfrak{a} \nsubseteq \mathfrak{m}^{n+1}$, there is an $a \in \mathfrak{a}b$ such that $a = bx^n$ with $b \notin \mathfrak{m}$; that is, b is a unit since the ring is local. It follows that $(x^n) \subseteq \mathfrak{a}$, and we are done.

$ii) \Rightarrow iii)$. Every PID is a UFD and all UFD's are integrally closed.

$iii) \Rightarrow i)$. Finally, assume that A is integrally closed in its fraction field K and let $x \in \mathfrak{m}$ be any element. Since A is Noetherian and of dimension one, there is an element $y \in A$ not in (x) such that $(x : y) = \mathfrak{m}$. This means that $yx^{-1}\mathfrak{m} \subseteq A$. We contend that $yx^{-1}\mathfrak{m} = A$. If not, one would have $yx^{-1}\mathfrak{m} \subseteq \mathfrak{m}$ and since $\mathfrak{m}$ is a finitely generated and faithful A-module it would follow that yx^{-1} is integral over A. Hence it holds that $yx^{-1} \in A$ since A is integrally closed, and therefore also $y \in (x)$, which is not the case.

$(i) \Leftrightarrow (iv)$. If $\mathfrak{m} = (x)$, then $\mathfrak{m}/\mathfrak{m}^2$ is generated by the class of x modulo $\mathfrak{m}^2$. We also have $\mathfrak{m} \neq \mathfrak{m}^2$ (since A has dimension 1), so $\dim_k \mathfrak{m}/\mathfrak{m}^2 = 1$. The converse implication follows by Nakayama's lemma. ❏

9.4 A ring as in the proposition is also a *discrete valuation ring*. If t is a generator for the maximal ideal $\mathfrak{m}$, one calls t a *uniformizing parameter* of A. All non-zero ideals in A are of the form (t^ν) with $\nu \in \mathbb{N}_0$, and therefore any non-zero element in the fraction field K may be written as αt^ν with α a unit in A and ν an integer. Indeed, if $f \in A$ and $f \neq 0$, we let $v(f)$ be the unique non-negative integer such that $(f) = \mathfrak{m}^{\nu(f)}$, then $f = \alpha t^{v(f)}$ with α being a unit, and for a general non-zero element fg^{-1} of the function field, one finds $fg^{-1} = \alpha t^{\nu(f)-\nu(g)}$ with α a unit.

Uniformizing parameters
uniformiserende parameter

9.5 The function $v\colon A\backslash\{0\} \to \mathbb{Z}$ sending f to the unique integer such that $f = \alpha t^{v(f)}$ with α a unit, is called the *valuation* associated to A. It resembles the well-known order function from complex function theory (recall that every meromorphic function has an order at a point, positive if its a zero and negative

Valuations
valuasjoner

in case of a pole), and it share several of its properties. For instance, the two following identities hold:

- $\nu(fg) = \nu(f) + \nu(g)$;
- $\nu(f+g) \geqslant \min\{\nu(f), \nu(g)\}$,

with equality in the latter when $\nu(f) \neq \nu(g)$. Any function $A\backslash\{0\} \to \mathbb{Z}$ satisfying these two properties is called a *discrete valuation* on A. We some times extend this definition to include 0, by assigning $v(0) = \infty$; in that case v is a map from $v : A \to \mathbb{Z} \cup \infty$. We will also sometimes extend the valuation to the whole fraction field $K = K(A)$ by defining $v(a/b) = v(a) - v(b)$.

Discrete valuations

EXAMPLE 9.6 For $A = \mathbb{Z}$, and $K = \mathbb{Q}$ and p a prime number we have the *p-adic valuation* on K defined by $v_p(a/b) = v_p(a) - v_p(b)$ where $v_p(a) = n$ if $a = p^n m$ and $p \nmid m$. ☆

EXAMPLE 9.7 The field $K = k(t)$ has a valuation $v : K^\times \to \mathbb{Z}$ given by $v(f) = n$ if $f \in k[t]$ is of the form $f = c(t)t^n$ and $c(0) \neq 0$. Note that $v^{-1}(\mathbb{Z}_{\geqslant 0}) = k[t]_{(t)}$ (the localization at the prime ideal (t)). ☆

9.8 Coming back to the geometric context, we let P be a non-singular point of the curve X. A uniformizing parameter t at P is a rational function on X which is regular at P and generates the maximal ideal $\mathfrak{m}_P$. Another common way of phrasing this is to say that t is regular and vanishes to first order at P, or that t has a simple zero at P. Every function f can be expressed as $f = \alpha t^{\nu_P(f)}$ with α a rational function on U that is regular and non-vanishing at P. One may think about the valuation $v_P(f)$ as the order of f at P, either the order of vanishing if f is regular at P or the oder of the pole if not.

EXERCISE 9.1 Assume that ν is a discrete valuation on a field K. Show that the set $A = \{\, x \in K \mid \nu(x) \geqslant 0 \,\}$ is discrete valuation ring by showing that $\{\, x \in K \mid \nu(x) > 0 \,\}$ is a maximal ideal generated by one element. ★

The extension lemma

9.9 We start with establishing the main property of curves in the present context, that any rational map from a curve into a projective varletry is defined at all non-singular points of the curve.

One may think about this as an advanced form of "l'Hôpital's" rule. The main idea of the proof is first to realise the mapping in a neighbourhood of P as the composition $\pi \circ \Phi$ where $\pi\colon \mathbb{A}^{n+1}\backslash\{0\} \to \mathbb{P}^n$ is the canonical projection, and where the map Φ is represented as $\Phi = (g_0, \ldots, g_n)$ with the g_i's regular functions near P, and then cancel out the common factors of the g_i's that vanish at P.

Lemma 9.10 *Let U be a curve and $P \in U$ a non-singular point. Assume that $\phi\colon U\backslash\{P\} \to \mathbb{P}^n$ is a morphism. Then there exists a morphism $\psi\colon U \to \mathbb{P}^n$ extending ϕ.*

Proof: The first observation is that it suffices to find an open $U_0 \subseteq U$ containing P over which ϕ extends. Indeed, if $\psi_0\colon U_0 \to \mathbb{P}^n$ is such an extension, the two morphisms ψ_0 and ϕ coincide on $U_0\backslash\{P\}$, and hence they patch together to a morphism on U. It follows that we may assume U to be affine.

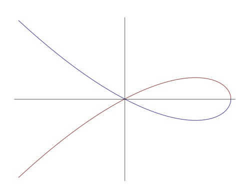
Secondly, we may, possible after having renumbered the coordinates, assume that the image $\phi(U\backslash\{P\})$ meets the basic open set $D = D_+(x_0)$; then the inverse image $V = \phi^{-1}(D)$ is a non-empty open subset of U. The basic open set D is an affine n-space with coordinates $x_1x_0^{-1}, \ldots, x_nx_0^{-1}$, and the map $\phi|_V$ is therefore given by n component functions regular on V. They are all rational functions on U, and may therefore be written as fractions $f_i = g_i/g_0$ of regular functions on U (for some fixed g_0)

Consider the morphism $\Phi(x) = (g_0, g_1, \ldots, g_n)$ from U into $\mathbb{A}^{n+1}$. It is well defined at the point P, but of course, it might be that it maps P to the origin. However, if this is not the case, the composition $\pi \circ \phi$ is defined at P and extends ϕ to the neighbourhood of P where the g_i's do not vanish simultaneously, and we will be done.

Now, the salient point is that we have the liberty to alter the morphism Φ by cancelling common factors of the g_i's without changing the composition $\pi \circ \Phi$: After such a modification the composition $\pi \circ \Phi$ and the original morphism ϕ coincide where they both are defined. Indeed, it holds true that $(hg_0 : \cdots : hg_n) = (g_0 : \cdots : g_n)$ where both sets of homogeneous coordinates are legitimate.

To get rid of common zeros the functions g_i's might have at the point P, we introduce a uniformizing parameter t at P; that is, a regular function t on some neighbourhood U_0 of P which generates the maximal ideal of the local ring $\mathcal{O}_{U,P}$. One may then write $g_i = \alpha_i(t)t^{\nu_i}$ with the $\alpha_i(t)$'s being regular functions on U_0 that do not vanish at P, and where the ν_i's are non-negative integers. Putting $\nu = \min_i \nu_i$, the differences $\mu_i = \nu_i - \nu$ will be non-negative, and at least one will be zero so that the corresponding g_i does not vanish at P. Hence replacing the g_i's by $g_it^{-\nu} = \alpha_i(t)t^{\nu_i-\nu}$ we arrive at the requested modification of Φ. ❑

Example 9.11 The assumption that P be a non-singular point is essential. For instance, let $U \subseteq \mathbb{A}^2$ be the "ordinary double point" given as $U = Z(y^2 - x^2(1 - x))$; and let $\phi(x, y) = yx^{-1}$. Then ϕ is defined on $U\backslash\{0\}$ but can not be extend. We have depicted the situation over the reals; when the absolute value $|x|$ is small, $|1 - x|$ is bonded away from zero, and the curve has two distinct (analytic) components parametrized as $y = x\sqrt{1-x}$ (the red one) and $y = -x\sqrt{1-x}$ (the blue one). The function yx^{-1} approaches 1 when x approaches zero along the red component, and it tends to -1 when x goes to zero while staying on the blue. This shows that there is not even a continuous extension. ☆



The Extension Theorems

Most of the work is done in proving the lemma, and we can collect the fruits:

THEOREM 9.12 (THE EXTENSION THEOREM) *Let X be a curve and $P \in X$ a non-singular point. Any rational map $\phi\colon X \dashrightarrow Y$ where Y is a projective variety, is defined at P.*

PROOF: Assume the projective variety Y is a closed subvariety of $\mathbb{P}^m$; that is, $Y \subseteq \mathbb{P}^m$. Let U be a neighbourhood of P such that ϕ is defined on $U \setminus \{P\}$. By the extension lemma (Lemma 9.10 above), the map ϕ composed with the inclusion Y into $\mathbb{P}^m$ extends to P, and the extension takes values in Y, since Y is closed in $\mathbb{P}^m$. ❑

As is illustrated in Example 9.11 above it is paramount that P be a non-singular point. If X has *e.g.* two different branches passing through P, the "limit" of ϕ at P along the two branches may be different.

THEOREM 9.13 *Assume that X and Y are two projective and non-singular curves that are birationally equivalent. Then they are isomorphic.*

PROOF: Let $U \subseteq X$ and $V \subseteq Y$ be two open sets such that there is an isomorphism $\phi\colon U \dashrightarrow V$. Since Y is projective and X is non-singular a repeated application of Theorem 9.12 above gives a morphism $\Phi\colon X \to Y$ extending ϕ. Similarly, there is morphism $\Psi\colon Y \to X$ extending ϕ^{-1}. Finally, the Hausdorff axiom holds for both X and Y, and one deduces that $\Phi \circ \Psi = \mathrm{id}_Y$ and $\Psi \circ \Phi = \mathrm{id}_X$ since they extend $\phi \circ \phi^{-1} = \mathrm{id}_V$ and $\phi^{-1} \circ \phi = \mathrm{id}_U$ respectively. ❑

Desingularization of curves

Every curve has a non-singular model. This is relatively easy to prove due to the fact that for curves being non-singular is the same as being normal. However it hinges on a non-trival result about normalization.

THEOREM 9.14 (FINITENESS OF INTEGRAL CLOSURE) *Let A be a domain finitely generated over the field k with fraction field K. For any finite field extension L of K, the the integral closure B of A in L is a finite module over A.*

In particular taking $L = K$, we see that the integral closure of A in K is finite over A. For domains other than those being finitely generated over a field, this theorem is subtle, and in positive characteristic it is not generally true, even for Noetherian domains.

9.15 Given a variety X, we apply this to the coordinate ring $A(X)$ of X, letting B be the integral closure of $A(X)$ in the function field $k(X)$. The theorem tells us that B is a finitely generated algebra over the ground field k, and hence there is an affine variety $\widetilde{X}$ whose coordinate ring equals B. The inclusion $A(X) \subseteq B$ induces a morphism $\widetilde{X} \to X$, which is finite because B is a finite $A(X)$-module, and since $A(X)$ and B have the same fraction field, this morphism is birational.

Proposition 9.16 *Any affine variety X has a normalisation $\widetilde{X}$; that is, a normal affine variety and a finite birational morphism $\pi\colon \widetilde{X} \to X$. It enjoys the universal property that any dominating morphism $\psi\colon Y \to X$ whose source Y is normal, factors through π; in other words, there is a morphism $\phi\colon Y \to \widetilde{X}$ such that $\pi = \psi \circ \phi$.*

Proof: Most is already accounted for, only the factorization remains. But ψ is dominant and induces an injection $k(X) \subseteq k(Y)$, and because $A(Y)$ is integrally closed in $k(Y)$, any element in $k(X)$ integral over $A(X)$ belongs to $A(Y)$. It follows that $A(\widetilde{X}) \subseteq A(Y)$ and the corresponding morphism $Y \to \widetilde{X}$ is the requested map. ❑

9.17 Recall that by Lying–Over (Proposition 7.11 on page 146) finite maps are closed, and they are surjective when dominating, so the same holds true for the normalisation morphisms $\pi_X\colon \widetilde{X} \to X$; they are closed and surjective.

9.18 Of course, one would desire a construction of a normalization for any variety, not only for affine ones, and indeed there is one. The proof consists of patching together the normalizations of members of an open affine cover of X. We shall not go through this time consuming process in general, but confine our attention to curves. Owing to the extension theorem, there is a shortcut making the gluing very easy in that case, which we learned from the book[1] by Igor Shafarevich.

[1]

Theorem 9.19 *Let X be a quasi-projective curve. Then X has a normalization.*

Proof: The first reduction is to find a cover of X of just two affine opens, U_1 and U_2. The curve X lies in some projective space $\mathbb{P}^m$ and we may choose a hyperplane h_1 not containing the curve. Then h_1 cuts the curve in finitely many points, and we choose a second hyperplane h_2 avoiding those finitely many points. Putting $U_i = D_+(h_i) \cap X$ we obtain our two affine open subset that cover X.

The next step is to introduce the normalization V_i of the U_i's and the normalization maps $\pi_i\colon V_i \to U_i$. Moreover, we shall need an affine open subset $U \subseteq X$ contained in $U_1 \cap U_2$ all whose points are non-singular. Then U lies naturally in both the V_i's as an open dense subset and $\pi_i|_U = \mathrm{id}_U$ for $i = 1, 2$.

The idea is now to "glue" V_1 and V_2 together along U, and this will amount to embedding them both in a variety W, with the the two embedding coinciding on U, and taking their union inside W.

To construct a suitable W, we begin with embedding each of the V_i's in some projective space, as big as one needs, and taking the projective closure there, we obtain two projective curves W_1 and W_2, having respectively V_1 and V_2 as open dense subsets. Our W will be the product $W_1 \times W_2$.

Now we come to the point where The Extension Theorem helps us. The Extension Theorem yields morphisms $\phi_1 \colon V_1 \to W_2$ and $\phi_2 \colon V_2 \to W_1$. Indeed, there are natural inclusions $U \subseteq V_1$ and $U \subseteq V_2$ and since V_1 is non-singular and W_2 projective, the inclusion of U into W_2 extends to a morphism $\phi_1 \colon V_1 \to W_2$. And in an analogous way we find a morphism $\phi_2 \colon V_2 \to W_1$.

It follows that we have morphism $V_1 \hookrightarrow W_1 \times W_2$ and $V_2 \hookrightarrow W_1 \times W_2$ both being injective, one of the component maps being the inclusion $V_i \subseteq W_i$. The first one, for instance, is the map sending x to $(x, \phi_1(x))$, and it factors as $V_1 \to V_1 \times W_2 \hookrightarrow W_1 \times W_2$ where the first map identifies V_1 with the graph of ϕ_1. The graph is a closed subvariety and V_1 is isomorphic to it. Closing up V_1 in the product $W_1 \times W_2$ gives closed subvariety $\overline{V_1}$ and since $\bar{V}_1 \cap V_1 \times W_2 = V_1$, we se that V_1 is an open subvariety of $\bar{V}_1$. Of course the same applies to V_2 and V_2 is an open subvariety of $\overline{V_2}$.

Since U is a common open dense set of V_1 and V_2, it holds true that $\bar{U} = \overline{V_1} = \overline{V_2}$.

We contend that $\widetilde{X} = V_1 \cup V_2$ is a normalization of X. Since both V_i's are normal and they cover $\widetilde{X}$, it is normal. It is birational to X and the two maps π_i coincide on U, and hence patch together to a map $\widetilde{X} \to X$. ❑

9.20 In proof from the previous paragraph we realized the normalization $\widetilde{X}$ as an open dense subset of the closed subvariety $\bar{U}$ of the projective variety $W_1 \times W_2$. The variety $\bar{U}$ is therefore a projective curve having $\widetilde{X}$ as a dense open subset, which can be all of $\bar{U}$, and in fact, when the original curve X is projective this will the case. This leads to:

> **Theorem 9.21** *If X is a projective curve, the normalization $\widetilde{X}$ is projective as well.*

Proof: Denote by $\pi_X \colon \widetilde{X} \to X$ the normalization map. In the remark preceeding the theorem, we described the inclusion $\widetilde{X} \subseteq \widetilde{U}$ of $\widetilde{X}$ as an open subset of a projective curve. Let $\widetilde{U}$ be the normalization of $\bar{U}$. The normalization map $\pi_{\bar{U}} \colon \widetilde{U} \to \bar{U}$ induces a rational map into X and since X is projective and $\widetilde{U}$ is non-singular it extends to morphism $\psi \colon \widetilde{U} \to X$. By the universal property of the normalization map $\phi_X \colon \widetilde{X} \to X$ the map ψ factors through $\widetilde{X}$; that is, there is a map $\phi \colon \widetilde{U} \to \widetilde{X}$ such that $\psi = \phi \circ \pi$. But $\pi_{\bar{U}}$ being surjective, it follows that $\widetilde{X} = \bar{U}$ and $\widetilde{X}$ is projective. ❑

The non-singular model

Theorem 9.22 (Fundamental theorem for curves) *Given a field K of transcendence degree one over k. Then there exists a non-singular projective curve X, unique up to isomorphism, such that $K \simeq k(X)$*

Proof: From field theory, we know that we can find an $x \in K$, so that K is a finite separable extension of $k(x)$. Then there is an element $f \in X$ so that $K = k(x)[f]$, and f satisfies an irreducible polynomial

$$y^n + a(x)y^{n-1} + \ldots + a_1(x)y + a_0(x) = 0 \tag{9.1}$$

over $k[x]$. When x and y are Interpreted as coordinates on the affine plane $\mathbb{A}^2$, the equation (9.1) is the equation of an irreducible curve Y whose function field is precisely K.

Now take the projective closure $\overline{Y} \subset \mathbb{P}^2$ and let X be the normalization of $\overline{Y}$. Then X is non-singular, projective and $k(X) = k(Y) = K$. ❑

9.2 Elliptic curves are not rational

Let k be an algebraically closed field of characteristic not equal to 2, and let X be the curve in $\mathbb{A}^2$ defined by

$$y^2 = p(x)$$

where $p(x)$ is a square-free polynomial, e.g., $p(x) = x(x-\lambda)(x-\mu)$ where $\lambda, \mu \in k^\times$ are different. The aim of this section is to prove the following result:

Theorem 9.23 *X is not rational.*

In other words, we need to prove that $k(X) \neq k(t)$. For that we will use an argument that uses valuations. The main characteristic property of $k(t)$ we will use is the following:

Lemma 9.24 *Let $g \in k(t)$. If $v(g)$ is even for every valuation $v : k(t)^\times \to \mathbb{Z}$, then g is a square.*

Proof: In $k(t)$ we may factor g as

$$g(t) = \prod_i (t - a_i)^{n_i}$$

for some $n_i \in \mathbb{Z}$ and $a_i \in k$. Then $n_i = \mathrm{ord}_{a_i}(g)$ are all even, so g is a square. ❑

Now, write $K = k(x)$ and $L = K(y) = k(X)$ where y satisfies $y^2 = x(x-\lambda)(x-\mu)$. We claim

Lemma 9.25 *$v(x)$ is even for every valuation $v : L^\times \to \mathbb{Z}$.*

PROOF: Let $v : L^\times \to \mathbb{Z}$ be a valuation. There are three cases:

If $v(x) = 0$, then the lemma clearly holds.

If $v(x) < 0$, then $v(x-\lambda) = \min(v(x), v(\lambda)) = v(x)$ and $v(x-\mu) = \min(v(x), v(\mu)) = v(x)$. Hence

$$2v(y) = v(y^2) = v(x(x-\lambda)(x-\mu)) = 3v(x)$$

so $v(x)$ is also even.

If $v(x) > 0$, then $v(x-\lambda) = \min(v(x), v(\lambda)) = v(\lambda) = 0$ and $v(x-\mu) = \min(v(x), v(\mu)) = v(\mu) = 0$. Hence

$$2v(y) = v(y^2) = v(x(x-\lambda)(x-\mu)) = v(x)$$

so $v(x)$ is even also in this case. ❑

Finally, to prove Theorem 9.23, we need only prove

LEMMA 9.26 *$x \in L$ is not a square in L.*

PROOF: Note that L is a 2-dimensional vector space over K with basis $1, y$.

If x is a square, we may write $x = (a+by)^2$ for some $a, b \in K$. In other words,

$$\begin{aligned} x &= (a+by)^2 \\ &= a^2 + 2aby + b^2y^2 \\ &= a^2 + 2aby + b^2x(x-\lambda)(x-\mu) \end{aligned}$$

Comparing the two sides, we must have either $a = 0$ or $b = 0$.

In the case $b = 0$, we must have $x = a^2$ is a square in $K = k(x)$, which is clearly not the case (e.g., as x has valuation 1 with respect to the valuation $v = \text{ord}_0$ at $x = 0$).

Likewise, in the case $a = 0$, we have $x = b^2x(x-\lambda)(x-\mu)$, so $(x-\lambda)(x-\mu)$ is a square in K. However, this is not possible, since $(x-\lambda)(x-\mu)$ has valuation 1 at $v = \text{ord}_\lambda$. Thus in both cases, we obtain a contradiction, so x is not a square in L. ❑

Chapter 10

The Structure of Maps

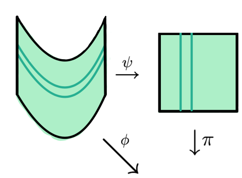
Topics in Chapter 10: *Generic structure of morphism—semi-continuity of fibre dimension—constructible sets—Chevalley's theorem —images of projective varieties—The Fundamental Theorem of Elimination Theory —degree of maps—more on finite maps*

Morphisms are the main tool to study varieties; heuristically speaking one breaks varieties up in the target space, often called the base, and the fibres, and one hopes to give a good description of the variety in terms of properties of the base and the fibres; but in general morphisms can be impenetrable, so even"hope" is a too strong word. However, there are some general principles that apply, and some general structure theorems which we will expound in this chapter; and of course many morphisms are good-natured, well understood and very useful.

10.1 Generic structure of morphisms

10.1 Among the many nice applications of Noether's Normalization lemma one finds a general structure theorem for dominating morphisms between varieties; or we should rather call it a generic structure theorem. Morphisms can be utterly intricate, but over a sufficiently small (but still dense) open subset of the target they are to a certain extent well behaved and factors as the composition of a projection and a finite map. Projections are harmless, but finite morphisms can be complex creatures, so the generic intricacies of the map are hidden in the finite part; be aware, however, that complications are mostly not not generic and appear outside the open set.

Theorem 10.2 (Generic structure of dominant morphisms) *Let X and Y be varieties. If a morphism $\phi\colon X \to Y$ is dominant, there exist open affine subsets $U \subseteq Y$ and $V \subseteq X$ such that V maps into U and such that $\phi|_V$ factors as $\phi|_V = \pi \circ \psi$ where $\pi\colon \mathbb{A}^n \times U \to U$ is the projection and $\psi\colon V \to \mathbb{A}^n \times U$ is a finite map.*



$$\begin{array}{ccccc} X & & \xrightarrow{\phi} & & Y \\ \uparrow & & & & \uparrow \\ V & \xrightarrow[\psi]{} & \mathbb{A}^n \times U & \xrightarrow[\pi]{} & U \end{array}$$

Finite maps preserve dimensions by Going–Up (Proposition 7.16 on page 147) from which ensues that $n + \dim U = \dim V$. Open dense sets have the same dimension as the surrounding space, and the integer n appearing in the theorem therefore equals the relative dimension $\dim X - \dim Y$.

PROOF: We let L be the function field of X and K that of Y. Since ϕ is dominating it gives rise to an extension $K \subseteq L$. Chose any open and affine subset $U \subseteq Y$ and denote its affine coordinate ring by A. Let $V \subseteq X$ be any open affine subset mapping into U. The coordinate ring B of V then contains A and is finitely generated as an A-algebra.

$$\begin{array}{ccc} K & \subset & L \\ \cup & & \cup \\ A_h & \subset & B_h \\ \cup & & \cup \\ A & \subset & B \end{array}$$

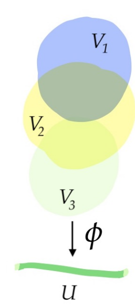
The algebra $B_K = B \otimes_A K$ is a finitely generated algebra over K as B is finitely generated over A. Noether's Normalization lemma applies, and there are elements $w_1, \ldots, w_n$ which are algebraically independent over K and are such that B_K is a finite module over K. We also pick generators $z_1, \ldots, z_r$ for B_K over $K[w_1, \ldots, w_n]$.

The basic trick is to replace U and V by smaller distinguished open affine subsets, U by $D(h)$ and V by $\phi|_V^{-1}(U_h) = D(h \circ \phi|_V)$, where h is the product of all denominators that might occur either in the w_i's or in the generators z_j's for B_K over $K[w_1, \ldots, w_n]$; the coordinate ring of $D(h)$ will be the localized ring A_h and that of $D(h \circ \phi|_v)$ will be B_h.

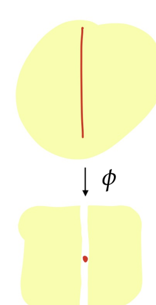
Each w_i may be written as $w_i = b_i s_i^{-1}$ with $b_i \in B$ and $s_i \in A \backslash \{0\}$, and the same for the z_i's, they are of the form $z_i = c_i t_i^{-1}$ with $c_i \in B$ and $t_i \in A \backslash \{0\}$. As our element h we take the product of all the s_i's and all the t_i's, then A_h is obtained by adjoining the denominators s_i^{-1} and t_i^{-1} to A, and the w_i's and the z_i's all lie in B_h. Moreover, $A_h[w_1, \ldots, w_n]$ is contained in B_h, and B_h is a finite module over it, and of course, the w_i's persist being algebraically independent so that $A_h[w_1, \ldots, w_n]$ is isomorphic to a polynomial ring over A_h. ❑

Actually, one has the slightly more general result.

THEOREM 10.3 *With the setting as in the theorem, there exists an open set U and a finite covering of the inverse image $\phi^{-1}(U)$ with affine open sets V_i such each restriction $\phi|_{V_i}$ factors as $\phi|_{V_i} = \pi \circ \psi_i$ where $\psi_i \colon V_i \to \mathbb{A}^n \times U$ is finite.*



PROOF: Start with any finite affine and open covering V_i of $\phi^{-1}(U)$. Shrink U sufficiently to work for all V_i. ❑



10.4 We have seen several instances of dominating morphisms having rather complicates images. The projection of a quadratic surface in $\mathbb{P}^3$ from a point on it for example, has as image $\mathbb{P}^2 \backslash L \cup \{p_1, p_2\}$ where $L \subseteq \mathbb{P}^2$ is a line and p_1 and p_2 are two points on the line, and a routine example is the map $\phi \colon \mathbb{A}^2 \to \mathbb{A}^2$ that sends (x, y) to (x, xy). Its hits all points in $\mathbb{A}^2$ save those on the y-axis different from the origin; that is, the image is $D(x) \cup \{(0, 0)\}$. However, the images of dominant maps will always contain a Zariski open set.

COROLLARY 10.5 *The image of a dominant morphism contains a Zariski open set.*

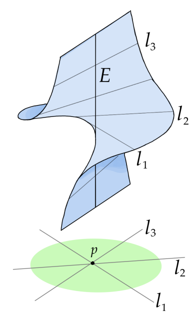
PROOF: The open set $U \subseteq Y$ that appears in the theorem is contained in the image of ϕ since both finite maps and a projections are surjective. ❑

The dimension of fibres

A good concept of dimension should comply to the principle of being "additive along maps". In other words, we expect formulas like

$$\dim X = \dim Y + \dim \phi^{-1}(y) \qquad (10.1)$$

for $y \in Y$. Such an equality holds for linear maps as we have encountered in courses on linear algebra; the dimensions of the kernel and the image add up to the dimension of the source. For differentiable manifolds the dimension is governed by tangent spaces, and the derivative of a map is expressed as a bunch of linear maps of tangent spaces, so one expects a similar relation between the dimension of fibres, image and source, at least for fibres over general points.



10.6 For varieties, the equality (10.1) does not always hold. Think for instance about the blow-up of a point p, where the problem is that the dimension of the exceptional fibre $\phi^{-1}(p)$ is too large. However, as we will see shortly, (10.1) holds for general points y, that is, for points belonging to an open dense subset of the target variety.

Without further hypotheses on the varieties there are not many limitations for the fibre dimensions over the "bad" points, but there is one governing principle: the fibre dimension is an upper semicontinuous function; of a fibre can only have higher dimension when compared with nearby fibres. This principle has as consequence of a theorem of Claude Chevalley asserting that the points whose fibres are of a given dimension, form a so-called *constructible set*. Thus the fibres of a morphisms do not behave too wildly.

Claude Chevalley (1909–1984) French Mathematician

10.7 We begin with the generic case which is an easy corollary of the structure theorem above.

THEOREM 10.8 *If $\phi\colon X \to Y$ is a dominant morphism of varieties, there is an open set $U \subseteq Y$ so that for for every $x \in U$ and every component Z of the fibre $\phi^{-1}(x)$ it holds true that*

$$\dim Z = \dim X - \dim Y.$$

PROOF: Let U be an open dense subset of Y as in the general version 10.3 of the structure theorem. For $x \in U$ and Z a component of $\phi^{-1}(x)$, at least one of the members V_i of the covering meets Z. Then the fibre of $\phi|_{V_i}$ over x equals $Z \cap V_i$, and $\dim Z = \dim Z \cap V_i$. Now one may factor ϕ as

$$V_i \longrightarrow \mathbb{A}^n \times U \longrightarrow U$$

with one map being finite and the other a projection. It follows that $\phi|_{Z\cap V_i}$ is a finite dominant map $Z \cap V_i \to \mathbb{A}^n \times \{x\}$, and by Going–Up (Proposition 7.16 on page 147) we infer that $\dim Z = n$. ❏

One should compare this with Proposition 7.49 on page 159, which says that for any point $y \in Y$ belonging to the image of ϕ, at least we have the inequality

$$\dim Z \geqslant \dim X - \dim Y$$

for each component Z of $\phi^{-1}(y)$.

10.9 The frequently occurring case that the two varieties X and Y are of the same dimension, merits a statement by itself. In that case the relative dimension is zero, and no projection is needed in the generic factorization. One has:

Proposition 10.10 *Assume that two varieties X and Y have the same dimension; that is, $\dim X = \dim Y$. Let $\phi\colon X \to Y$ be a dominant morphism. Then there is a dense open subset $U \subseteq Y$ so that if $V = \phi^{-1}(U)$, the restriction $\phi|_V$ is a finite map.*

Semi-continuity of the fibre dimension

10.11 An important part of the analysis of a morphism $\phi\colon X \to Y$ is to understand the partition of Y into the subsets where the fibres of ϕ have a given dimension. These sets can have a rather intricate topology, but the sets $W_r(\phi)$ where the fibre dimension is *at least* a given value r, are topologically simpler. They turn out to be closed. Formally, we define

$$W_r(\phi) = \{\, y \in Y \mid \dim \phi^{-1}(y) \geqslant r \,\}.$$

Recall that the dimension of an irreducible space is the supremum of the dimensions of the different components, so a point y belongs to $W_r(\phi)$ precisely when $\dim Z \geqslant r$ for at least one component Z of the fibre $\phi^{-1}(y)$. Bearing Equation (10.7) above in mind, it is clear that $W_r(\phi)$ equals the entire target Y for values of r less than the relative dimension $\dim X - \dim Y$

That $W_r(\phi)$ is a closed subsets of Y is commonly referred to by saying that the dimension is an *upper semi-continuous* function – or as geometers sometimes say it, that the fibre dimension *increases upon specialization*.

Proposition 10.12 (Principle of upper semi-continuity) *Let $\phi\colon X \to Y$ be a morphism. Then $W_r(\phi)$ is closed in Y.*

PROOF: We proceed by induction on the dimension $\dim Y$ and begin with picking an open dense subset U of Y as in the Structure Theorem. The complement of U is the union of a finite collection of closed irreducible sets $\{Z_i\}$, all of dimension less than $\dim Y$. Their inverse images $\phi^{-1}(Z_i)$ are again unions of closed irreducible subsets Z_{ij} of X. Clearly $W_r(\phi)$ is the union of the sets $W_r(\phi|_{Z_{ij}})$ which are all closed in Z_i by induction. They are thus closed in Y, since the Z_i's are, and it follows that $W_r(\phi)$ is closed being a finite union of closed sets. ❏

Constructible sets

10.13 In a topological space X a *locally closed subset* is a subset that is the intersection of an open and a closed set. In other words it is closed in an open set (or for that matter, open in a closed one). A subset of X is *constructible* if is the union of finitely many locally closed sets.

Locally closed sets
lokalt lukkede mengder

Constructible sets
konstruerbare mengder

COROLLARY 10.14 *The locus where the fibres of the morphism $\phi\colon X \to Y$ are of dimension r, is locally closed.*

PROOF: By Proposition 10.12, the locus of points y in Y where $\dim \phi^{-1}(y) = r$ equals $W_r(\phi) \backslash W_{r+1}(\phi)$ which is open in $W_r(\phi)$. ❏

10.15 The induction procedure from the proof of Proposition 10.12 above, with minor modifications, yields the following characterization of images of morphisms due to Claude Chevalley.

THEOREM 10.16 (CHEVALLEY'S THEOREM) *Let $\phi\colon X \to Y$ be a morphism of varieties. Then the image $\phi(X)$ is constructible.*

Chevalley's theorem is one of the basic facts in algebraic geometry asserting that the image of a polynomial map $\mathbb{A}^n \to \mathbb{A}^n$ can be described by a certain number of polynomial equations $g_1 = g_2 = \ldots = g_r = 0$ together with a certain number of "non-equations" $h_1 \neq h_2 \neq \ldots \neq h_s$.

PROOF: Again the proof is by induction on $\dim Y$. Since locally closed subsets of a closed subset are locally closed in the surrounding space, we may well assume that the morphism ϕ is dominating. Pick an open subset U as in the Structure Theorem. The complement is a finite union of components Z_i, all having dimension less than Y, and each inverse image $\phi^{-1}(Z_i)$ is the finite union of components Z_{ij}. By induction the image of $\phi|_{Z_{ij}}$ is constructible, and as $\phi(X) = U \cup \bigcup_{i,j} \phi|_{Z_{ij}}(Z_{ij})$, we are through as finite unions of constructible sets are constructible. ❏

Exercises

Recall that in a topological space X a locally closed subset is a subset that is the intersection of an open and a closed set, and that a subset of X is constructible if being the union of finitely many locally closed sets. A *Boolean algebra* $\mathcal{B}$ in X is a collection of subsets of X that is closed under finite set-theoretical operations: It contains the entire space and the empty set, finite unions and intersections of members of $\mathcal{B}$ belong to $\mathcal{B}$ and the set-theoretical difference of two members is a member as well.

Boolean algebra
Boolsk algebra

10.1 Let X be a topological space. Prove that the collection of constructible sets in X is the smallest Boolean algebra containing the open (or the closed) sets. Prove that inverse image of constructible sets under continuous maps are constructible.

10.2 *(Chevalley's Nullstellensatz.)* Show that the image of a constructible set under a morphism between varieties is constructible.

10.3 Let X be an affine variety. Let f and g be regular functions on $X \times \mathbb{P}^1$ such that $Z(f,g) \cap \{x\} \times \mathbb{P}^1$ is finite for all $x \in X$. Let $: X \times \mathbb{P}^1 \to X$ denote the projection. Show that $\pi(Z) \subseteq X$ has all components of codimension one. Conclude that if $A(X)$ is a UFD, then $\pi(Z) = Z(h)$ for some $h \in A(X)$.

10.4 *(The resultant.)* Let A be a ring. Denote by R_n the A-module consisting of polynomials in $A[t]$ of degree strictly less and n. Then R_n is free of rank n with the powers t^i for $i < n$ as a basis. Given two polynomials $f(t)$ and $g(t)$ in $A[t]$ of degree n and m respectively. Consider the k-linear map

$$R_n \times R_m \to R_{n+m}$$

that sends the pair (p,q) to $qf + pg$, and let Φ be its matrix in the bases whose elements are the powers of t. The determinant $\det \Phi$ is called the *resultant* of f and g and commonly written as $\mathrm{Res}(f,g)$.

The resultant
resultanten

a) Assume that A is a field k. Prove that $\det \Phi = 0$ if and only if f and g has a common root in some field extension of k. HINT: Start by showing that Φ has a kernel if and only if f and g has a common factor.
b) Let $\mathfrak{m}$ be a maximal ideal in A and let $k = A/\mathfrak{m}$. Denote by $\bar{h}$ the image in $k[t]$ of a polynomial $h \in A[t]$. Show that $\det \Phi \in \mathfrak{m}$ if and only $\bar{f}$ and $\bar{g}$ has a common root in an extension of k.
c) Show that $\mathrm{Res}(f,g)$ belongs to the ideal (f,g) in $A[t]$ generated by f and g. Compare with Exercise 10.3.

★

10.2 *Morphisms from projective varieties*

A very convenient property of compact topological spaces, and the main reason for their utility, is that they have closed images under continuous maps, and the projective varieties behave in this respect like compact spaces, their images are always closed.

For projective curves, this was established already in Chapter 9, and the proof we shall offer for the general case is a reduction to the curve case with the help of the generic structure theorem.

Another important aspect of projective varieties is that they do not have global regular functions other than constants—or locally constants when the variety is not connected—so if a variety X is fibered over an affine variety Y with fibres being connected projective varieties, one would expect every global function on X to come from a function on Y; and indeed, this is holds true. When the fibres are not connected (*e.g.* the case of curves) some finite cover is involved, and the global functions are no more forced to be constant, but they will be integral over the coordinate ring $A(Y)$. In particular, the global regular functions on X form a finite module over $A(Y)$.

The image is closed

10.17 The proof is a simplistic variant of what in scheme theory is called *the valuative criterion for properness*. The idea is rather simple, and a very rough sketch of the proof skipping most of the subtle points goes as follows: From the structure theorem we obtain a "good" open set U in the image. If y is a point not in that "good" open set, we pick a curve lying in U whose closure passes through y. This curve can be lifted to the target since we have strong control over the map over the good open set, and since the target is projective, the lifting extends to the entire closure of the curve; finally, knowing the result for curves, we obtain a point in the closure mapping to y.

THEOREM 10.18 *Let $\phi\colon X \to Y$ be a morphism between two varieties and let $Z \subseteq X$ be a closed subvariety. If Z is projective, then $\phi(Z)$ is closed.*

PROOF: Replacing Y by the closure of the image of ϕ we may assume that ϕ is dominating, the goal then being to prove that ϕ is surjective. Choose an open affine subset U of Y as in the Structure Theorem and denote by V an open affine subset of the inverse image $\phi^{-1}(U)$ as in that theorem. The restriction $\phi|_V$ then factors as

$$V \xrightarrow{\psi} U \times \mathbb{A}^n \xrightarrow{\rho} U, \qquad \phi|_V = \rho \circ \psi$$

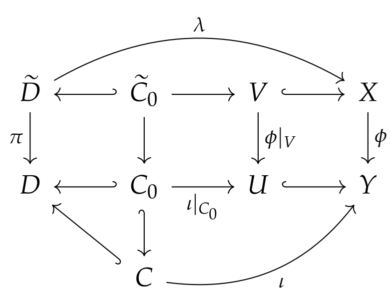
where ψ is finite and ρ the projection onto U. Finite maps being surjective, points in U belong to the image, so the issue is about points in $Y \backslash U$.

Pick then a point $y \in Y \backslash U$; we have to exhibit a point $x \in X$ with $\phi(x) = y$. As hinted at when introducing the theorem, the idea is to take a non-singular curve C and a morphism $\iota\colon C \to Y$ whose image is closed in Y, which passes trough y and which meets the open subset U. Lemma 10.19 below guarantees that such curves abound; just normalize the curves found there. To proceed, choose a $z \in C$ that maps to y, denote the inverse image of U in C by C_0 and let D be a projective and non-singular curve containing C.

The map $\iota|_{C_0}\colon C_0 \to U$ obviously lifts to an inclusion $\kappa\colon C_0 \to U \times \mathbb{A}^n$. We want to extend this lifting all the way up to V, but for that to be possible, we must replace C_0 by a finite cover.

To this end, consider $\psi^{-1}(\kappa(C_0))$ and let W be one its irreducible components. Then the function field $k(W)$ is a finite extension of the function field $k(C_0)$, hence the integral closure of C_0 in $k(W)$ maps into W and we have found the required lifting. The integral closure $\widetilde{D}$ of D in $k(W)$ is regular, and the induced normalization map $\pi\colon \widetilde{D} \to D$ is finite, hence surjective, and it extends the map $\widetilde{C}_0$ to C_0.

The slightly involved set up is summarized with the following tactical situation map where hooked arrows designate inclusions:



Pick a point $w \in \widetilde{D}$ such that $\pi(w) = z$. The crucial point is now that because the curve $\widetilde{D}$ is regular and the variety$X$ is projective, the map $\widetilde{C}_0 \to V$ extends to a map $\lambda\colon \widetilde{D} \to X$. It follows that $\phi(\lambda(w)) = \iota(\pi(w)) = \iota(z) = y$; so $x = \lambda(z)$ is our man, and y lies in the image of ϕ. ❑

Lemma 10.19 *Given a point x in the variety X and a closed subset Z containing x. Then there is an irreducible curve passing by x not contained in Z.*

Proof: Replacing X by an open affine neighbourhood of x, we may assume that X is affine (in general, just close up a curve found in this case).

Let f be a regular function with $Z \subseteq Z(f)$. It suffices to find a curve through x not lying in $Z(f)$. After Problem 7.9 on page 158 we may find a system of parameters $f_1, f_2, \ldots, f_n$ with $f_1 = f$ at x, and citing the same problem, we infer that all the components of $Z(f_2, \ldots, f_n)$ are of codimension $n-1$, so any one of them passing by x will be a curve as we search for. ❑

10.20 The following corollary goes under the name *The Fundamental Theorem of Elimination Theory*. The reason for the name is rooted in the classical situation of a variety $Z \subseteq \mathbb{A}^m \times \mathbb{P}^n$ given by the common zeros of a collection of polynomials $f_i(a_1, \ldots, a_m, x_0, \ldots, x_n)$ homogeneous in the x_i's. The image $\pi(Z)$ under the projection π onto $\mathbb{A}^m$ consists of points $(a_1, \ldots, a_m)$, so that the equations $f_i(a_1, \ldots, a_m, x_0, \ldots, x_n) = 0$ can be solved for the x_i's. The classical way to describe this set is to eliminate the x_i's from these equations, and the theorem states that this is doable (at least theoretically!).

Corollary 10.21 *Let X be a variety and let $Z \subseteq X \times \mathbb{P}^n$ be a closed subset. Let $\pi\colon X \times \mathbb{P}^n \to X$ denote the projection. Then $\pi(Z)$ is closed.*

Proof: Because every variety has an open cover whose members are quasi-projective (*e.g.* affine), the proposition is easily reduced to the case that X is quasi-projective. We may thus assume that X is an open subset of a projective variety W. Let $\overline{Z}$ denote the closure of Z in $W \times \mathbb{P}^n$. The theorem yields that $\pi(\overline{Z})$ is closed in W, but the equality $\pi(Z) = X \cap \pi(\overline{Z})$ holding true, $\pi(Z)$ will be closed in X. Indeed, we have $Z = \overline{Z} \cap (X \times \mathbb{P}^n)$ because Z is closed in the open set $X \times \mathbb{P}^n$, from which ensues that for $x \in X$ it holds true that

$$\{x\} \times \mathbb{P}^n \cap \overline{Z} = \{x\} \times \mathbb{P}^n \cap \overline{Z} \cap X \times \mathbb{P}^n = \{x\} \times \mathbb{P}^n \cap Z.$$

This entails the inclusion $X \cap \pi(\overline{Z}) \subseteq \pi(Z)$, and the reverse inclusion being trivial, we are through. ❑

10.22 The heuristic notion of a variety being fibered in projective varieties has as formal counterpart the notion of a projective morphism: a morphism $\phi\colon X \to Y$ is said to be *projective morphism* if there is a factorization

Projective morphism
projektive morfier

$$\begin{array}{ccc} X & \xhookrightarrow{\iota} & Y \times \mathbb{P}^n \\ \downarrow{\scriptstyle \phi} & \swarrow{\scriptstyle \mathrm{pr}_Y} & \\ Y & & \end{array}$$

where ι is a closed embedding and pr_Y is the projection. According to The Fundamental Theorem of Elimination the image of a projective morphism is closed. One also also verifies effortlessly that if ϕ is projective and Z is any variety, then the product map $\phi \times \mathrm{id}_Z\colon X \times Z \to Y \times Z$ is also projective.

A finiteness result

10.23 Besides being closed maps, projective maps enjoy a very important finiteness property which we now come to, which is a special case of a very general theorem of Alexander Grothendieck about for us such mysterious things as "higher directed images of coherent sheaves". The proof we give is quite simple, and it retains a certain flavour of the Rabinowitz trick from the proof of the Nullstellensatz.

Theorem 10.24 *Let Y be an affine variety with coordinate ring $A = A(Y)$ and let $\phi\colon X \to Y$ be a projective morphism. Then ϕ^* makes the ring of global regular functions $\Gamma(X, \mathcal{O}_X)$ on X a finitely generated module over A.*

Proof: We shall show that $\Gamma(X, \mathcal{O}_X)$ is an integral extension of A. It will then be a finite A module according to the general result that integral extensions of domains of finite type over a field are finite modules (see CA). We shall make use of the *hyperbola* $H \subseteq \mathbb{A}^1 \times \mathbb{A}^1$ given as the locus where $ut = 1$ with u and t being coordinates on $\mathbb{A}^1 \times \mathbb{A}^1 = \mathbb{A}^2$.

Pick a non-zero regular function f on X and consider the open set U where f does not vanish. The inverse function f^{-1} is regular on U and is in fact a morphism $f^{-1}\colon U \to \mathbb{A}^1$.

The graph $\Gamma \subseteq U \times \mathbb{A}^1 = \{\, (x, f(x)^{-1}) \mid x \in U \,\}$ of f^{-1} lies within $X \times \mathbb{A}^1$ and a crucial observation is that Γ is closed even in $X \times \mathbb{A}^1$ (as all graphs are closed, it is trivially closed in $U \times \mathbb{A}^1$). This hinges on Γ being the inverse image of the hyperbola H (which is closed) under the map

$$f \times \mathrm{id}_{\mathbb{A}^1}\colon X \times \mathbb{A}^1 \to \mathbb{A}^1 \times \mathbb{A}^1; \tag{10.2}$$

in other words, we have the equality $\Gamma = (f \times \mathrm{id}_{\mathbb{A}^1})^{-1}H$. Indeed, a point $(f(x), t)$ lies in the hyperbola H if and only if $f(x)t = 1$; that is, if and only if $t = f(x)^{-1}$. Note that in particular, $f(x) \neq 0$ so that $x \in U$, and furthermore that the second coordinate t never takes the value zero on H.

Consider then the map $\theta = \phi \times \mathrm{id}_{\mathbb{A}^1}\colon X \times \mathbb{A}^1 \to Y \times \mathbb{A}^1$. Points $(x, f(x)^{-1})$ in the graph Γ are send to points $(\phi(x), f(x)^{-1})$, so the second coordinate never vanishes. By the Fundamental Theorem of Elimination Theory (Corollary 10.21 above), the image $\theta(\Gamma)$ is closed and hence equals the zero set of an ideal $\mathfrak{a}$ in $A[t]$. Since $\theta(G) \cap Y \times \{0\} = \emptyset$, the Nullstellensatz gives a relation

$$1 = F(t) + tG(t)$$

with two polynomials F and G in $A[t]$. That θ sends Γ into $Z(\mathfrak{a})$ means that $P(1/f) = 0$ for all $P \in \mathfrak{a}$; in particular, it holds that $F(f^{-1}) = 0$, and bearing the relation (10.2) in mind, we find $1 = G(f^{-1})f^{-1}$. Multiplying through with a sufficiently high power of f we obtain an integral dependence for f over A. ❏

10.25 Theorem 10.24 opens the way for the following very practicable result of Claude Chevalley. A morphism $\phi\colon X \to Y$ is called *quasi-finite* if all the fibres

Quasi-finite morphisms
kvasiendelige morfier

are finite—morphism are often projective by nature and checking that fibres are finite, is a often low-hanging set-theoretical task.

THEOREM 10.26 *A projective and quasi-finite morphism $\phi\colon X \to Y$ is finite.*

PROOF: In view of the previous theorem, it will suffice to prove that each point y in Y has a neighbourhood U so that $\phi^{-1}(U)$ is affine. Since ϕ is projective, it factors as $\phi = \pi \circ \iota$ with $\iota\colon X \to Y \times \mathbb{P}^n$ being a closed embedding and $\pi\colon Y \times \mathbb{P}^n \to Y$ the projection, and identifying $\iota(X)$ and X, we may assume that X is a closed subset of $Y \times \mathbb{P}^n$. Since the fibre $\phi^{-1}(y)$ is finite, we may chose a hyperplane $H = Z(h) \subseteq \mathbb{P}^n$ disjoint from $\phi^{-1}(y)$. By Theorem 10.21 on page 199 the image $\pi(X \cap Y \times \mathbb{P}^n)$ is a closed subset of Y which by choice does not contain y, and so there is an open and affine neighbourhood U of y lying in the complement. Then $\phi^{-1}(U) = X \cap U \times \mathbb{P}^n = X \cap U \times D_+(h)$ is closed in the affine $U \times D_+(h)$, hence affine. ❑

EXAMPLE 10.27 Section 5.7, where the geometry of certain varieties of polynomials were studied, was based on a "multiplication map"

$$\phi : \mathbb{P}^1 \times \cdots \times \mathbb{P}^1 \to \mathbb{P}^d \tag{10.3}$$

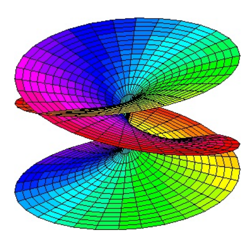
defined by the assignment

$$(u_1 : v_1) \times \cdots \times (u_r : v_r) \mapsto (u_1x + v_1)^{\lambda_1} \cdot \cdots \cdot (u_rx + v_r)^{\lambda_r}.$$

where $\lambda = (\lambda_1, \ldots, \lambda_r)$ is a partion of d. With Theorem 10.26 above, it is simple to verify that ϕ is a finite map. Obviously it is projective, and the fibres are all finite: the factorization of a polynomial in linear factors is unique up to order, so the maximal cardinality of a fibre is the number of permutation of d objects; that is, d! ☆

10.3 *Generically finite maps*

Those who have followed course in complex function theory and Riemann surfaces have certainly seen holomorphic maps depicted as a many sheeted covering of the plain like in the margin—which is typically how a finite map of complex curves look like near a branch point. Away from the branch points, they are locally just a bunch of stacked discs. Such iIlustrations presuppose we have small open sets (that is, disks) to our disposal. We do not have that in the Zariski topology, so these pictures must as usual be taken with a grain of salt (when working in positive characteristic, the grain ought to be rather large). Anyhow, some features are general. We notice that the number of points in most fibres are the same (three in picture), with a correct interpretation of how to count points in the fibre, this is generally true. It is called the degree of the map; and it may be defined for dominating maps whose generic fibres are finite; *e.g.* dominating maps between varieties of the same dimension.



There is however, a closed set of exceptional fibres, and how their cardinality relate to the cardinality of the generic fibre is unpredictable, only for very special maps are there general statements; one important instance is maps between projective non-singular curves. Morphisms between varieties of higher dimension, may have non-generic fibres with infinitely many points; a simple example is blow-up $\tilde{\mathbb{P}}^2 \to \mathbb{P}$ of $\mathbb{P}^2$ on one point p: all fibres are singletons except the one over p which is an entire projective line.

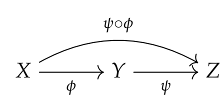
In characteristic zero the counting of the generic finite fibres is just the naive counting, but if the ground field has positive characteristic p, a genuine generic multiplicity may occur. The example to have in mind is $\mathbb{A}^1 \to \mathbb{A}^1$ sending x to x^p. It is polynomial map corresponding to the map $t \to t^p$. The equation $t^p = a$ has just one solution, and every fibre reduced to one point. However, the fibre multiplicity will be p—which is in concordance with the usual way to assign multiplicities to roots of polynomials.

The degree of a generically finite map

10.28 We start with a geometric set up with X and Y two varieties of the same dimension and $\phi\colon X \to Y$ a dominant morphism. As any dominant morphism does the morphism ϕ induces a field extension $k(Y) \subseteq k(X)$ which is algebraic since the transcendence degrees over k of the two fields are the same. Being function fields of varieties, both fields are finitely generated over k, and the extension is therefore finite. The degree $[k(X) : k(Y)]$ is called the *degree* of ϕ and denoted $\deg \phi$.

Degree of morphisms
graden til avbildninger

LEMMA 10.29 *Assume that ϕ and ψ are composable dominating morphisms between varieties of the same dimension. Then the composition $\psi \circ \phi$ is dominating and one has $\deg \psi \circ \phi = \deg \psi \cdot \deg \phi$.*



PROOF: We observed already in Paragraph 7.1 on page 144 that the composition of two dominant morphisms is dominant, so it merely remains to check the formula for the degrees. The field extensions of the function fields associated with the three involved maps constitute a tower of successive extensions:

$$k(Z) \subseteq k(Y) \subseteq k(X),$$

and since the degrees of field extensions are multiplicative in towers, it follows $[k(X) : k(Z)] = [k(X) : k(Y)][k(Y) : k(Z)]$; or in other words, that $\deg \psi \circ \phi = \deg \psi \cdot \deg \phi$. ❑

The heuristic geometric incarnation of the degree is "the number of points in the general fibre" which however is only viable in characteristic zero, the principle is not true for inseparable maps (though it holds for separable maps in positive characteristic). We shall shortly return to these questions.

EXAMPLE 10.30 *(Maps from $\mathbb{A}^1$ to $\mathbb{A}^1$)* The simplest example of the staging above is the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^1$ that sends the point t to $f(t)$ where $f(t)$ is a given polynomial of some degree n, say. In terms of a coordinate u on the target affine line the extension of function fields corresponding to ϕ is the extension $k(u) \subseteq k(t)$ where $u = f(t)$. It is generated by t and the minimal equation of t is $f(T) - u = 0$ (where T is a new variable) which is of degree n. So the (brand new) degree of the morphism agrees with the (good old) degree of the polynomial f.

The fibre of ϕ over a point a it is just given by the solutions of the equation $f(t) = a$. When f is separable, the derivative $f'(t)$ is a non-zero polynomial, and has finitely many roots, say $b_1, \ldots, b_s$. For points a not in the set $\{f(b_i)\}$ the derivative does not vanish at any of the solutions to $f(t) = a$, which therefore all are simple solutions. We conclude that for those a's there are exactly n-points in the fibre.

The inseparable case is more involved, and we merely illustrate it by the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^1$ that sends t to t^p, where p now is the characteristic of k. The equation $t^p = a$ has exactly one solution, since for any $b \in k$ the equation $t^p - b^p = (t-b)^p$ holds true. All fibres of ϕ are therefore reduced to a single point, but the degree of ϕ is p, so every points is counted with multiplicity p in its fibre. ☆

EXERCISE 10.5 Show that if $f(t)$ is an inseparable polynomial; that is, the derivative $f'(t)$ vanishes identically, then $f(t) = g(t^q)$ where $g(t)$ is a separable polynomial and q is a power of the characteristic p of the ground field k. Conclude that any morphism $\phi\colon \mathbb{A}^1 \to \mathbb{A}^1$ can be factored as $\phi = \psi \circ \mathrm{fr}$ where $g\colon \mathbb{A}^1 \to \mathbb{A}^1$ is separable and $\mathrm{fr}\colon \mathbb{A}^1 \to \mathbb{A}^1$ is the so-called *Frobenius map* given by $\mathrm{fr}(t) = t^q$. HINT: Confront lemma **??** on page **??** in CA. ★

Frobenius map
Frobenius avbildningen

Multiplicities

The ideal behaviour of a fibre of finite map $\phi\colon X \to Y$ is that the it has exactly as many points as the degree dictates, but of course, as examples show, this is certainly not true in general, already polynomial maps $\mathbb{A}^1 \to \mathbb{A}^1$ may have ramification points; *i.e.* points acquiring a multiplicity in the fibre. The next best behaviour is that an algebraic substitute, namely that the so-called "algebraic fibre" is of the right dimension; that is, equal to the degree of the map. But neither this holds generally true. The aim of this paragraph is to introduce the "algebraic fibre", and by that, the multiplicity of a fibre point.

10.31 In the currant paragraph we are concerned with the fibres of finite morphisms $\phi\colon X \to Y$, and since inverse images of open affines under a finite map are affine, we may as well assume that X and Y both are affine with coordinate rings $A(X)$ and $A(Y)$. Back in Chapter 7, in Lemma 2.58, we saw that the fibre of $\phi\colon X \to Y$ over a point $y \in Y$ consists of the points x so that $\phi^*\mathfrak{m}_y \subseteq \mathfrak{m}_x$, and for a lighter notation we let $\mathfrak{m} = \mathfrak{m}_y$ and write $\mathfrak{m}A(X)$ for $\phi^*\mathfrak{m}_y$. Then points belonging to the fibre over y correspond to maximal ideals in $A(X)$ containing $\mathfrak{m}A(X)$. Let $x_1, \ldots, x_r$ be these points and $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ the maximal ideals. The ring $A(X)/\mathfrak{m}A(X)$ is a finite algebra over $k(y)$, and hence is an Artinian ring. The primary decomposition of $\mathfrak{m}A(X)$ is shaped like

$$\mathfrak{m}A(X) = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r$$

with each $\mathfrak{q}_i$ being $\mathfrak{m}_i$-primary. Therefore, according to The Chinese Remainder Theorem, there is decomposition

$$A(X)/\mathfrak{m}A(X) = \prod_i A(X)/\mathfrak{q}_i.$$

The *multiplicity* of the point x_i in the fibre is then defined as $\mu(x_i) = \dim_k A(X)/\mathfrak{q}_i$ and we shall call the total dimension $\dim_k A(X)/\mathfrak{m}A(Y)$ the *algebraic cardinality* of the fibre. It is of course equal to the sum $\sum_{x \in \phi^{-1}(y)} \mu(x)$ over the fibre of the multiplicities.

Multiplicity of point in fibre
multiplisitet av punkt in fibre
The algebraic cardinality
den algebraiske kardinaliteten

EXAMPLE 10.32 It is worthwhile to compare this with Example 10.30 above where the map $\mathbb{A}^1 \to \mathbb{A}^1$ was defined by the assignment $t \mapsto f(t)$. On the level of coordinate rings it is incarnated as the homomorphism $k[t] \to k[t]$ that maps t to $f(t)$, so that a maximal ideal $(t-a)$ is transformed into the ideal $(f(t)-a)$. Now, the factorization of the polynomial $f(t)-a$ into linear factors is shaped like

$$f(t) - a = (t-b_1)^{n_1} \cdot \ldots \cdot (y-b_r)^{n_r},$$

where b_i's are the (distinct) points in the fibre over a and the n_i's natural numbers. Thus we find that

$$k[t]/(f(t)-a) = \prod_i k[t]/(t-b_i)^{n_i}.$$

Consequently the multiplicities are $\mu_p(b_i) = \dim_k k[t]/(t-b_i)^{n_i} = n_i$, and so we recuperate the multiplicities of the roots b_i in the polynomial $f(t)-a$. ☆

Examples

10.33 Consider the assignment $(x,y,z) \mapsto (x^2,y^2,z^2)$ which gives a morphism $\phi\colon \mathbb{A}^3 \to \mathbb{A}^3$. It is a finite map since $k[x,y,z]$ is generated over $k[x^2,y^2,z^2]$ by $1, x, y, z, xy, xz, yz$ and xyz; indeed, all monomials other than these must contain a square. So the degree will be eight (with a little effort, one shows the eight monomials are linearly independent over $k[x^2,y^2,z^2]$: use that every monomial is of shape P^2Q_i where Q_i is unique among the eight).

Assume first that k is not of characteristic two. Over points in the coordinate planes but not on any of the axes, the fibre has four points each of multiplicity 2. For instance, over the point $(0,1,1)$ one finds the four points $(0,\pm 1,\pm 1)$. Algebraically, it holds that

$$(x^2, y^2-1, z^2-1) = (x^2, y-1) \cap (x^2, y+1) \cap (x^2, z-1) \cap (x^2, z+1),$$

and the *algebraic fibre* will be the algebra $k[x]/x^2 \times k[x]/x^2 \times k[x]/x^2 \times k[x]/x^2$. Over points on the coordinate axes, the fibre has two points each of multiplicity 4, and over the origin lies only the origin, but with multiplicity eight.

It the ground field is of characteristic two, the situation changes dramatically. Then ϕ is bijective (in fact, a homeomorphism) but will still be of degree eight; $k[x,y,z]$ is still free of rank eight over $k[x^2,y^2,z^2]$ with the same basis, so every fibre will be of multiplicity eight. The fibre over $(0,1,1)$, for instance, reduces to $(0,1,1)$, and for the algebraic fibre we find $(x^2, y^2-1, x^2-1) = (x^2, (y-1)^2, (z-1)^2)$, which is $(x, y-1, z-1)$ primary.

10.34 Normalization maps have typically fibres with too many points. Our two old acquaintances, the cuspidal cubic and the rational double point, are good examples. Note that in contrast with for example maps $\mathbb{A}^1 \to \mathbb{A}^1$, the jumps in the fibre cardinality in theses cases are due to intrinsic geometric properties of the target of the map.

The cusp C is the image of the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^2$ defined by $x = t^2$ and $y = t^3$ and is given by the equation $y^2 = x^3$. On the level of rings the map $\phi^*\colon k[x,y] \to k[t]$ is given as $x \mapsto t^2$ and $y \mapsto t^3$. For a point $(a,b) \in C$ with $b^2 = a^3$, one finds when $a \neq 0$, that

$$(x-a, y-b)k[t] = (t^2-a, t^3-b) = (at-b) = (t-ba^{-1}).$$

The fibre over (a,b) is therefore just a single simple point, namely the point $b/a \in \mathbb{A}^1$. Over the origin, however, the algebraic fibre is $k[t]/(t^2,t^3) = k[t]/t^2$ which is two-dimensional. That fibre has also just one point, but with multiplicity two.

10.35 The double-point D is the image of $\mathbb{A}^1$ under the map $t \mapsto (t^2-1, t(t^2-1))$, and the equation of the image is $y^2 = x^2(x+1)$. To determine the fibres, let (a,b) be a point on D other than the origin. The equalities

$$(x-a, y-b)k[t] = (t^2-1-a, t(t^2-1)-b) = (at-b) = (t-b/a),$$

show there is a single simple point in the fibre. If $(a,b) = (0,0)$ however, on finds that algebraic fibre equals

$$k[t]/(t^2-1, t(t^2-1)) = k[t]/(t^2-1).$$

The structure of this ring depends on the characteristic of the ground field k. If the characteristic is different from two, it splits as the product of two copies of k, and the fibre over the origin consists of two distinct simple points. Is the characteristic two, however, the ring equals $k[t]/(t-1)^2$ which has just one maximal ideal. The fibre in that case consists of a single point but with multiplicity two.

☆

Exercises

10.6 Show that every morphism $\mathbb{A}^1 \to \mathbb{A}^1$ must have a ramification point; that is, a point of multiplicity at least two in its fibre.

10.7 Assume that the characteristic of k different from two and three. Consider the affine elliptic curve $y^2 - P(x)$ where $P(x)$ is a cubic polynomial with distinct roots. Determine the degree of the projections onto respectively the x-axis and the y-axis and find all ramification points (points with fibre multiplicity two or more).

10.8 Let $\phi\colon \mathbb{P}^2 \backslash (1:0:0) \to \mathbb{P}^3$ be the map that sends $(u:v:w)$ to $(uv : v^2 : w^2 : uw)$. Prove that image is contained in $Z_+(x_0^2 x_2 - x_3^2 x_1)$ and determine all fibres.

10.9 Assume that p and q are two relatively prime numbers. Let $C \subseteq \mathbb{A}^2$ be the image of the map $\phi\colon \mathbb{A}^1 \to \mathbb{A}^2$ given as $t \mapsto (t^p, t^q)$. Show that $C = Z(x^q - y^p)$. Prove that ϕ is a finite map and determine all fibres of ϕ.

10.10 Determine the fibres of the map $\mathbb{P}^2 \to \mathbb{P}^2$ given as $(x; y, z) \mapsto (x^6; y^6; z^6)$.

★

Generic freeness

As alluded to in the previous paragraph, there are examples without reduced fibres. However, even in that case, the dimension $\dim_k A(X)/\mathfrak{m}_y A(X)$ will be constant and equal the $\deg \phi$ for y in a sufficiently small open set (but of course dense). Generic fibres all have $\deg \phi$ points when appropriate multiplicities are taken into account; this ensues straight away from the following proposition. If the map ϕ is separable, one can say more. In that case, as we shall prove next, the generic fibres will be reduced; that is, all the points will be simple.

Proposition 10.36 (Generic freeness) *Let X and Y be two varieties and ϕ be a finite dominating map $\phi\colon X \to Y$. Then there is a dense open affine set U in Y such that $\phi^{-1}(U)$ is affine and $A(\phi^{-1}(U))$ is a free module over $A(U)$ of rank equal to $\deg \phi$.*

The proof is reduced to a piece of algebra contained in two subsequent lemmas.

Lemma 10.37 *Let $A \subseteq B$ be a finite extension of domains and let $K \subseteq L$ be the corresponding extension of fraction fields. Then the $K \subseteq L$ is a finite extension, and if S denotes the multiplicative set $S = A \backslash \{0\}$, it holds true that $L = S^{-1}B = B \otimes_A K$*

PROOF: The algebra B is a finite module over A and therefore all its elements are integral over A. Any given $f \in A$ thus satisfies an equation of integral dependence

$$f^n + a_{n-1}f^{n-1} + \ldots + a_1 f + a_0 = 0$$

with the constant term $a_0 \neq 0$. It follows that

$$f^{-1} = -a_0^{-1}(f^{n-1} + a_{\nu-1}f^{n-2} + \ldots + a_1),$$

and consequently that $f^{-1} \in B_S$. And since this holds for all $f \in B$, it ensues that $L = S^{-1}B_S$. In particular, any generating set of B as an A-module will be a generating set for L over K and L will be finite over K as B is finite over A. ❑

Recall that the dimension $\dim_K L$ of L as a vector space over K is called the *degree* of the field extension $K \subseteq L$ and that it is common usage to denote it by $[L : K]$.

Degree of field extensions
graden til kroppsutvidelser

LEMMA 10.38 *With the setting as in the previous lemma, there is an $g \in A$ such that $B_g = B[1/g] = B \otimes_A A_g$ is a free A_g-module of rank equal to the degree $[L : K]$.*

PROOF: This is just a matter of pinning down common denominators. From the previous lemma ensues that there is a basis for L over K of elements of the form $c_i a_i^{-1}$ with $c_i \in B$ and $a_i \in A$. Replacing the a_i's by their product, we may assume that they are all equal and that $c_i = b_i a^{-1}$. The c_i will not *a priori* generated B_a over A_a. They form however a basis for L over K and consequently all elements d_j from a finite generating set for B_a over A_a are sent into A when multiplied by appropriate elements g_j from A_a. Over the localized ring A_g, where g is the product of a and the g_j's, the elements c_i's form a basis for B_g. ❑

PROOF OF PROPOSITION 10.36: This is just a matter of translating Lemma 10.38 into geometry. We may assume that X and Y affine; just replaces Y by an open affine whose inverse image is affine and X by that inverse image. After Lemma 10.38 with $A = A(Y)$ and $B = A(X)$ we may find a regular function $g \in A(Y)$ so that the localization $A(X)_g = A(X)_{\phi \circ g}$ is free of rank $[k(X) : k(Y)]$ over $A(Y)_g$. Now, $A(Y)_g$ is the coordinate ring of the distinguished open set $D(g)$ in U, and obviously $\phi^{-1}(D(g)) = D(\phi \circ g)$ whose coordinate ring is $A(X)_{\phi \circ g}$, and that's it. ❑

Generic flatness

Combined with the structure theorem of morphisms, 10.3, results in the following theorem customary referred to as the Theorem of Generic Flatness (don't bother what this means).

Theorem 10.39 (Generic Flatness) *Let $\phi\colon X \to Y$ be a dominant morphism of varieties. Then there is an open affine U of Y and a finite covering $\{V_i\}$ of $\phi^{-1}(U)$ consisting of open affine subsets so that the coordinate rings $A(V_i)$ are finite free modules over $A(U)[t_1, \ldots, t_n]$ where the t_i's are elements algebraically independent over $A(U)$ and $n = \dim X - \dim Y$ denotes the relative dimension.*

Proof: Just couple Generic Freeness (Theorem 10.36 above) with the Structure Theorem (Theorem 10.3 on page 192). ❑

The separable case

As promised we shall take a closer look at the generic fibres of separable finite morphisms, which we shall prove are reduced.

Proposition 10.40 *Let X and Y be two varieties and ϕ be a finite dominating map $\phi\colon X \to Y$. If the field extension $k(Y) \subseteq k(X)$ is separable, there is an open affine $U \subseteq Y$ so that all fibres $\phi^{-1}(x)$ over points $x \in U$ have exactly $\deg \phi$ points; i.e. the multiplicity of each point in the fibre is one.*

The basic tool will be the Primitive Element Theorem which makes it easier to unveil the finer generic behaviour of finite morphisms based on the following lemma.

Recall that a field extension $K \subseteq L$ is called primitive if it has one single generator; that is $L = K(f)$ for an element f from K. The generator f is naturally called a *primitive element*. Every separable extension is primitive – this is the Primitive Element Theorem – and of course, every finite extension is a sequence of primitive ones.

Primitive element
primitive elementer

Lemma 10.41 *Let $\phi\colon X \to Y$ be a finite morphism between two varieties and assume that function field $k(X)$ is a primitive extension of $k(Y)$. Then there is an open affine U of Y such that $V = \phi^{-1}(U)$ is affine and such that*

$$A(V) \simeq A(U)[t]/(F(t))$$

for some monic polynomial $F(t)$ whose coefficient are regular functions on U.

Proof: Let f be the element in $k(X)$ that generates $k(X)$ over $k(Y)$. The field $k(X)$ is algebraic over $k(Y)$, so f satisfies an algebraic dependence equation

$$a_n f^n + a_{n-1} f^{n-1} + \ldots + a_0 = 0,$$

where the a_i's are rational functions on Y, and we may suppose that $F(t) = \sum_i a_i t^i$ is an irreducible polynomial over $k(Y)$. Now, we may find an open affine U in Y where the coefficients a_i all are regular and where a_n is without zeros. Since ϕ is finite the preimage $V = \phi^{-1}(U)$ is affine, and the ring $A(V)$ is a finite module over $A(U)$. It has a generating set whose members are linear combinations of powers of f with coefficients in $k(Y)$, and shrinking U further, we may assume that all these coefficients are regular in U. Then clearly

$$A(V) = A(U)(f) \simeq A(U)[t]/(F(t)).$$

with $F(t) = T^n + a_{n-1}a_n^{-1}T^{n-1} + \ldots + a_0 a_n^{-1}$. ❑

10.42 The isomorphism in the lemma is defined by sending the variable t to the function f. One may interpret this in a geometric way as V lying in the product $U \times \mathbb{A}^1$ embedded as the zero locus $Z(F(t))$, and the map ϕ is induced by the projection onto the first factor; remember that the coefficients of F are functions on U, so $F(y,t)$ would be a more precise notation. For a given point $y_0 \in U$, the fibre over y_0 is formed by the zeros of the polynomial $F(y_0,t)$, obtained by evaluating the coefficients a_i at the point y_0.

PROOF OF PROPOSITION 10.40: The derivative of $F'(y,t)$ with respect to t is as usual computed as $\sum_i i a_i t^{i-1}$. The hypothesis that f be separable ensures that the derivative $F'(t)$ share no common zero with $F(t)$ in $k(X)$. This shows that $Z(F'(y,t)) \cap Z(F(y,t))$ is a proper subset of $Z(F(y,t))$. Hence it does not dominate U, and the image D is a proper closed set of U (finite maps are closed according to Lying-Over, Proposition 7.11 on page 146). The points of the fibre over y_0 are the zeros of the polynomial $F(y_0,t)$ but when $y_0 \notin D$, the derivative $F'(y_0,t)$ does not vanish in any of the points, and they are all simple zeros. Consequently there are as many as the $\deg \phi$ indicates. ❑

EXAMPLE 10.43 Consider the map $\mathbb{A}^1$ to $\mathbb{A}^1$ sending x to x^p. On the level of coordinate rings it is given as $f(t) \mapsto f(t^p)$. Ir is finite og degree p, but all its fibres consist of just one point. $k[t]/(t^p - a)k[t]$. But $(t^p - a) = (t-b)^p$ so $(t-a)$ is the only maximal ideal containing $t^p - a$. ☆

EXERCISE 10.11 Given natural numbers n and let m with $m \leqslant n$ Construct a finite map of degree n having exactly m points in one of its fibres. ★

10.4 Curves over regular curves

When the target of a morphism is a non-singular curve, more can be said of its fibres. We illustrate this with finite maps, and the result will be useful when we attack the proof of Bezout's theorem in next section. The staging will be as follows. The givens are a regular curve C and a closed subset $Z \subseteq C \times \mathbb{P}^n$ whose components $Z_1, \ldots, Z_r$ are curves that all dominates C.

THEOREM 10.44 *Let C and D be irreducible curves with C non-singular, and let $\pi\colon D \to C$ be a projective morphism. Then all fibres of π has the same algebraic cardinality.*

When both curves C and D are projective, any morphism between them will be projective, and slightly more general, the restriction of a projective map π to open subsets of the form $\pi^{-1}(U)$ persists being projective. The statement is false if C is singular, the parameterization $\pi(t) = (t(t^2-1), t^2-1)$ of the simple double point is birational, so the generic fibres are singletons (also algebraically and even in characteristsic two), but over the singular point $(0,0)$ there lie the two points $t = \pm 1$ (which coalesce in characteristic two to a singleton, but with multiplicity two).

PROOF: Since D dominates C all the fibres of π are finite, and by 10.26 , projecive morphisms with finite fibres are finite , and we may thus cover C by open affines U_i so that $V_i = \pi^{-1}U_i$ are affine and each $A(V_i)$ is a finite module over $A(U_i)$. By the Proposition below, the algebraic fibre dimension will be constant within each U_i, but C being irreducible, any two U_i's will meet. ❑

Except for closed algebraic subsets of affine space, we have developed the general theory for varieties which all have been assumed to be irreducible. This has certainly made life agreeable, but we now have come to point where there is a price for this, and admittedly some inelegant wriggling will be necessary.

LEMMA 10.45 (PERMANENCE OF NUMBERS) *Let Z be a closed algebraic set whose irreducible components are $Z_1, \ldots, Z_r$. Let C be a non-singular affine curve and assume that $\pi\colon Z \to C$ is a finite morphism whose restrictions π_{Z_i} all are dominating. Then all fibres of π have the same algebraic cardinality; that is, $\dim_k A(Z)/\mathfrak{m}_x A(Z)$ is independent of the point $x \in C$.*

PROOF: The main point is that $A(Z)$ is a torsion free $A(C)$-module, and by assumption it is finitely generated. Now $A(C)$ is a Dedekind ring because C is non-singular, and it is a general fact that finitely generated modules over Dedekind rings are projective, and hence $A(Z) \otimes_{A(C)} k(x)$ has the same rank for all $x \in C$. (See *e.g.* Theorem **??** on page, **??** in CA)

To see that $A(Z)$ torsion free, observe that if $\mathfrak{p}_i$ be the ideals of the Z_i's in Z, it holds that $\bigcap_i \mathfrak{p}_i = 0$, and we have an inclusion

$$A(Z) \subseteq \prod_i A(Z_i).$$

Since each Z_i is assumed to dominate C, the natural maps $A(C) \to A(Z_i)$ are injective, but the $A(Z_i)$'s being integral domains, they are torsion free. ❑

Example: The divisor of rational functions on non-singular curves

Let a a non-singular curve C and a rational function $f \in k(C)$ be given. The maximal set U_f where f si regular into $\mathbb{A}^1$, and by xxx f extends to a morphism $C \to \mathbb{P}^1$ All the local rings $\mathcal{O}_{C,x}$ are DVR's. Denote the normalized valuation of $\mathcal{O}_{C,x}$ by v_x so that f has a local order $v_x(f)$ at each point x. It will be zero everywhere but at a finite number of points: if f is defined and non zero at a point, it holds that $v_x(f) = 0$, so only at points in the fibres $f^{-1}(0)$ and $f^{-1}(\infty)$ does f have non-zero order.

LEMMA 10.46 *The algebraic multiplicity of x as a point in the fibre $f^{-1}(0)$ respectively $f^{-1}(\infty)$ equals $v_x(f)$ respectively $-v_x(f)$.*

PROOF: Since f is finite, the preimage $f^{-1}(\mathbb{A}^1)$ is affine, say with coordinate ring A. Over $\mathbb{A}^1$ the algebraic incarnation of f is the algebra-map $k[t] \to A$ that sends t to f. Hence $\mathfrak{m}_0 A = (f)A$ and $A/\mathfrak{m}_0 A = A/(f)A = \mathcal{O}_{C,x}/(f)\mathcal{O}_{C,x}$ whose dimension equals $v_x(f)$. A similar argument, but with t replaced by $1/t$, takes care of points in the fibre over ∞. ❑

It follows that $\sum_{f(x)=0} v_x(f)$ and that $-\sum_{f(x)=\infty} v_x(f)$ are the algebraic cardinalities of the fibres $f^{-1}(0)$ and $f^{-1}(\infty)$, but these are equal. Thus we have established the following:

THEOREM 10.47 *Let C be a projective and non-singular curve and f a rational function on C. Then $\sum_{x \in C} v_x(f) = 0$.*

Chapter 11

Bézout's theorem

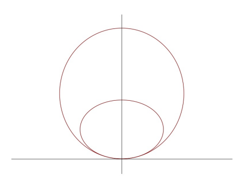
Topics in Chapter 11: *Divisors – Local multiplicities – Bézout's theorem – Pascal's and Pappus' theorems– Regular sequences and depth – Cohen–Macaulay rings – The Unmixedness theorem –*

Sir Isaac Newton observed in a note dated May 30th 1665, that the number of intersection points of two plane curves in equals the product of their degrees.

If one starts looking at examples, this pattern emerges almost immediately. Two lines meet in one point and two conics in four – at least if the two conics are in what one calls general position; that is, they are not tangent at the intersection points.

Étienne Bézout (1730–1783) French Mathematician

Of course, there are parallel lines and lines not meeting a given circle so the curves must be considered in the projective plane (then parallel lines meet at infinity) and the ground field must be algebraically closed (every line meets the circle in points with possibly complex coordinates). That a line L in $\mathbb{P}^2$ meets a curve X in as many points as the degree of the homogeneous form $F(x_0, x_1, x_2)$ defining the curve, is a direct consequence of the Fundamental Theorem of Algebra. Choosing appropriate coordinates we can parameterize the line as $(u : v : 0)$. The parameter values of the intersection points will be the roots of the equation $F(u, v, 0) = 0$, of which there are $\deg F$, unless, of course, x_2 is a factor of F, in which case the line L is a component of X. There is also an issue of multiplicities, roots need not be simple, and to get $\deg F$ intersection points these multiplicities must be taken into account. This issue persists in the general situation and is an inherent part of the problem.



Two ellipses with fourth order contact.

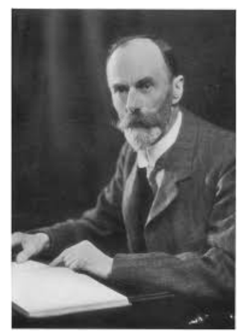
Example 11.1 The local multiplicity is, as this example shows, a quite subtle invariant even for conics. The two conics $zy - x^2$ and $zy - x^2 - y^2$ only intersect at $(0 : 0 : 1)$. Indeed, the difference of the two equations being y^2, it must hold that $y = 0$ at a common zero, and then x must vanish there as well. The ellipses have contact order four at $(0 : 0 : 1)$: inserting the parameterization (uv, u^2, v^2) of the first into the equation of the second yields the equation $u^4 = 0$, which has a quadruple root at $u = 0$. ☆

Exercise 11.1 Along the lines above, prove that a conic intersects a curve of degree n in $2n$ points multiplicities taken into account unless the conic is a component of the curve. HINT: Parameterize the conic as $(u^2 : uv : v^2)$. ★

11.2 What nowadays is called Bézout's theorem in the plane was, as we have indicated, known long time before Bézout published his famous paper *Théorie générale des équations algébriques* in 1779. His original contribution is the generalization to projective n-space $\mathbb{P}^n$. He asserted that the number of points n hypersurfaces in $\mathbb{P}^n$ have in common, when finite, is at most the product of the degrees of the hypersurfaces, and that equality holds when the hypersurfaces are general; that is, when they meet transversally. As usual there is an issue of multiplicities. Local multiplicities are part of the accounting, and with the correct definition of these multiplicities, the number of intersection points, when finite, will always be the product of the degrees. It seems that the first correct proofs of the full Bézout theorem were given by Georges-Henri Halphen and a little later by Adolf Hurwitz. Needless to say, Bézout's theorem marks the beginning of the vast subject called intersection theory, which lies at the hart of modern algebraic geometry. See *e.g.* Fulton's comprehensive book 'Intersection Theory'.

Adolf Hurwitz (1859–1919) German Mathematician

11.3 To prove Bézout's Theorem there are several possible lines of reasoning to follow. The classical technique used by Bézout and his contemporaries was by projections. They projected the intersection of the n hypersurfaces into a line where it was described by the vanishing of a certain polynomial, the so-called *resultant*. We fleetingly met a specimen of the kind in Exercise 10.4 on page 196. The proof we are about to present utilizes methods from commutative algebra. In particular, the proof involves properties of so-called *regular sequences*, and for the proof to work, the forms $F_1, \ldots, F_n$ defining the hypersurfaces must form such a regular sequence (see Paragraph 11.5 below). This is *a priori* a far stronger assumption than the locus of their common zeros being finite, but then the Macaulay's *Unmixedness Theorem* enters the scene. It asserts that whenever $Z_+(F_1, \ldots, F_n)$ is finite, the F_i's in fact form a regular sequence. Thus the algebraic condition that the hypersurfaces form a regular sequence (hard to check) is reduced to the geometric condition that the intersection be finite (substantially easier to check).



Francis Sowerby Macaulay (1862–1937) British Mathematician

11.1 *Bézout's Theorem*

Georges-Henri Halphen (1859–1919) French Mathematician

11.4 We fix the following set-up for the rest of the chapter. We are given n hypersurfaces $Z_1, \ldots, Z_n$ in $\mathbb{P}^n$ defined by the vanishing of the homogeneous polynomials $F_1, \ldots, F_n$. There will be a standing hypothesis that they intersect in finitely many points.

With every point $p \in \mathbb{P}^n$ we shall associate a multiplicity $\mu_p(Z_1, \ldots, Z_n)$ (see Section 11.8). This is a non-negative integer which is positive if and only if p belongs to the intersection $Z_1 \cap \ldots \cap Z_n$. With this in place, Bézout's theorem reads as follows

Theorem 11.5 (Bezout's theorem) *Let $Z_1, \ldots, Z_n$ be hypersurfaces in $\mathbb{P}^n$ with only finitely many points in common. Then*

$$\deg Z_1 \cdots \deg Z_n = \sum_p \mu_p(Z_1, \ldots, Z_n).$$

Notice that one can only hope for such a result when the number of hypersurfaces is n. If there are less, the intersection cannot be finite; indeed, by Krull's Hauptidalsatz the codimension of the intersection would be less than n and the dimension at least one, and the intersection would have an infinity of points. And if there are more hypersurfaces, we have little control on the number of intersection points, although there is an upper bound and the intersection will be empty for a general choice of hypersurfaces.

Exercise 11.2 Give examples of three conics in $\mathbb{P}^2$ that intersect in 0, 1, 2, 3 and 4 points. ★

11.6 It is quite natural to extend the scope of Bezout's theorem slightly to also encompass intersections of *effective divisors*[1]. Such an animal is a *formal* finite linear combination $\sum_i m_i Y_i$ of irreducible hypersurfaces Y_i with non-negative integral coefficients. This might look enigmatic at the first encounter, but it is merely a convenient and geometrically suggestive way to keep track of the irreducible components of the zero-locus of a homogeneous form. Indeed, if F is a homogeneous form of degree n that splits as $F = \prod_i F_i^{m_i}$ into a product of irreducible forms, the associate divisor is $\sum_i m_i Z_+(F_i)$.

Effective divisors
effektive divisorer

[1] The significance of the attribute *effective* is that the coefficients n_i are non-negative. A *divisor* is a linear combination $\sum_i n_i Z_i$ with integral coefficients.

The notion of divisors generalizes well. On any variety X a divisor is a formal finite linear combination $\sum_i n_i Y_i$ of irreducible subvarieties of codimension one; that is, an element in the free abelian group generated by the codimension one irreducible subvarieties of X.

The *degree* of a divisor $\sum_i m_i Y_i$ is defined to be the sum $\sum_i m_i$ which is nothing but the degree of the corresponding homogeneous form. Divisors can be added just by adding coefficients, and one clearly has $\deg(Z + Z') = \deg Z + \deg Z'$. The definition of the local multiplicities extends to effective divisors, and in terms of them Bézout's Theorem takes the form

Degree of divisors
graden til divisorer

Theorem 11.7 *Let $Z_1, \ldots, Z_n$ be effective divors in $\mathbb{P}^n$ with only finitely many points in common. Then*

$$\deg Z_1 \cdots \deg Z_n = \sum_p \mu_p(Z_1, \ldots, Z_n).$$

11.2 *The local intersection multiplicity*

We continue working with the given hypersurfaces $Z_1, \ldots, Z_n$ and their homogeneous equations $F_1, \ldots, F_n$. The F_i's need not be irreducible and can even have factors with exponents higher than one, but there is a standing assumption that the intersection $Z_1 \cap \ldots \cap Z_n$ is finite.

11.8 To set the stage, we choose homogeneous coordinates $(x_0 : \ldots : x_n)$ on $\mathbb{P}^n$ so that the intersection $Z_1 \cap \ldots \cap Z_n$ is contained in the distinguished open set $U = D_+(x_0)$. One may in fact use any open affine containing the intersection, but for the presentation it is convenient to use one of the standard affine ones.

The set U is an affine space $\mathbb{A}^n$ and the coordinates we shall use will be $t_i = x_i/x_0$. The equations of the subvarieties $Z_i \cap U$ of U are the dehomogenized polynomials $f_i(t_1, \ldots, t_n) = F_i(1, x_1/x_0, \ldots, x_n/x_0)$. They live in the coordinate ring $A(U) = k[t_1, \ldots, t_n]$.

In this setting, the intersection $Z_1 \cap \ldots \cap Z_n$ equals the closed algebraic subset $Z(f_1, \ldots, f_n)$ of $D_+(x_0)$ of $\mathbb{A}^n$, and by the standing hypothesis is a finite set. Consequently, the ring[2]

$$\mathcal{O}_{Z_1 \cap \ldots \cap Z_n} = k[t_1, \ldots, t_n]/(f_1, \ldots, f_n)$$

is Artinian. As any Artinian ring, it is isomorphic to the direct product of its localizations; that is, one has a natural isomorphism

$$\mathcal{O}_{Z_1 \cap \ldots \cap Z_n} \simeq \prod_p \mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p} \qquad (11.1)$$

where the product extends over points p from the intersection $Z_1 \cap \ldots \cap Z_n$. This is just the structure theorem for modules of finite length (Proposition **??** on page **??** in CA). Moreover each factor $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p}$ is given as

$$\mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p} = \mathcal{O}_{\mathbb{A}^n, p}/(f_1, \ldots, f_n)\mathcal{O}_{\mathbb{A}^n, p}$$

where $\mathcal{O}_{\mathbb{A}^n, p} = k[t_1, \ldots, t_n]_{\mathfrak{m}_p}$ as usual denotes the local ring of $\mathbb{A}^n$ at the point p. This leads us to the definition of the *intersection multiplicity* at p, also called the *local intersection number* at p. The local rings $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p}$ are all finite dimensional vector spaces over the ground field k, and we define

The intersection multiplicity
snittmultiplisitet
Local intersection numbers
lokale snitt-tall

$$\mu_p(Z_1, \ldots, Z_n) = \dim_k \mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p}.$$

If p does not belong to the intersection, we let $\mu_p(Z_1, \ldots, Z_n) = 0$.

Examples

The higher multiplicities are caused by two different phenomena, tangency and singularity. The hypersurfaces involved can be tangent at the intersection point, or one or more of them can have a singularity there; that is, the defining form

[2] The notation might be slightly misleading, but is standard when working with schemes. The ring $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n}$ is not the ring of functions on the intersection $Z_1 \cap \ldots \cap Z_n$; it could certainly have nilpotent elements for instance.

vanishes to the second order at the point. If neither of the two phenomena occur, the contribution of the point is just one, the intersection is said to be *transversal* at the point.

Transversal intersections
transversale snitt

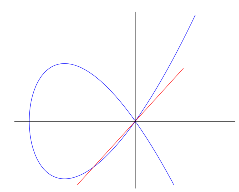
11.9 *(A parabola and a tangent)* The simplest example of tangency is the standard parabola and the x-axis. Consider the two curves $C = Z(f)$ and $D = Z(g)$ in $\mathbb{A}^2$ given by $f = y - x^2$ and $g = y$. Clearly, $(f, g) = (y - x^2, y) = (x^2, y)$. At the origin $p = (0, 0)$, we thus have

$$\begin{aligned}\dim_k \mathcal{O}_{\mathbb{A}^2,p}/(f,g) &= \dim_k \big(k[x,y]/(x^2,y)\big)_{(x,y)}\\ &= \dim_k \big(k[x]/(x^2)\big)_{(x)} = \dim_k k[x]/(x^2) = 2\end{aligned}$$

Hence the local intersection multiplicity is $\mu_p(C, D) = 2$.

11.10 *(A nodal cubic and a line through the node)* As an example of a singularity, consider the nodal cubic C with equation $f = y^2 - x^2(x+1)$ and a line L with equation $g = y - ax$. Then, with p denoting the origin, $\mu_p(C, L) = 3$ when $a = \pm 1$ and $\mu_p(C, L) = 2$ else: we find $(y^2 - x^2(x+1), y - ax) = (y - ax, x^2(a^2 - 1 - x))$, and hence

$$\dim_k \mathcal{O}_{\mathbb{A}^2,p}/(f,g) = \dim_k k[x]/\big(x^2(a^2-1-x)\big) = \begin{cases} 2 & \text{when } a \neq \pm 1;\\ 3 & \text{when } a = \pm 1.\end{cases}$$



A nodal cubic and a line through the origin.

When the parameter $a = \pm 1$, the line L is tangent to C and the two phenomena occur simultaneously, which makes the intersection multiplicity increase.

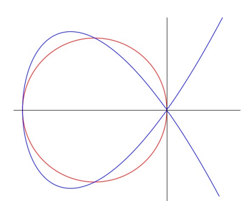
When $a \neq \pm 1$, there is also a second intersection point $q = (1 - a^2, a(1 - a^2))$. Observe that x^2 is invertible in the local ring $\mathcal{O}_{\mathbb{A}^2,q}$ and therefore

$$(f,g)\mathcal{O}_{\mathbb{A}^2,q} = \big(y - ax, x - (1-a^2)\big)\mathcal{O}_{\mathbb{A}^2,q} = \big(y - a(1-a^2), x - (1-a^2)\big)\mathcal{O}_{\mathbb{A}^2,q}$$

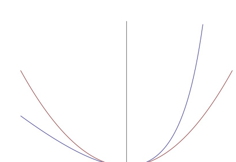
which is just the maximal ideal at q. Hence $\mu_q(C, L) = 1$.

11.11 *(A cuspidal tangent)* The cuspidal cubic $y^2 = x^3$ is singular at the origin and has the x-axis as tangent, and their intersection number at the origin equals three: we have $(y^2 - x^3, y) = (y, x^3)$, so that $k[x,y]/(y^2 - x^3, y) = k[x]/(x^3)$ which is three-dimensional.

11.12 In the margin we have depicted two curves, the circle $(x+1)^2 + y^2 = 1$ and the cubic $y^2 = x^2(x+2)$. They intersect in four points. At the leftmost point there is a tangency and the multiplicity is two, then come two points where the intersection is transversal each contributing one to the total, and finally in the rightmost point the cubic aquires a double point, and the local multiplicity is two.



11.13 *(Two parabolas with high contact)* Let $f(x,y) = y - x^2$ and $g(x,y) = y - x^2 - xy$. Then the two parabolas $X = Z(f)$ and $Y = Z(g)$ have triple contact at the origin; that is, $\mu_p(X, Y) = 3$. One finds

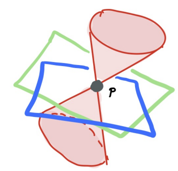


Two parabolas with triple contact at the origin.

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

and hence $\mathcal{O}_{X\cap Y,p} = \big(k[x,y]/(f,g)\big)_{(x,y)} \simeq k[x]/(x^3)$.

11.14 *(A quadratic cone and two planes)* Consider the three polynomials $f_1 = xy - z^2$, $f_2 = x - z$ and $f_3 = x + y$ and let $Z_i = Z(f_i)$ be their zero-loci in $\mathbb{A}^3$. Let p denote the origin and assume that k is not of characteristic two. Then the intersection multiplicity $\mu_p(Z_1, Z_2, Z_3)$ equals two; indeed, we have the equality $(xy - z^2, x - z, x + y) = (-2x^2, x - z, x + y)$ so that



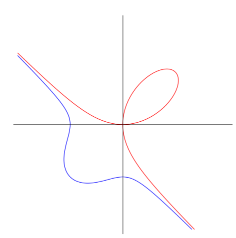
$$\begin{aligned}\dim_k \mathcal{O}_{\mathbb{A}^3,p}/(f_1, f_2, f_3) &= \dim_k \big(k[x,y,z]/(xy - z^2, x - z, x + y)\big)_{(x,y,z)} \\ &= \dim_k k[x]/(x^2) = 2.\end{aligned}$$

Hence $\mu_p(Z_1, Z_2, Z_3) = 2$. The locus $Z(x - z, x + y)$ is a line (parameterized by $t \mapsto (t, -t, t)$) which meets the surfaces $Z_1 = Z(xy - z^2)$ only at the origin. There the surface has a double point, and this explains the multiplicity. When the ground field k is of characteristic two, the line is entirely contained in $Z(xy - z^2)$, so in that case the local multiplicity is infinite.

11.15 Let us do a more involved example and determine the points of intersections and the local intersection multiplicities of the two cubic curves C and D whose equations are $F = (x^2 + y^2)z + x^3 + y^3$ and $G = x^3 + y^3 - 2xyz$. It is meant as an illustration of the working of local intersections numbers, but with some machinery at hand, it can be done easier. We assume that the characteristic of k is not 2 or 3, and we let η be a primitive cube root of -1.

Their real points in $D_+(z)$ are depicted in margin, C in blue and D in red, but be suspicious! The picture is an example of how deceiving real pictures can be. The blue curve has an acnode; *i.e.* an isolated node, at the origin, which is invisible in the picture, and the curves meet at infinity in the complex conjugate points $(\eta : 1 : 0)$ and $(\eta^{-1} : 1 : 0)$ in addition to $(-1 : 1 : 0)$. The latter is manifested in the picture as the common asymptote $y + x = 0$, but the two former are not visible.

In the affine piece $D_+(z)$, where $z \neq 0$, we have with f and g the dehomogenized polynomials $F(x,y,1)$ and $G(x,y,1)$ the following equality in $k[x,y]$:



The two cubics

$$\begin{aligned}\mathfrak{a} = (f,g) &= \big(x^2 + y^2 + x^3 + y^3, x^3 + y^3 - 2xy\big) = \big(x^2 + y^2 + 2xy, x^3 + y^3 - 2xy\big) \\ &= \big((x+y)^2, x^3 + y^3 - 2xy)\big).\end{aligned}$$

Using that $x^3 + y^3 = (x + y)((x + y)^2 - 3xy)$, we find

$$\mathfrak{a} = \big((x+y)^2, xy(3(x+y) - 2)\big).$$

One easily sees that $Z(\mathfrak{a})$ is the origin (denoted by p). Near the origin $3(x + y) - 2$ is invertible, and we conclude that

$$\mu_p(C, D) = \dim_k \mathcal{O}_{C\cap D,p} = \dim_k k[x,y]/\big((x+y)^2, xy\big) = 4,$$

since 1, x, y, x^2 form a basis for the last algebra.

We already established that $C \cap D$ has three points in the affine piece $D_+(y)$, namely the points $q = (-1 : 1 : 0)$, $r_1 = (\eta : 1 : 0)$ and $r_2 = (\eta^{-1} : 1 : 0)$. With f and g this time being $F(x, 1, z)$ and $G(x, 1, z)$ the equality

$$\mathfrak{a} = (f, g) = \big((x+1)^2 z, x^3 + 1 - 2xz\big)$$

holds in the ring $k[x, z]$. Near the point $q = (-1 : 1 : 0)$ both $x^2 - x + y^2$ and $2x$ are invertible, and so writing $u = x + 1$, we arrive at $\mathfrak{a}\mathcal{O}_{\mathbb{A}^2,q} = (u^2 z, z - cu)$ with c a unit, and consequently it holds true that

$$\mu_q(C, D) = \dim \mathcal{O}_{C \cap D, q} = \dim_k k[x, z]/u^3 = 3.$$

Near the point $r_1 = (\eta : 1 : 0)$ it holds that $(x+1)^2$ is invertible, and therefore $\mathfrak{a}\mathcal{O}_{A^2,q} = \big(z, (x+\eta)(x^2 + (\eta^{-1}+1)x + \eta^{-1})\big) = (z, x+\eta)$. Hence the intersection multiplicity equals one. By symmetry, the same reasoning works for the point $r_2 = (\eta^{-1} : 1 : 0)$, so we may conclude that

$$\mu_{r_1}(C, D) = \mu_{r_2}(C, D) = 1.$$

Summing up, there are four intersection points; one with multiplicity 4, one with 3 and two with 1. This adds up to 9 which is precisely what Bézout's theorem predicts.

☆

Exercises

11.3 Find the intersection and the local multiplicities of the three surfaces in $\mathbb{P}^3$ given by $xy - zw$, $xz - yw$ and $xw - yz$.

11.4 Prove that $xy - zw$ and $x^2 y - z^2 x$ intersect along five lines. Find the intersection of $y - x$, $xy - zw$ and $x^2 y - z^2 x$.

11.5 With the set up from Example 11.14, show that any line through the origin that is not contained in $Z(xy - z^2)$, meets $Z(xy - z^2)$ with multiplicity two there.

11.6 Fill in the details in Example 11.12 above.

★

Transversal intersections

11.16 Recall that a hypersurface X in $\mathbb{A}^n$ given by the polynomial f which passes through the point p, is *regular* or *non-singular* at p if f does not vanish to the second order there. In other words, if $\mathfrak{m}$ denotes the maximal ideal at p, the polynomial f is required not to belong to the square $\mathfrak{m}^2$; that is, $f \notin \mathfrak{m}^2$.

Non-singular or regular points
ikkesingulære eller regulære punkter

Recall also that we defined the *differential* df of f as the class of f in the cotangent space $\mathfrak{m}/\mathfrak{m}^2$. Thus $Z(f)$ is non-singular at p if and only if $df \neq 0$ in $\mathfrak{m}/\mathfrak{m}^2$.

It is common usage to say that r hypersurfaces $Z_1, \ldots, Z_r$ with equations $f_1, \ldots, f_r$ meet *transversally* at the point p if *(i)* they are non-singular at p; and *(ii)* the differentials $df_1, \ldots df_r \in \mathfrak{m}/\mathfrak{m}^2$ are linearly independent. We already used the term *transversal intersection* when the local intersection multiplicity of the hypersurfaces equals one, and of course, the two interpretations coincide, which is the content of the next lemma. It is almost tautological, and states that the differentials of the f_i's are linearly independent if and only if $\mu_p(Z_1, \ldots, Z_n) = 1$.

Transversal hypersurfaces
transversale hyperflater

LEMMA 11.17 *Let $A = k[t_1, \ldots, t_n]_{(t_1,\ldots,t_n)} = \mathcal{O}_{\mathbb{A}^n,\mathfrak{m}}$ where $\mathfrak{m} = (t_1, \ldots, t_n)$ is the maximal ideal at the origin, and let $f_1, \ldots, f_n \in \mathfrak{m}$. Then the following are equivalent:*

i) *$f_1, \ldots, f_n$ meet transversely at $p = (0, \ldots, 0)$;*
ii) *$(f_1, \ldots, f_n) = \mathfrak{m}$;*
iii) *$\mu_p(Z_1, \ldots, Z_n) = 1$, where $Z_i = Z(f_i)$ for $i = 1, \ldots, n$.*

PROOF: By Nakayama's lemma, the differentials $df_1, \ldots, df_n$ generate $\mathfrak{m}/\mathfrak{m}^2$ if and only if $f_1, \ldots, f_n$ generate $\mathfrak{m}$. And $df_1, \ldots, df_n$ generate if and only if they are linearly independent, so $(i) \Leftrightarrow (ii)$. The implication $(ii) \Leftrightarrow (iii)$ follows immediately from the definition of μ_p. ❑

Additivity

The local intersection number is additive in the sense that if one of the divisors splits into as sum, say that the last one decomposes as $Z_n = Z_n' + Z_n''$, the following addition formula holds true

$$\mu_p(Z_1, \ldots, Z_n' + Z_n'') = \mu_p(Z_1, .., Z_n') + \mu_p(Z_1, \ldots, Z_n'').$$

However it is astonishingly subtle to prove and hinges on the Unmixedness Theorem of Macaulay.

Let f_n and g_n be the equations of Z_n' and Z_n'', and let A be the local ring $A = \big(k[t_1, \ldots, t_n]/(f_1, \ldots, f_{n-1})\big)_{\mathfrak{m}_p}$. One has the exact sequence

$$A/(g_n)A \xrightarrow{\alpha} A/(f_n g_n)A \longrightarrow A/(f_n)A \longrightarrow 0$$

of Artinian algebras where α is just multiplication by f_n; that is, $\alpha(a) = f_n \cdot a$, and the rightmost map is the natural surjection. The sequence is short exact –

that is, α is injective – precisely when the vector space dimensions over k of the involved algebras add up; in other words, when the local intersection numbers add up.

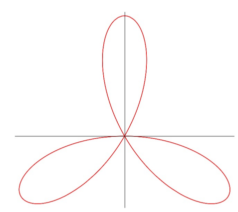
However, α being injective requires that f_n be a non-zero divisor in A so that $f_n a = b f_n g_n$ implies that $a = b g_n$. This is not generally true for one-dimensional local rings even if neither of f_n and g_n lies in any of the minimal primes of A. Luckily Macaulay's Unmixedness theorem saves the day: There are no embedded components, and additivity holds true.

Plane curves: multiplicities and intersections

11.18 Let us study more closely the intersection of two curves C and D at a point p in $\mathbb{P}^2$, and we shall relate the local intersection number $\mu_p(C, D)$ to the multiplicities of the curves at p. If the equation of the curve C near p is f, the *multiplicity* $m_p(C)$ is the smallest integer ν so that $f \in \mathfrak{m}^\nu$ where $\mathfrak{m} \subseteq \mathcal{O}_{\mathbb{P}^2,p}$ is the maximal ideal in the local ring of $\mathbb{P}^2$ at the point.

Multiplicity of a point on a curve
multiplisiteten til et punkt på en kurve

Choose a distinguished affine open neighbourhood of p which is an affine space $\mathbb{A}^2$ equipped with coordinates x and y so that p corresponds to the origin, and develop f in its homogeneous components


The trifulium

$$f = f_m + f_{m+1} + \ldots + f_i + \ldots,$$

where each f_i is a homogeneous polynomial of degree i in x and y, and where $m = m_p(C)$ is the multiplicity of C at p. Every homogeneous polynomial in two variables can be factored into a product of linear forms, and we may therefore write $f_m = l_1 \cdot \ldots l_m$, where the l_i's are linear forms. The lines $Z(l_i)$ (or their closures $Z_+(l_i)$ in the projective plane $\mathbb{P}^2$) are called the *tangents* of C at p. They are unique up to order, but need not be different. For instance, the cusp $y^2 - x^3 = 0$ has just one tangent line $y = 0$, but it appears with multiplicity **two**.

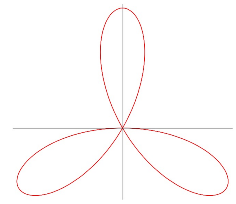
Tangent lines
tangenter

11.19 Note that this notion of a tangent line is a refinement of the notion of a tangent spaces introduced in Paragraph 8.1 on page 165 as the null-space of the Jacobian matrix. In a singular point p of $C = Z(f)$, both partials of f vanish, and hence $T_pC = \mathbb{A}^2$.

Examples

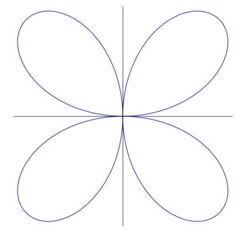
11.20 The ordinary double point $y^2 - x^2(x + a) = 0$ with $a \neq 0$ has of course multiplicity two at the origin and the two tangents are $y = \pm\sqrt{a}x$, which are distinct when the characteristic of k is not two.

11.21 The *trifolium*, whose equation is $(x^2 + y^2)^2 + 3x^2y - y^3 = 0$, has a triple point at the origin, and the lowest term of the equation factors as $3x^2y - y^3 = y(\sqrt{3}x - y)(\sqrt{3}x + y)$. Thus, when the characteristic is different from two and



The trifulium

three, the trifolium has three simple tangents at the origin; the x-axis and the two lines $y = \pm x/\sqrt{3}$. When the characteristic is two, the two lines $y = \pm x/\sqrt{3}$ coalesce and give a double tangent, and when it equals three, the lowest term becomes y^3, and the x-axis is a triple tangent.



the quadrifolium

11.22 The *quadrifolium*, whose equation is $(x^2+y^2)^3 - x^2y^2 = 0$, has a quadruple point at the origin, and the lowest term of the equation factors as x^2y^2. Thus, when the characteristic is different from two, the quadrifolium has two double tangents at the origin, the x-axis and the y-axis. When the characteristic is two, the defining polynomial becomes the square of $(x+y)^3 - xy$, and the origin is an ordinary double point with tangents the two axes.

☆

11.23 Inspired by pictures of real points of singular curves – like the ones in the margin – one says loosely[3] that a curve has a "branch" corresponding to each tangent vector at a point p. For instance the, nodal cubic has two "branches" at the origin (you might find the picture of the blow-up on Page 177 convincing), and the trifolium has three.

[3] The term may be made precise in several ways; *e.g.* by considering fibre of the normalization map over the point, by blowing up the point or passing to the so-called $\mathfrak{m}$-adic completion of the local ring

In terms of "branches" the heuristics behind the next result are as follows: Given two curves C and D through p. Each "branch" of one of the curves C passing by p meets every "branch" of the other curve D, and this incisce contributes at least one to the the total intersection multiplicity (the contribution exceeds one when the two branches touch each other). This indicates clearly that the intersection multiplicity is at least equal to the product of the number of "branches" of the two curves. A "branch" can however contribute more than one to the multiplicity, as for a cusp, but the heuristics are still sound:

Proposition 11.24 *Let $p \in \mathbb{P}^2$ be a point and and assume two curves C and D be given, both passing through p. Then $\mu_p(C,D) \geqslant m_p(C)\cdot m_p(D)$, and equality occurs when and only when C and D do not have common tangent lines through p.*

Proof: Let f and g be equations for C and D in a distinguished open affine neighbourhood of p as above, and to ease the notation, let $\mu = m_p(C)$ and $\nu = m_p(D)$; then $f \in \mathfrak{m}^\mu$ and $g \in \mathfrak{m}^\nu$. Consider the exact sequence

$$k[x,y]/\mathfrak{m}^\mu \oplus k[x,y]/\mathfrak{m}^\nu \xrightarrow{\alpha} k[x,y]/\mathfrak{m}^{\mu+\nu} \longrightarrow (k[x,y]/(f,g))_{\mathfrak{m}} \longrightarrow 0,$$

where $\alpha([a],[b]) = [ag - bf]$ (as usual, brackets indicate appropriate residue classes). The map α is well defined since $f\mathfrak{m}^\nu \subseteq \mathfrak{m}^{\mu+\nu}$ and $g\mathfrak{m}^\mu \subseteq \mathfrak{m}^{\mu+\nu}$. Taking dimension we get

$$\begin{aligned}\mu_p(C,D) &= \dim_k \left(k[x,y]/(f,g)\right)_{\mathfrak{m}} = \dim_k k[x,y]/\mathfrak{m}^{\mu+\nu} - \dim_k \operatorname{im}\alpha \geqslant \\ &\geqslant \dim_k k[x,y]/\mathfrak{m}^{\mu+\nu} - \dim_k k[x,y]/\mathfrak{m}^{\nu} - \dim_k k[x,y]/\mathfrak{m}^{\mu} = \\ &= \binom{\mu+\nu}{2} - \binom{\mu}{2} - \binom{\nu}{2} = \mu\nu.\end{aligned}$$

Assume then that all the tangents of the two curves are distinct; in other words, that the lowest terms f_μ and g_ν of f and g are without common factors. We contend that α is injective in that case; and hence the inequalities above become equalities, and we will be through.

Suppose next that the curves do not have a common tangent at p. We are to show that α is injective. So let a and b be polynomials with $ag - bf \in \mathfrak{m}^{\mu+\nu}$, and write $a = a_s + a'$ and $b = b_t = b'$ with a_s and b_t being the homogeneous terms of lowest degree.

If $s + \nu < t + \mu$, the lowest term of $ag - bf$ will be $a_s g_\nu$, and it follows that $s \geqslant \mu$; hence $[a] = 0$. Then $bf \in \mathfrak{m}^{\mu+\nu}$ and $t \geqslant \mu$, and we infer that $[b] = 0$. Similarly, if $t + \mu < s + \nu$, it ensues that $[a] = [b] = 0$.

Finally, in the remaining case $s + \nu = t + \mu$ there are two possibilities: it may be that as a polynomial $a_s g_\nu - b_t f_\mu \neq 0$, and it then ensues that $s \geqslant \mu$ and $t \geqslant \nu$, so that $[a] = [b] = 0$. If not, the polynomial identity $a_s g_\nu - b_t f_\mu = 0$ holds true. Consequently f_μ divides a_s and $[a] = 0$ since by hypothesis g_ν and f_μ are without common factors. For the same reason g_ν divides b_t, so that $[b] = 0$.

We leave it to the students to check that α is not injective when f_μ and g_ν have a common factor. ❑

Example 11.25 Let us find the intersection number of the *quadrifolium* Q with equation $(x^2+y^2)^3 - x^2y^2 = 0$ and the standard cuspidal cubic C given as $y^2 = x^3$. In the local ring $k[x,y]_{(x,y)}$ we have the equalities

$$\begin{aligned}\left((x^2+y^2)^3 - x^2y^2, y^2 - x^3\right) &= \left((x^2+x^3)^3 - x^5, y^2 - x^3\right) = \\ &= \left(x^5(x(1+x)^3) - 1, y^2 - x^3\right) = (x^5, y^2 - x^3).\end{aligned}$$

Hence $\mu_p(C,Q) = \dim_k k[x,y]/(x^5, y^2 - x^3) = 10$ since, as one easily verifies, the ten monomials $y^\epsilon x^\eta$ for $0 \leqslant \epsilon \leqslant 1$ and $0 \leqslant \eta \leqslant 4$ form a vector space basis for $k[x,y]/(x^3, y^2 - x^3)$. ☆

Exercises

11.7 Let C be a plane curve and $L = Z_+(l)$ a line in $\mathbb{P}^2$ that meets C in a point p.

a) Show that $\mu_p(C,L)$ equals the exponent of the linear form l in the factorization of the lowest term of the equation for C near p;
b) Show that L is tangent to C at p if and only if $\mu_p(C,L) > 1$.

11.8 In case $n = 2$, the unmixedness theorem is almost trivial, and additivity comes for free. Prove additivity for local intersections multiplicities for two effective divisors in $\mathbb{P}^2$.

11.9 The relative behaviour of two intersecting curve can be rather complicated. The following example is taken from William Fulton's book[4] Show that the local intersection number at the origin of the two curves $(x^2+y^2)^2+3x^2y-y^3=0$ and $(x^2+y^2)^3-4x^2y^2=0$ equals 14. Where else do they intersect? HINT: Additivity can be useful.

[4]

★

11.3 *Hilbert functions and the degree of a variety*

In this section, we let $R = k[x_0,\dots,x_n]$ and M is a finitely generated R-module. We define the *Hilbert function* of M as the function $h_M : \mathbb{Z} \to \mathbb{Z}$ given by

Hilbert function
Hilbert-funksjonen

$$h_M(i) = \dim_k M_i$$

In short, h_M gives the dimensions of the graded pieces of M.

EXAMPLE 11.26 For $M = R$, we have $h_M(d) = \binom{n+d}{n}$ for $d \geqslant 0$, that is, the number of monomials of degree d in $x_0,\dots,x_n$.

More generally, for the *twisted module* $M = R(m)$ (whose underlying module structure is just R, but with the grading shifted: $M_i = R_{i+m}$), we have $h_M = \binom{n+m+d}{n}$. ☆

EXAMPLE 11.27 Let $R = k[x_0,x_1]$ and $M = R/(x_0^3, x_0x_1, x_1^5)$. Then the values of the Hilbert function is given in the following table:

i	0	1	2	3	4	5	6	7	$\cdots$
h_M	1	2	2	1	1	0	0	0	$\cdots$

For $R = k[x_0,x_1,x_2,x_3]$ and $M = R/(x_0^2, x_1^3, x_2^5)$, the corresponding Hilbert function is given by

i	0	1	2	3	4	5	6	7	$\cdots$
h_M	1	4	9	15	21	26	30	30	$\cdots$

☆

The two last examples are typical in the sense that the Hilbert function becomes equal to a polynomial for large values. This is the content of the following theorem.

Theorem 11.28 (Hilbert–Serre) *Let M be a finitely generated graded R-module. Then there is a unique polynomial $\chi_M \in \mathbb{Q}[z]$ such that*

$$h_M(m) = \chi_M(m) \quad \text{for all } m \gg 0.$$

Furthermore,

i) $\deg \chi_M = \dim Z_+(\operatorname{Ann} M) \subset \mathbb{P}^n$;

ii) If $Z_+(\operatorname{Ann} M) \neq \emptyset$, then the leading coefficient of χ_M is of the form

$$\frac{a}{(\deg \chi_M)!}$$

where a is an integer.

Proof: This is theorem **??** on page **??** in CA; see the proof there. ❑

We will in particular be interested in the case where $M = R/\mathfrak{a}$ for some homogeneous ideal $\mathfrak{a} \subset R$; in particular the case that $\mathfrak{a} = I(X)$ for a closed projective set $X = Z_+(\mathfrak{a}) \subset \mathbb{P}^n$ will be of interest. Then $Z_+(\operatorname{Ann} M) = X$, and $\deg \chi_M = \dim X$. In this case, we define the *degree* $\deg X$ of X to be the integer a that figures in the theorem; or in other words, with $d = \dim X$, the degree $\deg X$ equals $d!$ times the leading coefficient of $\chi_{R/\mathfrak{a}}(z)$. That is, it holds that

The degree of a projective variety
graden til en projektiv varietet

$$\chi_{R/\mathfrak{a}}(z) = \frac{\deg X}{d!} z^d + \text{lower order terms}.$$

Note that if X is non-empty, the degree $\deg X$ is a positive number. Indeed, if $X \neq \emptyset$, then the ideal $\mathfrak{a}$ is not primary for the irrelevant ideal $\mathfrak{m}_+$, and $\mathfrak{a}_\nu \neq R_\nu$, for large ν. It follows that $\chi_{R/\mathfrak{a}}(\nu) > 0$ for large ν, and hence $\chi_{R/\mathfrak{a}}$ has a positive leading coefficient. Conversely, if $X = \emptyset$, then by the Projective Nullstellensatz (Theorem 4.17 on page 79), $\mathfrak{a}$ is primary for the irrelevant ideal $\mathfrak{m}_+$, and there is an integer $N > 0$ so that $x_i^N \in \mathfrak{a}$ for every $i = 0, \dots, n$; hence $(R/\mathfrak{a})_\nu = R_\nu/\mathfrak{a}_\nu = 0$ for $\nu \geqslant N(n+1) + 1$, and consequently $\chi_M = 0$.

11.29 Let us check that the degree of a hypersurface $X = Z_+(F) \subset \mathbb{P}^n$ indeed equals the degree of the defining polynomial F. If F has degree d, there is an exact sequence

$$0 \to R(-d) \to R \to R/(F) \to 0$$

of graded modules (where the first map is multiplication by F). Hilbert functions are additive on exact sequences, so for $z \geqslant d$ it follows that

$$\begin{aligned} h_{R/(F)}(z) &= h_R(z) - h_{R(-d)}(z) \\ &= \binom{z+n}{n} - \binom{z-d+n}{n} = \frac{d}{(n-1)!} z^{n-1} + \dots \end{aligned}$$

and hence $\deg X = d$.

EXAMPLE 11.30 *(The twisted cubic)* Let $\mathfrak{a} \subset k[x_0, x_1, x_2, x_3]$ be the ideal of the twisted cubic $X = Z(\mathfrak{a}) \subset \mathbb{P}^3$; that is, the ideal generated by the 2×2-minors of the matrix

$$A = \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{pmatrix}$$

In other words $\mathfrak{a} = (q_0, q_1, q_2)$ where $q_0 = x_1x_3 - x_2^2$, $q_1 = x_1x_2 - x_0x_3$ and $q_2 = x_0x_2 - x_1^2$. There is an exact sequence (see Exercise **??** on page **??** in CA for a sketch of a proof; there is a solution provided.)

$$0 \to R(-3)^2 \xrightarrow{\alpha} R(-2)^3 \xrightarrow{\beta} R \to R/I \to 0$$

where $\alpha : R^2 \to R^3$ is given by the matrix A and β sends the i-th basis vector e_i to q_i. This gives the following Hilbert polynomial for R/I:

$$\begin{aligned} h_{R/I}(z) &= h_R(z) - 3h_R(z-2) + 2h_R(z-3) \\ &= \binom{z+3}{3} - 3\binom{z+1}{3} + 2\binom{m}{3} \\ &= 3m + 1. \end{aligned}$$

Since X has dimension 1, the degree is equal to 3 (as the name suggests). ☆

EXERCISE 11.10 Let $X, Y \subset \mathbb{P}^n$ be two closed algebraic sets of the same dimension m with no common component. Then show that

$$\deg(X \cup Y) = \deg X + \deg Y. \tag{11.2}$$

HINT: : Use a relevant exact sequence relating the ideals $I(X), I(Y), I(X \cap Y)$ and $I(X \cup Y)$. ★

11.4 *Proof of Bezout's theorem*

Bearing the Unmixedness Theorem in mind, we proceed with proving Bézout's theorem under the assumption that the polynomials $F_1, \ldots, F_n$ form a regular sequence. And there will two parts: firstly, we shall identify the product of the degrees of the n hypersurfaces as the Hilbert polynomial of the graded ring $S = k[x_0, \ldots, x_n]/(F_1, \ldots, F_n)$ (which is constant), and subsequently show that this constant equals the sum of the local multiplicities.

The Hilbert polynomial

The first step is thus to establish the following formula:

LEMMA 11.31 $\chi_S(t) = \deg Z_1 \cdots \deg Z_n$,

which follows as a special case of Lemma 11.33 below.

11.32 To prepare for Lemma 11.33, we introduce, for each r with $1 \leqslant r \leqslant n$, the quotient ring $S_r = k[x_0, \ldots, x_n]/(F_1, \ldots, F_r)$, and for convenience, we let $S_0 = k[x_0, \ldots, x_n]$; moreover, we put $d_i = \deg Z_i$ for each $1 \leqslant i \leqslant r$, and we let $d_0 = 1$. The sequence $F_1, \ldots, F_n$ being regular means by definition there are short exact sequences

$$0 \longrightarrow S_r(-d_{r+1}) \xrightarrow{F_{r+1}} S_r \longrightarrow S_{r+1} \longrightarrow 0,$$

one for each $0 \leqslant r \leqslant n-1$. The indicated map is just multiplication by F_{r+1}, and both maps in the sequences are homogeneous of degee 0, so that the following identity between Hilbert polynomials follows

$$\chi_{S_{r+1}}(t) = \chi_{S_r}(t) - \chi_{S_r}(t - d_{r+1}).$$

Now, it is an elementary fact that for any polynomial $P(t)$ of degree r with leading coefficient a, the difference $P(t) - P(t-d)$ is of degree $r-1$ with leading coefficient rda; indeed, for any natural number m the Binomial Theorem yields the equality

$$t^m - (t-d)^m = md \cdot t^{m-1} + o(m-2),$$

where $o(m-2)$ stands for a polynomial term of degree less than $m-2$. Using this, a straightforward induction gives the following (remember that $d_0 = 1$):

Lemma 11.33 *For any r with $0 \leqslant r \leqslant n$ it holds true that*

$$\chi_{S_r}(t) = d_0 \cdot \cdots \cdot d_r \frac{t^{n-r}}{(n-r)!} + o(n-r-1)$$

where the term $o(n-r-1)$ is a polynomial of degree at most $n-r-1$. In particular for $r = n$, we have $\chi_S(t) = d_1 \ldots d_n$.

Proof: Induction on r. ❑

From Hilbert functions to local multiplicities

We have now come to the point where we establish the link between the graded pieces of S and the algebra $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n}$. Recall that when setting the scene, we chose coordinates so that the entire intersection $Z_1 \cap \ldots \cap Z_n$ was contained in the basic open set $D = D_+(x_0)$, and it is very inviting to localize at x_0 and to consider the localized algebra S_{x_0}. Since x_0 is homogeneous, S_{x_0} is a graded algebra whose homogeneous elements are of the form $H \cdot x_0^{-r}$ with H (the residue class of) a homogeneous polynomial. The degree of the homogeneous element $H \cdot x_0^{-r}$ is of course equal to $\deg H - r$.

The first crucial fact is the following description of the graded pieces of S_{x_0}; they are all isomorphic to $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n}$.

Lemma 11.34 *The degree zero part of S_{x_0} equals $\mathcal{O}_{Z_1\cap\ldots\cap Z_n}$, and the decomposition of S_{x_0} into homogeneous pieces takes the form:*

$$S_{x_0} = \bigoplus_{i\in\mathbb{Z}} k[x_1x_0^{-1}, \ldots, x_nx_0^{-1}]/(f_1, \ldots, f_n)\cdot x_0^i = \bigoplus_{i\in\mathbb{Z}} \mathcal{O}_{Z_1\cap\ldots\cap Z_n}\cdot x_0^i.$$

Proof: The first thing to observe is that in the ring S_{x_0} where x_0 is invertible, the equality $(F_1, \ldots, F_n) = (f_1, .., f_n)$ holds true; indeed, $x_0^{d_i} f_i = F_i$. The second is that all the f_i's are homogeneous of degree zero.

The first part of the argument is quite general; in fact, it holds water for any graded ring having a unit of degree one. The inclusion $(S_{x_0})_0 \cdot x_0^d \subseteq (S_{x_0})_d$ is obvious, and because x_0 is invertible, the multiplication-by-x_0-map is injective. To see that is surjective, observe that any homogeneous element z in S_{x_0} is of the form $z = ax_0^s$ where a is homogeneous of degree zero and $s \in \mathbb{Z}$: Indeed, z equals $H(x_0, \ldots, x_n)x_0^{-r}$ for some homogeneous polynomial H and some integer r, and therefore $z = H(x_1x_0^{-1}, \ldots, x_nx_0^{-1}, 1)x_0^{d-r}$ where d denotes the degree of H. Hence $S_{x_0} = \bigoplus_{d\in\mathbb{Z}}(S_{x_0})_0 \cdot x_0^d$.

What remains, is to identify the degree zero piece $(S_{x_0})_0$ as $\mathcal{O}_{Z_1\cap\ldots\cap Z_n}$. To that end, consider the quotient map $R \to R/(F_1, \ldots, F_n)$. When localized in x_0 it yields the quotient map

$$R_{x_0} \to R_{x_0}/(f_1, \ldots, f_n)R_{x_0} = S_{x_0}.$$

Considering the degree zero part of this map, and observing that the degree zero part of the ideal $(f_1, \ldots, f_n)$ in R_{x_0} is the ideal $(f_1, \ldots, f_n)(R_{x_0})_0$ in $(R_{x_0})_0$ since the f_i's all are of degree zero, we are done; indeed, the degree zero part of R_{x_0} equals $(R_{x_0})_0 = k[x_1x_0^{-1}, \ldots, x_nx_0^{-1}]$. ❑

The second, and last, crucial element in the proof of Bézout's Theorem is the following:

Lemma 11.35 *For d sufficiently large, the localization map $S \to S_{x_0}$ induces an isomorphism $S_d \to (S_{x_0})_d$ between the graded pieces of degree d.*

Proof: There are two things to prove: that the map is injective and that it is surjective.

First of all, the kernel of the localization map is supported at the origin because the locus $Z(x_0)$ and the support of S only has the origin in common: since $Z_+(F_1, \ldots, F_n, x_0) = \emptyset$, the Projective Nullstellensatz implies that the ideal $(F_1, \ldots, F_n, x_0)$ is $\mathfrak{m}_+$-primary. Any element in the kernel of the localization map is killed by some power of x_0 and being an element in S, it is killed by the F_i's as well. Hence the kernel is killed by some power $\mathfrak{m}_+^N$, and it is therefore of finite dimension as a vector space over k. Being of finite dimension over k, the kernel can merely have finitely many graded pieces different from zero, and hence the localization maps induce injections $S_d \to (S_{x_0})_d$ in large degrees.

$$\begin{array}{ccc} S_d & \longrightarrow & (S_{x_0})_d \\ {\scriptstyle x_0^d}\uparrow & & \simeq\uparrow{\scriptstyle x_0^d} \\ S_0 & \longrightarrow & (S_{x_0})_0 \end{array}$$

Turning to the surjectivity, we observe that any homogeneous element $ax_0^r \in S_{x_0}$ with a of degree zero, can be expressed as a product $ax_0^r = Hx_0^{r-d}$ where H is the residue class mod $(F_1, \ldots, F_n)$ of a homogeneous polynomial of degree d, and consequently, when $r > d$, the element ax_0^r lies in the image of the localization map. So, take any basis $a_1, \ldots, a_r$ for $(S_{x_0})_0$ and write the members as products $a_j = H_j x_0^{-d_j}$ where H_j is the residue class of a homogeneous polynomial of degree d_j. If now $d > \max d_j$, all the products $a_j x_0^d$ lie in the image by our observation above; and since multiplication by x_0 is an isomorphism $(S_{x_0})_0 \to (S_{x_0})_d$, this shows that the localization map $S_d \to (S_{x_0})_d$ is onto, and we are done. ❑

11.36 Summing up, the two lemmas combined yields the result we want:

Proposition 11.37 *For $d \gg 0$, the localization map $S \to S_{x_0}$ induces an isomorphism between the graded piece S_d of S and $\mathcal{O}_{Z_1 \cap \ldots \cap Z_n} \cdot x_0^d$. In particular, the following equality holds true*

$$\dim_k S_d = \dim_k \mathcal{O}_{Z_1 \cap \ldots \cap Z_n}.$$

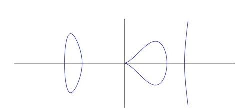
Proof of Bézout's theorem: Finally, to finish the proof of Bézout's theorem observe that the lemmas we have established and the definitions we have given, yield the following sequence of equalities:

$$d_1 \cdot \ldots \cdot d_n = \chi_S(d) = \dim S_d = \dim_k \mathcal{O}_{Z_1 \cap \ldots \cap Z_n} =$$
$$= \sum_p \dim_k \mathcal{O}_{Z_1 \cap \ldots \cap Z_n, p} = \sum_p \mu_p(Z_1, \ldots, Z_n).$$

❑

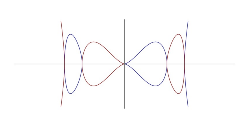
Exercises

11.11 Let $n > m$ be two natural numbers and let $\alpha(x)$ and $\beta(x)$ be two polynomials which do not vanish at $x = 0$. Determine the local intersection multiplicity at the origin of the two curves defined respectively by $y - \alpha(x)x^n$ and $y - \beta(x)x^m$. If $m = n$, show by exhibiting an example that the local multiplicity can take any integral value larger than n.



The affine pieces in $D_+(z)$ of one the two curves in problem 11.13

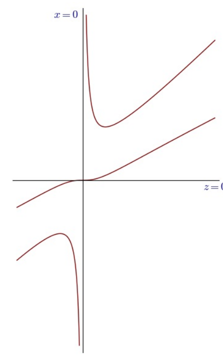
11.12 Find all intersection points of the two cubic curves defined by the forms $zy^2 - x^3$ and $zy^2 + x^3$ (we assume the characteristic of the ground field to be different from two). Determine all the local intersection multiplicities of the two curves.



The affine pieces in $D_+(z)$ of the two curves in problem 11.13

11.13 Let X and Y be two curves in $\mathbb{P}^2$ being the zero loci of the polynomials $z^5y^2 - x^3(z^2 - x^2)(2z^2 - x^2)$ and $z^5y^2 + x^3(z^2 - x^2)(2z^2 - x^2)$. Determine all intersection points and the local multiplicities in all the intersection points of X and Y

11.14 Let C be the curve given as $zy^2 - x(x-z)(x-2z)$. Determine the intersection points and the local multiplicities that C has withe line $z = 0$. Same task, but with the line $x - z = 0$.



11.15 There are several proofs of Bézout's theorem of quite different flavours. In this exercise you will led through a degeneration argument, a kind of result called a moving lemma. We shall do it in the plane $\mathbb{P}^2$, and treat the intersection of two irreducible curves. The idea is to degenerate one of curves say Z_1 into a union of lines each intersecting Z_2 properly, and the show that intersection number stays the same throughout the degeneration; a principle called the permanence of numbers.

a) For each $1 \leqslant i \leqslant d$ let L_i be a line that intersects each component of Z_2 in finite set choosen in the way that the different intersections $L_i \cap Z_i$ are disjoint. Show that $d_1 d_2 = \sum_p^{\mu_P}$.
b) Consider $W = Z(tF(x) + (t-1)L(x), G(x)) \subseteq \mathbb{A}^1 \times \mathbb{P}^2$. It decomposes into a union of components, which all map to $\mathbb{P}^1$, by Macaulay they are all of codimension two, that is dimension one. Some dominates $\mathbb{P}^1$, and keep those, and throw away the rest. The resulting vareity is Z.
c) Show that if A is DVR and B is a finite A-algebra such that t is not a zero-divisor in A, then B is a free A-module. Conclude that $\dim_K B \otimes_A K = \dim_k B \otimes_A k$.
d) Show that $B = A(Z)$ is a finite module over $k[t]$ and that $\dim_k B \otimes_A k(0) = d_1 d_2$ and $\dim_k B \otimes_k k(1) = \sum_P \mu_p(Z_1, Z_2)$.

★

11.5 Appendix: Depth, regular sequences and unmixedness

An important ingredient in the full proof of Bézout's theorem is the concept of so-called unmixed rings. These are Noetherian rings all whose associated prime ideals are of the same height, or what amounts to the same in our context of algebras of finite type over a field, that $\dim A/\mathfrak{p}$ is the same for all associated primes $\mathfrak{p}$. In particular A has no embedded components, the height of an embedded prime would of course be larger than the height of at least one of the others. In geometric terms, if $A = k[x_1, \ldots, x_n]/\mathfrak{a}$, all the components of the closed algebraic subset $X = Z(\mathfrak{a})$ are of the same dimension, and A has no embedded component.

Among other things Francis Sowerby Macaulay showed that if the ideal $(f_1, \ldots, f_r)$ is of height r, then the algebra $A = k[x_1, \ldots, x_n]/(f_1, \ldots, f_r)$ is unmixed. That the irreducible components of the closed algebraic set $Z = Z(f_1, \ldots, f_r)$ all are of codimension r is clear: the height is the smallest codimension of any of the components, and Krull's Hauptidealsatz tells us that every component is of codimension at most r, so the subtle content of Macaulay's result is that there are no embedded components. This has the consequence that if f_{r+1} is a new polynomial not vanishing along any of the components, then f_{r+1} is a non-zero divisor in A; indeed, the set of zero divisors in A equals the union of the associated prime ideals, and these are just prime ideals of the different irreducibel components of Z.

Clearly, $A_s = k[x_1, \ldots, x_n]/(f_1, \ldots, f_s)$ is also unmixed for $s < r$; indeed, any oversized component in $Z(f_1, \ldots, f_s)$ would stay oversized when intersected with the hypersurfaces $Z(f_{s+1}), \ldots, Z(f_r)$, and hence $(f_1, \ldots, f_s)$ is of height s and hence unmixed by Macaulay's result. By the observation above, f_{s+1} is not a zero-divisor in A_s.

Turning thus property into a definition and saying that a sequence $\{f_i\}$ of polynomials so that f_s is not a zero-divisors in A_s is a regular sequence, Macaulay's result may thus be paraphrased as a sequence $(f_1, \ldots, f_r)$ being of height r is equivalent to $f_1, \ldots, f_r$ being a regular sequence. This result was later vastly generalized by Irvin Cohen to other rings than polynomial rings.

Regular sequences

The theory of Cohen–Macaulay rings and more generally of the Cohen–Macaulay modules, is based on the concept of *regular sequences* which was introduced by Jean Pierre Serre in 1955. Their basic properties are described in this paragraph.

11.38 The stage is set as follows: we are given a ring A together with a proper ideal $\mathfrak{a}$ in A and an A-module M. Most of the time A will be local and Noetherian and M will be finitely generated over A.

A sequence $x_1, \ldots, x_r$ of elements belonging to the ideal $\mathfrak{a}$ is said to be *regular for M*, or *M-regular* for short, if the following two conditions are fulfilled, where we for notational convenience let $x_0 = 0$.

Regular sequences
regulære følger

i) $\mathfrak{a}M \neq M$;

ii) For any i with $1 \leqslant i \leqslant r$ the multiplication-by-x_i map

$$M/(x_1, \ldots, x_{i-1}) \longrightarrow M/(x_1, \ldots, x_{i-1})$$

is injective.

In other words, each quotient $M/(x_1, \ldots, x_{i-1})$ is non-zero, and x_i is a not a zero-divisor in $M/(x_1, \ldots, x_{i-1})$. In particular, x_1 is not a zero-divisor in M. The first condition is always verified in situations where Nakayama's lemma is valid; *e.g.* when M is finitely generated and A is local, or when A is positively graded and M is a finitely generated graded S-module.

11.39 A regular sequence $x_1, \ldots, x_r$ is said to be *maximal* if it does not remain regular when an element from $\mathfrak{a}$ is appended. For a Noetherian module M this is equivalent to $\mathfrak{a}$ being contained in one of the associated primes of $M/(x_1, \ldots, x_r)$; indeed, the union of these associated primes is precisely the set of zero-divisors

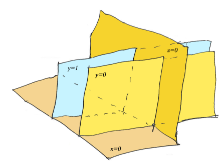
Maximal regular sequences
maksimale regulære følger

in $M/(x_1, \ldots, x_r)$, and prime avoidance tells us that $\mathfrak{a}$ is contained in the union precisely when it is contained in one of the associated primes.

EXERCISE 11.16 Show that maximal regular sequences for Noetherian modules are finite. Exhibit a counterexample when M is not Noetherian. HINT: Consider the ascending chain $(x_1, \ldots, x_i)$ of ideals. ★

11.40 Be aware that in general the order of the x_i's is important: permute them and the sequence may no more be regular. However, in situations where Nakayama's lemma holds true, regular sequences always remain regular after an arbitrary permutation. Henceforth, we shall exclusively stick to such a situation (*e.g.* rings are Noetherian and local and modules finitely generated over A (but with a sideways glimpse into the graded case, so important for projective geometry).

EXAMPLE 11.41 The simplest example of a sequence that ceases being regular when permuted is as follows. Start with the three coordinate planes in $\mathbb{A}^3$; they are given as the zero loci of x, y, z. Add a plane disjoint from one of them to the two others; *e.g.* consider the zero loci of the three polynomials $x(y-1)$, y and $z(y-1)$.



The sequence $x(y-1), z(y-1), y$ is *not* a regular sequence in $k[x, y, x]$: the point is that $z(y-1)$ kills any function on $Z(x(y-1))$ that vanishes on the component $Z(x)$ (for example x), and is thus not a zero-divisor in $k[x, y, z]/(x(y-1))$.

On the other hand, the sequence $x(y-1), y, z(y-1)$ is regular. Indeed, it holds that $k[x, y, z]/(x(y-1), y) = k[z]$, and in that ring $z(y-1)$ is congruent to $-z$, and thence is not a zero-divisor. Geometrically, capping $Z(x(y-1))$ with $Z(y)$ makes the villain component $Z(y-1)$ go away.

This example is in fact arche-typical. The troubles occur when two of the involved closed algebraic sets have a common component disjoint from one of the components of a third. If all components of all the closed algebraic subsets involved have a point in common, one is basically in a local situation, and permutations are permitted. ☆

Permutations permitted

11.42 As mentioned in the previous example, in local Noetherian rings a sequence being regular is a property insensitive to order. The same holds true in a graded setting, and in both cases Nakayama's lemma is the tool that makes it work.

LEMMA 11.43 *Assume that A is a local Noetherian ring with maximal ideal $\mathfrak{m}$ and M a finitely generated A-module. If x_1, x_2 is a regular sequence in $\mathfrak{m}$ for M, then x_2, x_1 is one as well.*

PROOF: There are two things to be checked. Firstly, that x_2 is a non-zero divisor in M. The annihilator $(0 : x_2)_M = \{ a \in M \mid x_2 a = 0 \}$ must map to zero in M/x_1M because multiplication by x_2 in M/x_1M is injective. Hence $(0 : x_2)_M + x_1M = x_1M$, and since $x_1 \in \mathfrak{m}$ and M is finitely generated, Nakayama's lemma applies and $(0 : x_2)_M = 0$.

Secondly, we are to see that multiplication by x_1 is injective on M/x_2M, so assume that $x_1a = x_2b$. But multiplication by x_2 is injective on M/x_1M, and it follows that $b = cx_1$ for some c; that is, $x_1a = x_1x_2c$. Cancelling x_1, which is legal since x_1 is a non-zero divisor in M, we obtain $a = cx_2$. ❑

PROPOSITION 11.44 *Let A be a local Noetherian ring with maximal ideal $\mathfrak{m}$ and M a finitely generated A-module. Assume that $x_1, \ldots, x_r$ is a regular sequence in $\mathfrak{m}$ for M. Then for any permutation σ the sequence $x_{\sigma(1)}, \ldots, x_{\sigma(n)}$ is regular.*

PROOF: It suffices to say that any permutation can be achieved by successively swapping neighbours. ❑

11.45 The graded version reads as follows:

PROPOSITION 11.46 *Let A be a graded ring satisfying $A_i = 0$ when $i < 0$, and let M be a finitely generated graded A-module. If $x_1, \ldots, x_r$ is a sequence of elements from A, homogeneous of positive degree, that form a regular sequence in M, then for any permutation σ the sequence $x_{\sigma(1)}, \ldots, x_{\sigma(r)}$ is also a regular in M.*

PROOF: As above, one may assume that $r = 2$. The proof of Lemma 11.43 goes through *mutatis mutandis*; the submodule $(0 : x_2)_M$ will be a graded submodule because x_2 is homogeneous, and a version of Nakayma's lemma for graded modules is available (see for instance Exercise **??** on page **??** in CA). ❑

EXERCISE 11.17 With assumptions as in 11.44 or 11.46, prove that if $x_1, \ldots, x_r$ is a regular sequence for M and $\nu_1, \ldots, \nu_r$ is a sequence of natural numbers, then $x_1^{\nu_1}, \ldots, x_r^{\nu_r}$ will be a regular sequence as well. HINT: Reduce to the case of $x_1, \ldots, x_{r-1}, x_r^{\nu}$. ★

EXERCISE 11.18 Jean Dieudonné gave the following example of a regular sequence x_1, x_2 in local non-Noetherian ring such that x_2, x_1 is not regular. Consider the ring B of germs of C^∞-functions near 0 in $\mathbb{R}$. It is a local ring whose maximal ideal $\mathfrak{m}$ constists of the functions vanishing at zero. Let $\mathfrak{a}$ be the ideal $\mathfrak{a} = \bigcap_i \mathfrak{m}^i$ of functions all whose derivatives vanish at the origin. Let $A = B[T]/\mathfrak{a}TB[T]$. Let I be the function $I(x) = x$. Show that the sequence I, T is a regular sequence in A whereas T, I is not. ★

The depth

It is natural to introduce the number $\operatorname{depth}_{\mathfrak{a}} M$ as the length of the longest regular M-sequence in $\mathfrak{a}$. It is called the *depth* of M in $\mathfrak{a}$. In the end, it turns out that all maximal M-sequences in $\mathfrak{a}$ have the same length, but for the moment we do not know that, and *a priori* the number is not even bounded. However, we have:

The depth of a module
dybden til en modul

LEMMA 11.47 *If A is a local Noetherian ring with maximal ideal $\mathfrak{m}$, and $\mathfrak{a}$ a proper ideal and M a finitely generated A-module, then $\operatorname{depth}_{\mathfrak{a}} M \leqslant \dim M$. In particular, $\operatorname{depth}_{\mathfrak{a}} M$ is finite and $\operatorname{depth}_{\mathfrak{a}} M \leqslant \dim A/\mathfrak{p}$ for all prime ideals associated to M.*

PROOF: Induction on $\dim M$ (which is finite!). If $\dim M = 0$, the maximal ideal $\mathfrak{m}$ is the only associated prime of M. Therefore every element in $\mathfrak{m}$ is a zero divisor and $\operatorname{depth}_{\mathfrak{a}} M = 0$.

Next, observe that if x is a non-zero divisor in M, it holds true that $\dim M/xM < \dim M$, and by induction one may infer that

$$\operatorname{depth}_{\mathfrak{a}} M/xM \leqslant \dim M/xM < \dim M. \tag{11.3}$$

So if $x_1, \ldots, x_r$ is a maximal regular sequence in M (they are all finite after Problem 11.16), the sequence $x_2, \ldots, x_r$ will be one for M/x_1M, and by (11.3) $r - 1 < \dim M$; that is $r \leqslant \dim M$. ❏

11.48 In a situation when permuting regular sequences is permitted; that is over a local Noetherian ring or in the graded case, it is not very difficult to prove that maximal regular sequences have the same length. This follows from a homological characterization of the depth, important by itself, but here we give an abecedarian proof found by David Rees.

PROPOSITION 11.49 *Let A be a Noetherian local ring, $\mathfrak{a}$ a proper ideal and M a finitely generated A-module. Then all maximal regular M-sequences in $\mathfrak{a}$ have the same length; which, of course, equals $\operatorname{depth}_{\mathfrak{a}} M$.*

PROOF: Let $x_1, \ldots, x_r$ be a maximal regular sequence and let $y_1, \ldots, y_s$ be another one. The proof will be by induction on r. The crux of the induction step is to replace the last elements in the sequences by a common element: We contend that there is an element $z \in A$ so that $x_1, \ldots, x_{r-1}, z$ and $y_1, \ldots, y_{s-1}, z$ both are regular sequences. If this is the case, since permutations are permitted, the sequences $z, x_1, \ldots, x_{r-1}$ and $z, y_1, \ldots, y_{s-1}$ are regular M-sequences. Hence $x_1, \ldots, x_{r-1}$ and $y_1, \ldots, y_{s-1}$ are regular sequences on M/zM. Induction applies, and we infer that $r - 1 = s - 1$; that is, $r = s$.

To check that ring elements as desired are about, remember that the zero-divisors on $M/(x_1, \ldots, x_{r-1})M$ and those on $M/(y_1, \ldots, y_{s-1})$ are the the unions of the respective associated primes, and by prime avoidance, the union of all these prime ideals is not equal to maximal ideal: if it where, one of them would equal the maximal ideal $\mathfrak{m}$; which would contradict that x_r is not a zero-divisor on $M/(x_1, \ldots, x_{r-1})$ and x_s not one on $M/(y_1, \ldots, y_{s-1})M$.

The more subtle part is the start of the induction; the case that $r = 1$. So given two regular elements x and y on M, and assume that x is a maximal regular sequence. This means that $\mathfrak{m}$ is associated to M/xM, so there is an element $z \in M$ with $\mathfrak{m}z \subseteq xM$ but $z \notin xM$. Now, $y \in \mathfrak{m}$, and hence $yz = xv$ for some $v \in M$. We contend that $v \notin yM$ and $\mathfrak{m}v \subseteq yM$; which means that $\mathfrak{m}$ is associated to M/yM, and y is a maximal sequence as well.

The first claim follows since $v = yw$ would imply that $yz = xv = xyw$ hence $z = xw$ which is not the case. For the second, observe that $x\mathfrak{m}v = y\mathfrak{m}z \subseteq xyM$, and since x is a non-zero devisor on M it ensues that $\mathfrak{m}v \subseteq yM$. ❑

11.50 Nakayama's lemma is valid in a graded situation, permutation is permitted, and hence one obtains the following graded analogy to Proposition 11.49 ; and as usual, the proof is *mutatis mutandis* as the one just given.

PROPOSITION 11.51 *Let A be a Noetherian graded ring satisfying $A_i = 0$ when $i < 0$, and let M be a finitely generated graded A-module. Then all homogeneous maximal regular M-sequences have the same length.*

Cohen-Macaualy rings

Above in Lemma 11.47 we established that $\operatorname{depth}_{\mathfrak{m}} M \leqslant \dim M$ whenever A is local and Noetherian with maximal ideal $\mathfrak{m}$ and M is a finitely generated A-module. Modules for which the equality $\operatorname{depth}_{\mathfrak{m}} M = \dim M$, holds true, have several particularly good properties; they are called *Cohen-Macaulay modules*. In particular, the local ring A itself is *Cohen-Macaulay* if $\operatorname{depth}_{\mathfrak{m}} A = \dim A$. In general, a Noetherian ring is called Cohen–Macaualy if all localizations $A_{\mathfrak{p}}$ in prime ideals are.

Cohen-Macaulay modules
Cohen-Macaulay moduler

11.52 An important property is that quotient of Cohen–Macaulay modules by non-zero divisors persist being Cohen–Macaulay; in fact, more holds true:

LEMMA 11.53 *Assume that A is a local Noetherian ring with maximal ideal $\mathfrak{m}$. Let M be finitely generated A-module and $x \in \mathfrak{m}$ a non-zero divisor on M. Then M is Cohen–Macaulay if an only if M/xM is.*

PROOF: Both the depth and the dimension of M/xM are one less than of M, and hence $\operatorname{depth}_{\mathfrak{m}} M = \dim M$ if and only if $\operatorname{depth}_{\mathfrak{m}} M/xM = \dim M/xM$. ❑

11.54 Cohen-Macaulay rings are unmixed:

Theorem 11.55 (Unmixedness Theorem) *Assume that A is a local Noetherian Cohen-Macaulay ring. Then A is unmixed. That is* $\dim A/\mathfrak{p} = \dim A$ *for all associated primes $\mathfrak{p}$ of A; in particular, A has no embedded components.*

Proof: In view of Lemma 11.47 this is almost a tautology since it yields the inequalities

$$\operatorname{depth}_{\mathfrak{m}} A \leqslant \dim A/\mathfrak{p} \leqslant \dim A.$$

For a Cohen–Macaulay ring, the lower and the upper bounds coincide, hence all the dimensions $\dim A(\mathfrak{p})$ coincide. ❑

11.56 Recall that a *system of parameters* in a local ring A of dimension n is a sequence $x_1, \ldots, x_n$ so that the quotient $A/(x_1, \ldots, x_n)$ is of dimension zero. In geometric terms a sequence of functions $f_1, \ldots, f_n$ in a coordinate ring $A(X)$ is a system of parameters at a point p if in a neighbourhood of p their only common zero is p. One nice thing about Cohen–Macaulay rings is that this property, which is geometric in the sense that it involves only dimensions, in fact implies that the sequence is regular, which is an *a priori* much stronger algebraic property. This was exactly what was needed in the proof of Bézout's theorem.

Theorem 11.57 *The following holds true for a local Noetherian ring A:*

i) A is Cohen–Macaulay if one system of parameters form a regular sequence;

ii) If A is Cohen–Macaulay, every system of parameters form a regular sequence.

Proof: Statement *i)* is proved by induction on $\dim A$; since all zero-dimensional local rings are Cohen–Macaulay, it holds true for $n = 0$. Let then $x_1, \ldots, x_n$ be a system of parameters that forms a regular sequence, $\dim A/x_1 = \dim A - 1$ and $x_2, \ldots, x_n$ is a both a regular sequence and a system of parameters on A/x_1. Hence A/x_1 is Cohen–Macaulay, and we are done by induction citing Lemma 11.53.

To verify *ii)* we us induction on $\dim A$ as well. Assume that $x_1, \ldots, x_r$ is a system of parameters; it must hold that $\dim A/x_1A < \dim A$. Indeed, by Krull's Hauptidealsatz the dimension can drop by at most one each time we mod out by an x_i, and to reach zero after n steps, it must drop every time.

Since A is a Cohen–Macaulay ring, all its associated primes $\mathfrak{p}$ are of dimension $\dim A/\mathfrak{p} = \dim A$, hence x_1 can not belong to any of the $\mathfrak{p}$'s, and as the zero-divisors in A equals the union of all the associated primes, it follows that x_1 is non-zero divisor. Thence A/x_1A is Cohen–Macaulay, the induction hypothesis ensures that $x_2, \ldots, x_r$ is a regular sequence in A/x_1A, and we are through. ❑

For instance, all the local rings $A_n = k[x_1, \ldots, x_n]_{\mathfrak{m}_n}$ where $\mathfrak{m}_n = (x_1, \ldots, x_n)$ are Cohen–Macaulay since the sequence $x_1, \ldots, x_n$ is regular. This follows easily by induction because there are natural isomorphisms $A_n/x_nA_n \simeq A_{n-1}$ induced by the maps $k[x_1, \ldots, x_n] \to k[x_1, \ldots, x_{n-1}]$ that send x_n to zero.

Chapter 12

Applications of Bézout's theorem

12.1 *Applications of Bezout's theorem*

In this chapter we touch a series of themes of different sorts, with a few exceptions about plane curves. They are all basic results in algebraic geometry, worthy of being included in any beginners course, and of course, they all have in common that they rely on Bezout's theorem.

12.2 *Automorphisms of* $\mathbb{P}^n$

We begin with describing all automorphisms of the projective spaces. Back in Paragraph 4.51 on page 91 we saw that each invertible linear map $\Phi\colon \mathbb{A}^{n+1} \to \mathbb{A}^{n+1}$ induces an isomorphism $\mathbb{P}^n \to \mathbb{P}^n$, which we called a linear automorphism, simply because $\Phi(x) \neq 0$ when $x \neq 0$ and $\Phi(ax) = a\Phi(x)$ for all $a \in k$. Clearly the composition of two linear maps induces the composition of the corresponding automorphisms, so we have a group homomorphism $\mathrm{GL}(n+1, k) \to \mathrm{Aut}(\mathbb{P}^n)$. Two such invertible linear maps are proportional – that is $\Phi = a\Phi'$ for $a \in k^*$ – if and only they induce coinciding automorphisms of $\mathbb{P}^n$, and the kernel of the map therefore consists of scalar multiples $a\,\mathrm{id}_{\mathbb{A}^{n+1}}$ of the identity. Passing to the quotient group $\mathrm{PGL}(n+1, k)$ we arrive at an injective map

$$\rho\colon \mathrm{PGL}(n+1, k) \to \mathrm{Aut}(\mathbb{P}^n). \qquad (12.1)$$

And we are about to show it is an isomorphism:

Theorem 12.1 *Any automorphism* $\phi : \mathbb{P}^n \to \mathbb{P}^n$ *is a linear transformation. That is, the map* ρ *is an isomorphism , so that*

$$\mathrm{Aut}(\mathbb{P}^n) = PGL_{n+1}(k).$$

Proof: The task is to show that ρ is surjective. So let $\phi : \mathbb{P}^n \to \mathbb{P}^n$ be an automorphism, and consider the hyperplanes $H_i = Z_+(x_i)$ for $i = 1, \ldots, n$. These meet in the single point $p = (1 : \ldots : 0 : 0)$.

Since ϕ is an isomorphism, each preimage $D_i = \phi^{-1}(H_i) \subset \mathbb{P}^n$ is a subvariety of $\mathbb{P}^n$ of codimension one; that is, a hypersurface of, say, degree d_i. Now, we have

$$\begin{aligned} \phi^{-1}(p) &= \phi^{-1}(H_1) \cap \phi^{-1}(H_2) \cap \ldots \cap \phi^{-1}(H_n) \\ &= D_1 \cap \cdots \cap D_n. \end{aligned}$$

Since ϕ is an isomorphism, it preserves intersection multiplicities, so by Bézout's theorem, we find that

$$1 = \deg D_1 \cdots \deg D_n = d_1 \cdots d_n.$$

The only way this is possible is that all the D_i are hypersurfaces of degree 1, i.e., ϕ sends the coordinate hyperplanes to hyperplanes. By conjugating with a linear coordinate change (which we are allowed to do!), we may assume that ϕ induces isomorphisms $D_+(x_i) \to D_+(x_i)$ for each $i = 0, \ldots, n$, that is, isomorphisms $\phi_i : \mathbb{A}^n \to \mathbb{A}^n$.

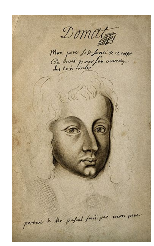
Each such isomorphism of affine spaces is given by an n-tuple of polynomials in $x_i^{-1}x_0, \ldots, x_i^{-1}x_n$. Since, for instance, ϕ_0^* sends each coordinate space $Z(x_0^{-1}x_j)$ to itself, we must have that $\phi_i^*(x_0^{-1}x_j) = \left(x_0^{-1}x_j\right)^{N_j}$ for each $j = 1, \ldots, n$ and some power N_j. Since ϕ is an isomorphism, it follows that $N_j = 1$, and ϕ_i is the identity map. From there we deduce that ϕ must also be the identity map; hence ρ is surjective. ❑

The theorems of Pappus and Pascal

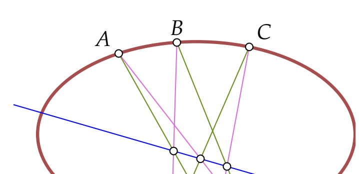
As an illustration of how one can apply Bézout's theorem, we shall prove two central results from classic projective geometry (there are of course myriads of other different proofs).

Pappus' theorem

12.2 In 1639 when he was merely 16 years old, Blaise Pascal wrote a short one page notice (printed as a broadside the following year) entitled *Essay pour les Coniques* where the first in his series of famous theorems was announced. Earlier the same year one saw the release of a note by the founder of projective geometry, Girard Desargues, with the longish title *Brouillon project d'une Atteinte aux evenemens des rencontres du Cone avec un Plan*, which inspired Pascal. As all theorems about conics Pascal's comes in two dual versions, but we shall only treat the one about conics circumscribed a hexagon (the dual is about conics inscribed in a hexagon). As a matter of notation, the line through two points A and B is denoted AB.



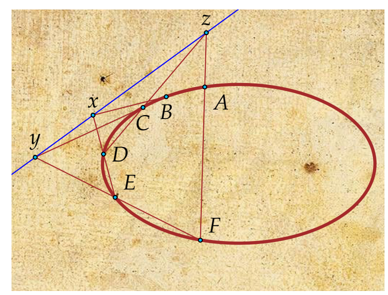
Blaise Pascal (1623–1623)
French mathematician



THEOREM 12.3 (PASCAL'S THEOREM) *Let Q be an irreducible plane conic and let A, B, C and a, b, c be two groups of points on Q all being different. Then the points $y = Ab \cap aB$, $x = Bc \cap bC$ and $z = Ca \cap cA$ are collinear.*

PROOF: Let f be the equation of the cubic which is the union $Ab \cup aC \cup Bc$ of three of the lines passing by the six points, and g the one for the union $aB \cup Ac \cup cB$. Both cubics pass through the nine points from the theorem. Let P be a general point on Q. In the family $af + bg$ (where $a, v \in k$) of cubics at least one, say R, passes by P; indeed, the space $V = \{\, af + bg \mid a, b \in k \,\}$ is a two dimensional subspace of the space of all cubic forms, and the evaluation map at P, which is a linear map $V \to k$, must have a non-trivial kernel. The cubic R has then seven points in common with the conic Q, and by Bézout's theorem, R and Q must have a common component. Since Q is irreducible, R must decompose as $R = Q \cup L$ where L is a line, and since none of the points x, y and z lies on Q, but lies on R, they must lie on L. ❑

12.4 The original way of stating the theorem is that points where opposite sides of a hexagon inscribed in a conic meet, are collinear: in the formulation above we make precise what the opposite sides are without mentioning the hexagon. In the old-style picture in the margin the hexagon is evident, but still has somehow weird proportions. The Pascal line gets further away from the conic the more regular the hexagon is (a regular hexagon inscribed in a circle gives the line at infinity), and this makes illustrating Pascal's theorem a challenging exercise.



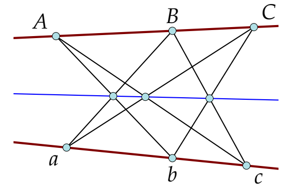
As a curio, given six points on the conic Q one easily verifies that there are 60 different ways of organizing them in a Pascal configuration as in the theorem, and so 60 lines, the so-called Pascal lines, emerge.

Pascal's theorem

12.5 The theorem of Pappus is a version of Pascal's for a degenerate conic: the union of two lines. And each of the two groups of three points lies on its own line. Pappus' theorem goes far back in the history. Pappus is an elusive character of the history of mathematics, not much is known about his life. He lived and worked in Alexandria, was born about 290 AD and died about 350 AD.

THEOREM 12.6 *Let A, B, C and a, b, c be two groups of different collinear points lying on the different lines L and l respectively. Then the points $y = Ab \cap aB$, $x = Bc \cap bC$ and $z = Ca \cap cA$ are collinear*

PROOF: The proof starts out just as the proof of Pascal's Theorem does, and the two proofs are identical up to the point where R and Q are found have a common component. Proceeding from there and remembering that Q the union of two lines, we deduce that one of them, say L, must be a component of R.



Hence $R = L \cup C$ with C a conic. Now, at least three of the original points do not lie on the line L, and lying on R, they must lie on C. Again by Bézout's theorem, as $Q = L \cup l$, the line l must be a component of C. It follows that $R \backslash (L \cup l)$ is a line as well, and it must contain x, y and z. ❏

Colin Maclaurin (1698–1746) Scottish Mathematician

An upper bound for the number of singular points

Our third application of Bézout's theorem is an upper bound for the number of singular points an irreducible plane curve can have—this is also an old result, going back at least to Colin Maclaurin who proved it in paper in 1720. If p is a singular point on a curve C and D is another curve passing through p, the local intersection number exceeds two: $\mu_p(C,D) \geqslant 2$. The idea is to chose an auxiliary curve X passing by all the singular points of C and a certain number of additional points. If N denotes the number of singular points C has and M the number of additional points X goes through, Bézout's theorem yields the estimate

$$\deg X \cdot \deg C = \sum_p \mu_p(C,X) \geqslant 2N + M, \tag{12.2}$$

and the subtlety is then to find X fulfilling the two opposing wishes: we want $\deg X$ small, but M large. This procedure gives the following bound:

PROPOSITION 12.7 *An irreducible curve $C \subseteq \mathbb{P}^2$ of degree d can not have more than $\binom{d-1}{2}$ singular points.*

PROOF: Let $N = \binom{d-1}{2}$, and going for a contradiction we suppose that C is singular in the $N+1$ points $p_1, \ldots, p_{N+1}$. Choose $d-3$ new additional points $p_{N+2}, \ldots, p_{N+d-2}$ on C. The crux is that one may always find a curve X of degree $d-2$ (not necessarily irreducible) that passes through $N+d-1$ given points, and Bézout's theorem, as expressed in (12.2), applied to C and X, then yields the impossible inequality

$$d(d-2) \geqslant 2(N+1) + d - 3 = 2(\binom{d-1}{2} + 1) + d - 3 = (d-1)(d-2) + 1.$$

To see there is an auxiliary curve X as desired, consider the space V_{d-2} of cubic forms of degree $d-2$. For each point $p \in \mathbb{P}^2$, there is an evaluation map[1] $V_{d-2} \to k$, which is linear, and summing these up over all the p_i's, gives a linear map

$$\Phi\colon V_{d-2} \to \sum_{p_i} k \simeq k^{N+d-1},$$

whose kernel consists of the forms that vanish at all the p_i's. Now, the dimensions are $\dim V_{d-2} = \binom{d}{2}$ and $N + d - 1 = \binom{d-1}{2} + d - 1 = \binom{d}{2} - 1$, so Φ has a kernel, and the wanted X can be found. ❏

[1] Be aware that this map is not canonical: one has to choose a point on the line in $\mathbb{A}^3$ corresponding to p and evaluate at that point; for the matter of vanishing or not, however, the choice is not important.

Note that the bound from (12.2) can be refined when closer estimates of the multiplicities of C at the singular points are know, as done in the following exercise.

EXERCISE 12.1 Let C be an irreducible curve of degree d whose equation is F. The aim of this exercise is to establish the inequality

$$\sum_p m_p(m_p - 1)/2 \leqslant (d-1)(d-2)/2$$

through the following points:

a) Show that $\sum_p m_p(C)(m_p(C) - 1) \leqslant d(d-1)$ HINT: Intersect C by the curve defined by one of the partial derivatives of F;
b) Given a point $p \in \mathbb{P}^2$. Show that requiring that a form should vanish to the order $m-1$ at p imposes $m(m-1)/2$ conditions on forms of a given degree;
c) Show that $r = (d-1)(d+2)/2 - \sum_p m_p(C)(m_p(C) - 1)/2 \geqslant 0$;
d) Pick r points on C and show that there is a curve of degree $n-1$ that passes through the r points and vanishes to the order $m_p(C) - 1$ in each singular point of C;
e) Conclude by Bézout's theorem HINT: You will also need Proposition 11.24.

★

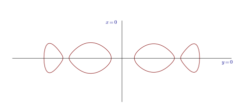
Harnack's theorem — a touch of reality

These notes are for the most concerned about varieties over an algebraically closed field; *e.g.* over the field $\mathbb{C}$ of complex numbers. Real algebraic geometry, the study of the real points of varieties, is however a large branch of algebraic geometry with a lot beautiful results, and a lot of intriguingly difficult questions.

Carl Gustav Axel Harnack (1851–1888) German Mathematician

We just want to give a glimpse into that fascinating world, an "amuse-bouche" as a French cook would say, and shall offer a short story about real curves in $\mathbb{RP}^2$; or more precisely: about the set of real points $C(\mathbb{R})$ of curves C in $\mathbb{P}^2_{\mathbb{C}}$ whose equations have real coefficients, *i.e.* it is the zero-locus of a homogeneous polynomials F in $\mathbb{R}[x, y, z]$.

12.8 The Jacobian criterion (Proposition 8.15 on page 173) tells us that C is non-singular when the partials F_x, F_y and F_z do not vanish simultaneously. In a standard distinguished open affine, *e.g.* $D_+(z)$, this means that the partials f_x and f_y do not vanish simultaneously at any point in $C \cap D_+(z)$, where f is the dehomogenized polynomial $f(x, y) = F(x, y, 1)$. This is precisely the condition of the implicit function theorem from analysis, so we conclude that when C is non-singular, $C(\mathbb{R})$ is a real, one-dimensional differential manifold. It is compact being closed in the compact space $\mathbb{RP}^2$. Compact one-dimensional manifolds are easily classified: they are diffeomorphic to the disjoint union of finitely



The real part of a hyperelliptic curve of degree 8.

many circles. So the set of real points $C(\mathbb{R})$ decomposes as a disjoin union $C(\mathbb{R}) = \gamma_1 \cup \ldots \cup \gamma_r$ where the γ_i's are circles embedded in $\mathbb{RP}^2$ as differential submanifolds. The circles γ_i's are called the *ovals* or the *circuits* of $C(\mathbb{R})$.

Ovals of a curve
ovalene til en kure
circuits of a curve
kretsene til en kurve

Be aware, that viewed in one of the standard affines of $\mathbb{RP}^2$, which are isomorphic to $\mathbb{R}^2$, a circuit belonging to $C(\mathbb{R})$ need not be a closed curve. It can stretch out to infinity and can very well have several components (see *e.g.* Example 12.13 below).

The first natural question that arises is: how many ovals can a non-singular curve of degree d have? It was answered by Carl Harnack already in 1876: the maximal number is $r \leqslant (d-1)(d-2)/2+1$ and in fact, for any number $r \leqslant (d-1)(d-2)/2+1$ he exhibited examples of curves of degree d with exactly r ovals, however, we shall contend ourself with proving Harnack's upper bound. The next question is what possible relative position the circuits can have; *e.g.* to which extend can they be nested? A complete answer to this is still only known in degrees $d \leqslant 7$.

Proposition 12.9 (Harnack's bound) *Let $C\mathbb{P}^2$ be a real non-singular curve of degree d. Then $C(\mathbb{R})$ has at most $(d-1)(d-2)/2+1$ ovals.*

The proof of the proposition is very similar to the proof of Proposition 12.7 where we gave a bound on the number of singular points. In that proof a factor two arose from the intersection multiplicity at singular points of the original and the auxiliary curve, in the present case that factor comes from the topology of the so-called even ovals; apart from this, the proofs are identical. We begin with recalling a few rudimentary topological facts.

12.10 The basic topological invariants of $\mathbb{RP}^2$, the fundamental group $\pi_1(\mathbb{RP}^2)$ and the homology group $H_1(\mathbb{RP}^2, \mathbb{Z})$ are both isomorphic to $\mathbb{Z}/2\mathbb{Z}$; that is we have isomorphism

$$\pi_1(\mathbb{RP}^2) \simeq H_1(\mathbb{RP}^2, \mathbb{Z}) \simeq H_1(\mathbb{RP}^2, \mathbb{Z}/2\mathbb{Z}) \simeq \mathbb{Z}/2\mathbb{Z}.$$

This follows readily from $\mathbb{RP}^2$ being obtained from the sphere $\mathbb{S}^2$ by identifying antipodal points, so that there is a two-to-one covering map $\mathbb{S}^2 \to \mathbb{RP}^2$ (remember, $\mathbb{S}^2$ is simply connected). Along that map, a line in $\mathbb{RP}^2$ pulls back to a great circle in $\mathbb{S}^2$,which is connected and covers the line doubly. The line is therefore not null-homotopic and generates $\pi_1(\mathbb{RP}^2)$. Thus we may take the class of a line as a generator of $H_1(\mathbb{RP}^2, \mathbb{Z})$ as well.

As for every real surface, there is an intersection pairing on $H_1(\mathbb{RP}^2, \mathbb{Z}/2\mathbb{Z})$ with values in $\mathbb{Z}/2\mathbb{Z}$. It merely keeps track of the parity of the intersections and is computed by the following rule: choose representatives for the two classes that meet transversally and count their intersection points mod two. This pairing is non-degenerate; *i.e.* the generator squares to one. Indeed, two distinct lines meet in one point.

12.11 The ovals γ of a curve C come in two flavours: the class of γ in the homology group $H_1(\mathbb{RP}^2, \mathbb{Z})$ may or may not be non-zero. The ovals in a non-zero class are called *odd ovals*. Two odd ovals must meet, since the generating class squares to one. In case they meet transversally, the intersection consists of an odd number of points. Since the different ovals of C are disjoint, we infer that at most one of them can be odd. Moreover, the number of real points where a general line meets C, has the same parity as the degree $\deg C$; so a curve of odd degree has exactly one odd oval, whereas one of even degree has none.

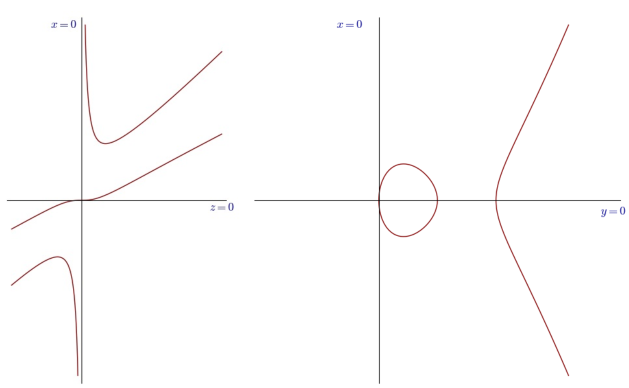
Odd ovals
odde ovale

The ovals of C that are not odd, are said to be *even*. Their classes in $H_1(\mathbb{RP}^2, \mathbb{Z})$ vanish, and consequently they meet any curve in an even number of points: if the curves meet transversal, they have an even number of common points (which might be zero), and if they meet in an odd number of points, the curves are forced to be tangent somewhere.

Even ovals
jevne ovaler

EXAMPLE 12.12 *(Quadrics)* A quadric Q has always one even circuit, which one sees perfectly when Q intersects the line at infinity in two non-real points; the quadric then appears as an ellipse in the corresponding standard affine piece $\mathbb{R}^2$. If the finite part is a parabola — which happens when the line at infinity is tangent to Q — the picture is also convincing. However, when Q has two real points at infinity, the finite part is disconnected: it will be a hyperbola, which has two branches: if you travel along one branch and reache infinity approaching one asymptote, you will reappear at the other end of that asymptote and continue you journey along the other branch, eventually reaching infinity again, but approaching the other asymptote. ☆

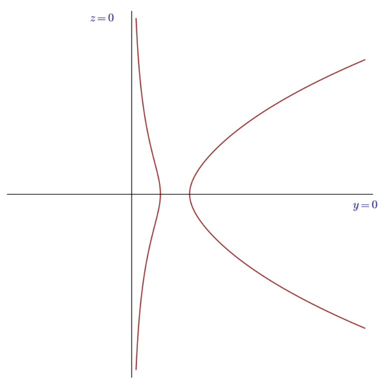
EXAMPLE 12.13 In general, an oval that meets the line at infinity will when drawn in $\mathbb{R}^2$ — that is, in the real points of the distinguished open set $D_+(z)$ — not appear as a simple closed curve, but as a curve with asymptotes. Odd ovals behave always like this; they must meet any line, in particular the line at infinity. The asymptotes correspond to points at infinity of the curve, and since the line at infinity disconnects the curve when there are more than one intersection point, the finite part of the curve will then have several branches.



In the figure we have depicted the real points of the elliptic curve C given as

$zy^2 = x(x-z)(x-2z)$ in two different perspectives. To the left we have drawn the part in the distinguished open set $D_+(y)$. The line at infinity is $y = 0$, which C meets in the three points $(0:0:1)$, $(1:0:1)$ and $(2:0:1)$. There are two circuits, one with the two asymptotes $z = 0$ and $z = 2x$, and one with the asymptote $z = x$. To the right we see the familiar picture of the cubic in $D_+(z)$; there is one even oval not meeting the line at infinity, and one having triple contact with it at the point $(0:1:0)$.

Then, of course, one wonders what C looks like in the affine piece $D_+(x)$; as the picture below shows, there are two infinite circuits. One is odd and has the y-axis as asymptote, and the other is even and is tangent to the the line $x = 0$ at infinity in $(0:0:1)$. ☆



Proof of Harnack's theorem: Let r be the number of ovals C has, and assume that r exceeds the Harnack bound; that is, $r > (d-1)(d-2)/2+1$, and as in the proof of Proposition 12.7 we let $N = (d-1)(d-2)/2+1$. Let $\gamma_1, \ldots, \gamma_r$ be the ovals of C, arranged in a way that at least the $r-1$ first $\gamma_1, \ldots, \gamma_{r-1}$ are even (there is at most one odd one). As noted above, the proof uses an auxiliary curve which we find as follows: pick a point in each of the ovals $\gamma_1, \ldots, \gamma_N$, and chose $d-3$ additional points in the oval γ_r. As in the proof of Proposition 12.7, we may find a curve X of degree $d-2$ passing through all these points. The topological considerations above tell us that for each of the even ovals $\gamma_1, \ldots, \gamma_{r-1}$ either X meets it in two points or is tangent to it, in both cases the contribution to the total intersection number is at least two. Bézout's theorem then gives the contradictory inequality:

$$d(d-2) = \sum_{p \in C \cap X} \mu_p(C, X) \geqslant 2N + d - 3 = (d-1)(d-2) + 2 + d - 3 = d(d-2) + 1,$$

and that's it. ❑

12.14 We now return to second natural question we asked about $C(\mathbb{R})$: What are the possible configuration the even ovals can have? How can they be nested? The complement $\mathbb{RP}^2 \backslash \gamma$ of an even oval has two connected components. One which is homeomorphic to a disk, and the other which is homeomorphic to the Möbius band. This is obvious if γ is a parallel close to the equator; *i.e.* a great circle slightly moved downwards (or upwards). What lying it within an oval in the finite part $\mathbb{R}^2$ means, is obvious, and due to the different topologies of the components of the complement, it is still meaningful for even ovals in $\mathbb{RP}^2$: one says that an even oval γ' lies within another one γ if γ' is contained in the disk-like component of the complement of γ. One also says that the two are *nested*; more generally, a sequence $\gamma_1, \ldots, \gamma_r$ of even ovals is said to be *nested* if each γ_i lies within the preceding one. The number r is often called the *depth* of the nesting.

PROPOSITION 12.15 *If C be non-singular real curve of degree d, then the maximal depth of a nested sequence of even ovals of $\mathbb{C}(\mathbb{R})$ is $d/2$.*

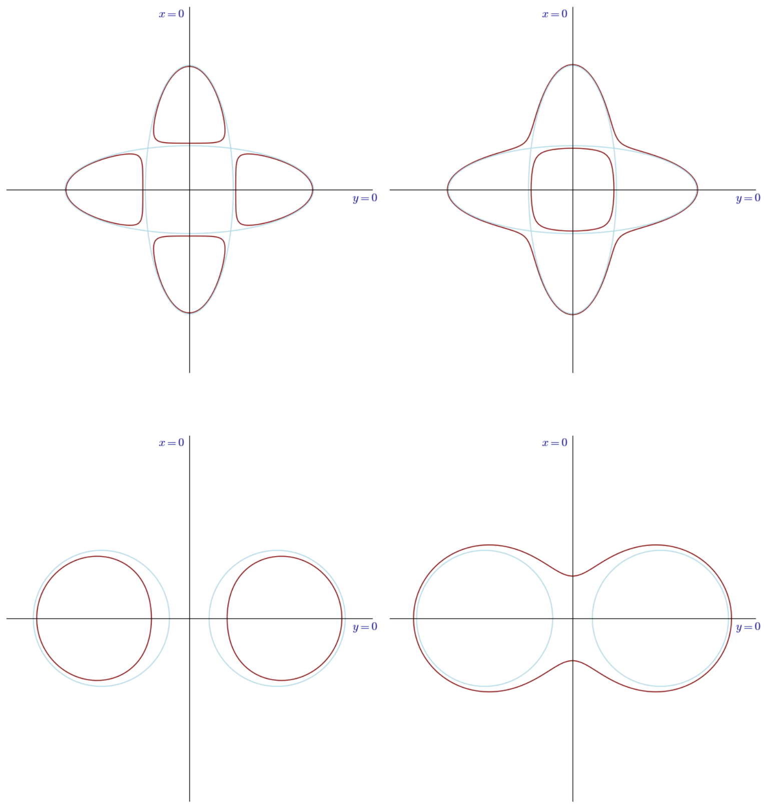
PROOF: Let $\gamma_1, \ldots, \gamma_r$ is the nested sequence of even ovals belonging to C. Pick a point x lying within γ_r, the smallest of the γ_i's, and a point y outside γ_1, the largest. The point x lies within and the point y lies outside all the γ_i, hence a line L connecting x and y intersects each γ_i, and the γ_i's being even, they meet in at least two points. It follows that $\#C \cap L \geqslant 2r$, so by Bezout, $2r \leqslant d$. ❑

EXAMPLE 12.16 The different configuration of the ovals of a quartic are as follows. There is at most four by Harnack's theorem, and there can 1,2,3 and 4, and, of course, a quartic need not have any real points at all. No more than two can be nested, and if two are, they are the sole ovals. Indeed, if there were a third, one could draw a line from a point within it to a point within the smallest of the two nested ones, and this line would meet C in at least six points, which contradicti Bézout's theorem. And this exhausts all possibilities.

The following examples are all found by considering two ellipses (or two circles) and slightly perturbing the equation of their union. The ellipses divide the plane into several regions, and to figure out the distribution of ovals, boils down to an analysis of the signs the two quadratic forms that define the ellipses, have in the different regions.

The two first share the pair off ellipses, but the perturbations are of opposite sign; the equations are $Q = (y^2 + 8x^2 - 2)(8y^2 + x^2 - 2) + \epsilon = 0$ with $\epsilon > 0$ for the one to the right and $\epsilon < 0$ for the one to the left: To exhibit the next two, examples of quartics with two even ovals and with just a singel one, we use two disjoint circles. The equations of the quartics are of the form $(y^2 + (x-1)^2 - r)(y^2 + (x+1)^2 - r) + \epsilon = 0$ with r a number less than one, and ϵ small in absolute value. A negative ϵ gives the curve to the left with two ovals, and when ϵ is positive, one obtains a curve with just one oval: Finally, to have an example with three ovals, we resort to to ellipses again. The equation is $(x^2 + 10(y^2 - y))(6x^2 + (y-2)^2 - 4) + \epsilon = 0$ with $\epsilon > 0$. ☆

EXERCISE 12.2 Since the proofs of Harnack's bound and the bound on the number of singular points are so similar, it is tempting to combine them. For



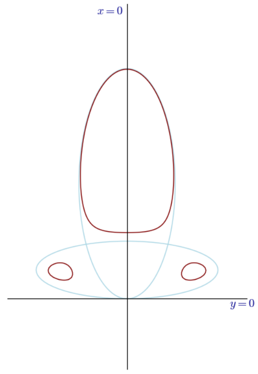
simplicity we shall assume that C is a real irreducible curve in $\mathbb{P}^2_{\mathbb{C}}$ all whose singularities are so-called *crunodes*; they are ordinary double points with complex conjugate tangents (may be this is the empty set, but any how!). In the real picture such singularities appear as isolated points of $C(\mathbb{R})$. Let g denote the difference between $(d-1)(d-2)/2$ and the number of double points. Show that the number of ovals of C is at most $g+1$. This is in fact a general statement about the real part of any projective non-singular curve with g being an invariant of the curve called the genus. ★

Inflexion points and the Hessian

In early calculus courses one meets the notion of points of inflexion of a function, points where the curvature of the graph changes sign or where the tangent moves from one side of the graph to the other. There is an analogue concept in the theory of plane curves called a *flex* or an *inflexion point* of a curve. This is a regular point on C where the tangent line T to C at p has a order of contact superior to two; that is, points where $\mu_p(T,C) \geqslant 3$. When equality holds, one says that flex is *ordinary*.

Flexes
fleks
Inflexion points
infleksjonspunkter,vendepunkter

In school we learned that a necessary condition for a point on a graph of a



curve to be a flex is that the double derivative of the function vanishes there. For plane algebraic curves C there is a similar criterion, which involves a certain locally defined 'double derivative'. Flexes on curves in $\mathbb{P}^2$ are even described by a *global* criterion; there is another curve in $\mathbb{P}^2$, the *Hessian* of the given curve C, which cuts out the flexes on C.

The Hessian curve
den Hessiske kurven, Hesse-kurven

12.17 Let F be the homogeneous polynomial with C as zero locus and let $(x_0 : x_1 : x_2)$ be homogeneous coordinates on $\mathbb{P}^2$. The equation of the Hessian is the determinant of the matrix $(F_{x_i x_j})_{ij}$ of the double derivatives of F. It will be denoted H_F and is named the *Hessian determinant* or simply the *Hessian* of F. That is:

$$H_F = \begin{vmatrix} F_{x_0x_0} & F_{x_0x_1} & F_{x_0x_2} \\ F_{x_1x_0} & F_{x_1x_1} & F_{x_1x_2} \\ F_{x_2x_0} & F_{x_2x_1} & F_{x_2x_2} \end{vmatrix}$$

Clearly the determinant depends on the coordinates used, but the drudgery of twice applying the chain rule shows that $H'_F = (\det A)^2 H_F$ where A is the matrix of the coordinate change, and the prime indicates that the Hessian is computed with respect to the new coordinates. So the curve $Z_+(H_F)$ is canonically associated to C; we shall denote it by H_C.

12.18 The Hessian H_F is a homogeneous form of degree $3(d-2)$ if the curve C is of degree d. For $d = 1$ all the double partials vanish, so $H_F = 0$, and for conics the Hessian is equally uninteresting being a non-zero scalar only defined up to scaling. So cubics are the first interesting case; then the Hessian will we a cubic as well, so a new cubic canonically associated to the one given emerges!.

12.19 There is a form of the Hessian determinant that significantly makes dehomogenization easier and which therefore is well adapted to a local analysis. Singling out one of the coordinates, say x_2, one may bring the Hessian on the following form

$$x_2^2 \cdot H_F = \begin{vmatrix} F_{x_0x_0} & F_{x_0x_1} & (d-1)F_{x_0} \\ F_{x_0x_1} & F_{x_1x_1} & (d-1)F_{x_1} \\ (d-1)F_{x_0} & (d-1)F_{x_1} & d(d-1)F \end{vmatrix} \tag{12.3}$$

Indeed, using the Euler formula $x_0 G_{x_0} + x_1 G_{x_1} + x_2 G_{x_2} = \deg G \cdot G$, valid for any homogeneous polynomial G, and performing appropriate elementary row and column operations gives the equality.

Note that this reveals that in case the characteristic of k divides $d-1$, the Hessian vanishes identically, and is of no use. For instance, this happens if C is a cubic and the characteristic is two. As well it reveals that the Hessian vanishes in each singular point of C as the entire lower row vanishes (or rightmost column) there.

12.20 Let $C \subseteq \mathbb{P}^2$ be plane curve and let p be a point on C. We aim at explaining that the Hessian of C measures wether p is a flex or not. We shall work with homogeneous coordinates $(x_0 : x_2 : x_3)$ such that $p = (0:0:1)$ and, to ease the notation, we shall denote the induced coordinates on $D_+(x_2) = \mathbb{A}^2$ by $x = x_0 x_2^{-1}$ and $y = x_1 x_2^{-1}$. Moreover, the coordinates should be chosen in a way that the tangent of C at p is the x-axis $y = 0$. Under these conditions, the equation f of C may easily be brought on the form

$$f = y \cdot a(x,y) + x^2 b(x), \tag{12.4}$$

where $a(0,0) \neq 0$, and we have:

Lemma 12.21 *The point p is a flex if and only if $b(0) = 0$; that is, if and only if the local equation is of shape $f = y \cdot a(x,y) + x^3 c(x)$.*

Proof: The local intersection number of the curve and the tangent is given as $\dim_k k[x,y]/(y,f)$. As $k[x,y]/(y,f) = k[x]/(x^2 b(x))$, this vector space dimension exceeds two precisely when $b(0) = 0$. ❑

Proposition 12.22 *Let $C \subseteq \mathbb{P}^2$ be a curve of degree d and assume that the characteristic of k does not divide $d-1$ and is not two. Then a non-singular point p on C is a flex if and only if the Hessian vanishes at p; that is, if and only if $H(p) = 0$.*

Proof: To compute the dehomogenized Hessian h we use the form from (12.3) with $x_2 = 1$, $x_0 = x$ and $x_1 = y$, and find

$$h = \begin{vmatrix} f_{xx} & f_{xy} & (d-1)f_x \\ f_{xy} & f_{yy} & (d-1)f_y \\ (d-1)f_x & (d-1)f_y & d(d-1)f \end{vmatrix} \tag{12.5}$$

From (12.4) we decuce, after some trivial computations which give $f_x(0,0) = 0$, $f_{xx}(0,0) = 2b(0)$ and $f_y(0,0) = a(0,0)$, that

$$h(0,0) = (d-1)^2 \begin{vmatrix} 2b(0) & \star & 0 \\ \star & \star & a(0,0) \\ 0 & a(0,0) & 0 \end{vmatrix} = 2(d-1)^2 b(0) a(0,0)^2, \tag{12.6}$$

where the values of star-marked entries are irrelevant. Since $d-1 \neq 0$ and $a(0,0) \neq 0$, we infer that the Hessian vanishes at p if and only if $b(0)=0$; that is, if and only if p is a flex. ❑

Corollary 12.23 *Under the assumption that the characteristic of k does not divide $d-1$ and is not two, every non-singular plane curve has an inflexion point. The number of flexes, when finite and counted with appropriate multiplicities, is equal to $3d(d-2)$.*

It may happen that all points of a curve C are flexes. This is however a phenomenon occurring merely over fields of positive characteristic — in characteristic zero the Hessian does not vanish identically, and hence the number is finite. An eye-opening example is the curve with equation $F = x^p y + y^p z + z^p x$. In characteristic p the partial derivatives of F are $F_x = z^p$, $F_y = x^p$ and $F_z = y^p$. They do not have common zeros in $\mathbb{P}^2$, so the curve is non-singular. The Hessian vanishes identically, but as the degree is $p+1$, the link between the Hessian and the flexes is broken, and we have to compute intersection numbers. It turns out that if $p \geqslant 3$ all points on C will be inflexion points, but if $p=2$, there will be only 9.

Let us begin by seeing what takes place in the affine piece $D_+(z)$. There the dehomogenized equation is $f = x + y^p + x^p y$, and if (a,b) is a point on the curve, performing the coordinate change $u = x-a$ and $v = y-b$ brings f on the form $u + a^p v + u^p v + v^p + bu^p$ To ease notation, let $\alpha = -a^p$. The tangent equation at (a,b) is $u - \alpha v$, and we find

$$\mathfrak{a} = (u-\alpha v, u-\alpha v + u^p v + v^p + bu^p) = (u-\alpha v, v^p(1+\alpha^p b + \alpha^p v)).$$

Hence in the points where $\alpha^p b \neq -1$ it holds that

$$\mu_q(C,T) = \dim_k k[u,v]/(u-\alpha v, v^p(1+\alpha^p b+\alpha^p v)) = \dim_k k[v]/v^p = p.$$

If $p \geqslant 3$, the point q is thus a flex. However, when $p=2$, it is an ordinary tangent! To be a flex in that case, the equation $a^{p^2} b = 1$ must be satisfied. Combined with the relation between a and b imposed by the equation of the curve, the condition becomes $a^9 + a^6 + 1 = 0$. This a separable polynomial and hence C has nine flexes.

Proof of Corollary 12.23: If the number of inflexion points is finite, the Hessian does not vanish identically. Its degree being $3(d-2)$, the corollary then follows directly from Bézout's theorem (non-singular curves are irreducible, and C is obviously not a component of the Hessian). ❑

Corollary 12.24 *A non-singular cubic curve in $\mathbb{P}^2$ over an algebraically closed field whose characteristic is not two or three, has exactly nine points of (ordinary) inflexion.*

In fact, with other methods one can prove that cubics over fields of characteristic two still have nine flexes (as in the example above), in characteristic three however, there will either be only or only three (see Exercise 12.26 below).

PROOF: The Hessian of C is of degree three, and so by Bézout's theorem it holds true that

$$9 = \deg C \cdot \deg H_C = \sum_p \mu_p(C, H_C),$$

so our task will be to see that C and H_C meets transversally, and this requires a closer local analysis. At a flex of our cubic curve the local equation (12.4) takes, after a possible scaling of the coordinate x, the form $f = x^3 + ya(x,y)$ where the function a is a quadratic polynomial. We are searching after the component along dx of the differential dh of the local expression (12.5) for the Hessian. Computing modulo (x^2, y) — that is, proceeding as usual, but with the luxury of discarding all terms of degree two and all those containing y — and using (12.5) we arrive at the equality

$$h \equiv \begin{vmatrix} 6x & a_x & 0 \\ a_x & 2a_y & a \\ 0 & a & 0 \end{vmatrix} = 6xa^2. \tag{12.7}$$

As longs the characteristic is not two or three, the linear part of h, *i.e.* the differential dh, has the non-zero term $6a(0)^2 dx$ modulo dy, and since dy is the differential of the equation of the curve, the Hessian is transversal to C at p. ❑

EXAMPLE 12.25 *(Flexes of the Fermat cubic)* Let us find the flexes of the Fermat cubic C given by $x^3 + y^2 + z^3 = 0$ when k is of characteristic distinct from two and three. The Hessian of C is trivially found to be

$$H = \begin{vmatrix} 6x & 0 & 0 \\ 0 & 6y & 0 \\ 0 & 0 & 6z \end{vmatrix} = 216 \cdot xyz.$$

The inflexion points are therefore the intersections point of the lines $Z_+(x)$, $Z_+(y)$ and $Z_+(z)$ with C. Setting *e.g.* $z = 0$ in the equation for C yields $y^3 = -x^3$. Hence the line $Z_+(z)$ meets C in the three points $(1 : -1 : 0)$, $(1 : \eta : 0)$ and $(1 : \eta^2 : 0)$ where η is primitive a cube root of -1. By symmetry, one finds the six other flexes by permuting the coordinates; they are the points $(1 : 0 : -1)$, $(1 : 0 : \eta)$, $(1 : 0 : \eta^2)$, $(0 : 1 : -1)$, $(0 : 1 : \eta)$ and $(0 : 1 : \eta^2)$.

Note that only three of the nine are real, and that the three real ones are collinear—they lie on the line $x + y + z = 0$.

In characteristic two the Fermat cubic still has nine flexes: Factoring for example $y^3 + z^3$ in the product of three linear forms, one infers, *e.g.* by Lemma 12.21, that the intersection points between each line given by one of these factors and $Z_+(x)$ will be flexes[2]. If the characteristic is two, the actual factorization is $z^3 + y^3 = (z + y)(z + \epsilon y)(z + \epsilon^2)$ where ϵ is a solution of the equation $t^2 + t + 1 = 0$. (The field $\mathbb{F}_4$ with four elements has the elements $0, 1, \epsilon, \epsilon^2$, and $\epsilon^2 = \epsilon^{-1} = \epsilon + 1$, so the flexes are defined over $\mathbb{F}_4$). ☆

[2] Of course, this methods works as well in all other cases.

EXAMPLE 12.26 *(Flexes in characteristic three)* The proposition does not hold in characteristic three as shows the curve C given by $f = (x + y + z)^3 - xyz = 0$. One finds $f_x = -yz$, $f_y = -xz$ and $f_z = -xz$. In a common zero of three the first partial two of the coordinates must vanish, so the point can not lie on C as then the third coordinate would vanish as well. So C is non-singular. An easy computation shows that the Hessian equals xyz. So the flexes lie where either $Z_+(x)$, $Z_+(y)$ or $Z_+(Z)$ meets C; that is, in one of three points $(0 : -1 : 1)$, $(-1 : 1 : 0)$ or $(1 : 0 : -1)$. At each point the intersection multiplicity equals three.

The example in the previous paragraph is not quit prototypical. There are also non-singular cubics in characteristic three with just a single flex — they are called *supersinguar* curves — but that is all what can happen, there are either one or three flexes (See Exercise 12.4 below). For instance, the curve C with equation $F = zy^2 - (x^3 + axz^2 + bz^3) = 0$ with $a \neq 0$ is of the kind. One finds $F_x = -2az^2$, $F_y = 2yz$; it ensues that

$$z^2 H = \begin{vmatrix} 0 & 0 & -4az^2 \\ 0 & 2z & \star \\ -4az^2 & \star & \star \end{vmatrix} = -32z^5$$

and hence that $H = z^3$. The line at infinity $Z_+(z) = 0$ meets C only at $(0 : 1 : 0)$, so this is the only flex C has. ☆

12.27 The configuration of the flexes of cubics has the following remarkable property—which in fact holds in any characteristic.

PROPOSITION 12.28 *On a line that passes through two of the flexes of C lies a third flex.*

PROOF: Choose homogenous coordinates $(x : y : z)$ on $\mathbb{P}^2$ so that the two flexes are $p = (0 : 0 : 1)$ and $q = (0 : 1 : 0)$ and such that C has the tangent line Z_+(y) at p and $Z_+(z)$ at q. By the description in Lemma 12.21 of the local equations at a flexe, the cubic homogeneous equation of C is, because p is a flex, shaped like $\alpha \cdot x^3 + yA(x, y, z)$ where $A(x, y, z)$ is a quadric form and $\alpha \in k^*$, and since q is a flex, it is also on the form $\alpha \cdot x^3 + zB(x, y, z)$ with B a quadric. It follows that there is a linear form w, proportional to neither y nor z, so that $a = zw$ and $b = yw$. The equation then takes the form $x^3 + yzw$, and the intersection of the line $Z_+(w)$ with $Z_+(x)$ will be a third point of inflexion. ❑

12.29 It is interesting that such a configuration of nine points with each triple collinear can not exists in the Euclidean plane $\mathbb{R}^2$. This is a consequence of the so-called Sylvester–Gallai theorem that states that for any finite set of points in $\mathbb{R}^2$ either all lie on a line or there is a line passing by exactly two of the points. This entails that all the nine (complex) inflexion points of a *real* cubic curve cannot be real (*i.e.* have real coordinates).

Exercises

12.3 Assume that C is real cubic. Show that C either has one or three real inflexion points (in fact, one may show that the case of a single flex does not occur). HINT: Show first that if a line intersects C in two real points, the third intersection point is real as well, then appeal to the Sylvester–Gallai theorem.

12.4 Show that a non-singular cubic in characteristic three either has one or three flexes. HINT: If it has more than three flexes, it has more than six, and by (12.7) the Hessian has local intersection numbers with C exceeding two in each flex. Conclude by Bézout's theorem.

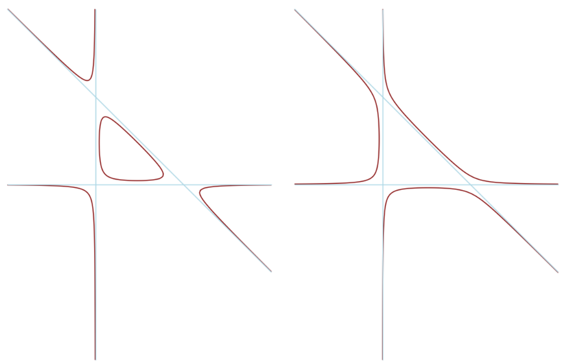
★

12.3 *Plane cubics*

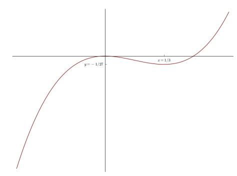
EXAMPLE 12.30 In the figure below two other version of cubics drawn in $D_+(z)$. Their equations are shaped like $xy(x+y-z)+\epsilon z^3$ where ϵ is small in absolute value; the cubic to the left has $\epsilon > 0$ and the one to the right $\epsilon < 0$. Starting with the equation a reducible curve — in this case the union of three lines — and perturbing the equation slightly is a fruitful way of constructing real curves with many circuits.

This family of curves has several interesting aspects, and we shall later show that almost every plane non-singular curve may be expressed in the form, and that the exceptions are easily classified (there is basically but one). One feature is that the parameter ϵ discriminates between curves having a single oval and the bicircuited ones. And the limt value is $\epsilon = 1/27$. For this value C_ϵ is singular, with an isolated singularity at $(1:1:3)$. For $\epsilon > 1/27$ the curve has two ovals, and when ϵ approaches $1/27$ the even oval shrinks to the singular point $(1:1:3)$ and disappers (it is swallowed by the singularity, as a cosmologist would say) and for $\epsilon < 1/27$ there is only one.

Expounding on this, we begin by locating the singular members of the pencil. The dehomogenized equation of the curve is $xy(x+y-1)+\epsilon = 0$. The partials are $f_x = 2xy + y^2 - y$ and $f_y = 2xy + x^2 - x$, and one easily shows that $(0,1)$, $(1,0)$, $(1,1)$ and $(1/3,1/3)$ are the only of their common zeros that can lie on the curve. The three first do when $\epsilon = 0$, and the last if $3^{-2}(3^{-1}-1)+\epsilon = 0$; that is when $\epsilon = 1/27$. The distribution of the signs of the product $xy(x+y-1)$ prohibits a second and even component when $\epsilon < 0$. Now, if $\epsilon > 0$, intersect the curve with the line L where $y = x$. The curve is symmetric about L, so if L did not intersect an even oval, the oval would lie entirely in one of the connected



components of L^c, and would have a mirror image in the other component, which is impossible since C has at most one even oval; we infer that L must intersect a non-empty even oval. The intersection is given as $x^2(2x-1) = -\epsilon$, but the function $2x^3 - x^2$ has a local minimum at $x = 1/3$ with value $-1/27$, hence it attains each value less than $-1/27$ just once. That is, for $\epsilon > 1/27$ there is no second intersection between the curve and the line $y = x$; *i.e.* no even oval.



The graph of $2x^3 - x^2$

☆

Real flexes of a real cubic

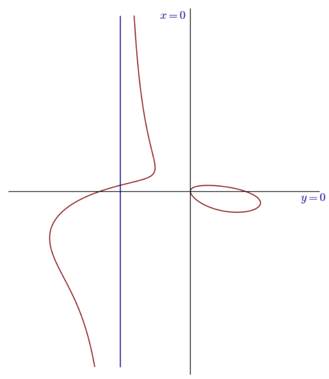
Time has come to finish the story about the real flexes of a real cubic. A real cubic has, as we have seen, at most three flexes, and in fact, as we are about to see, the number is always three. Due to Proposition 12.28 it will suffice to find two, and the three will be collinear. The proof we shall give is quit elementary, almost at calculus level.

The cubic curve will be denoted C, it is real and non-singular. We begin with choose appropriate homogenous coordinates $(x : y : z)$ in $\mathbb{P}^2_{\mathbb{C}}$: the line at infinity $Z_+(z)$ shall meet C in just one real point $(0 : 1 : 0)$, and the y-axis $Z_+(x)$ shall be tangent to C at $(0 : 1 : 0)$.

The curve has one or two real ovals, but only one, the odd one, meets the line at infinity. We let γ be the portion of the odd oval contained in $\mathbb{R}^2$, the real points of the affine piece $D_+(z)$. It is a one-dimensional, non-compact and connected real manifold, and so is diffeomorphic to the real line $\mathbb{R}$, and we choose an injective regular parameterization $\phi(t) = (x(t), y(t))$ of γ.

The y-axis is an asymptote of γ and $\phi(t)$ tends to $(0 : 1 : 0)$ when t tends to ∞ or $-\infty$, that is, $y(t)$ tends to $\pm\infty$ when $t \to \pm\infty$. We choose the orientation so that $\lim_{t\to\infty} y(t) = \infty$.

The y-axis meets C twice at infinity and hence by Bézout, it has precisely one point q in common with γ in $\mathbb{R}^2$. Thus the y-axis divides the circuit γ in two parts, where $y(t) > 0$ and $y(t) < 0$.



THEOREM 12.31 *A non-singular real cubic curve has exactly three inflexion points. They are collinear and lie on the odd oval.*

PROOF: With the set up as above, consider the x-coordinate $x(t)$. It tends to zero when $t \to \pm\infty$, and hence $x(t)$ has a maximum and a minimum. By Bézout's theorem a vertical line $x = a$ cuts the curve in at most two points since it passes by $(0 : 1 : 0)$, and this shows that x has a unique local maximum, which also is global, say at t_0. Similarly, $x(t)$ has a unique local minimum, which also is global, say at t_1. By symmetry we may assume that $t_0 > t_1$.

By Bézout's theorem, every tangent to γ meets γ in exactly one more point $g(t)$ (which can lie on the line at infinity and coincides with the point of tangency precisely at the flexes). This is necessarily a real point, and $g(t)$ is a continuous function of t where it is defined. It has 'poles' where the tangent meets γ at infinity: since the tangent is real, the intersection point must be a real point; that is, it equals $(0 : 1 : 0)$. This happens exactly when the tangent is parallel to the y-axis, and this occurs precisely in the two extreme points $\phi(t_0)$ and $\phi(t_1)$ of the x-coordinate. Observe that $g(t) \to q$ when $t \to \pm\infty$.

Since the unit tangent $\tau(t)$ of γ is oriented, we may measure the distance from $\phi(t)$ to $g(t)$ algebraically (*i.e.* with a sign, positive when $g(t)$ lies in "front" of $\phi(t)$ and negative when it lies behind), and we call the distance $d(t)$. This is a differentiable function of t that vanishes precisely at the flexes; moreover it tends to infinity whenever $\phi(t)$ or $g(t)$ tends to infinity.

In fact, $d(t) \to -\infty$ as $t \to \infty$, since then $g(t) \to q$. So if we can show that $d(t) \to \infty$ when t approaches t_0 from above, $g(t)$ has to vanish for some $t > t_0$, and the oval g will ahve a flex somewhere between $\phi(t_0)$ and infinity. A symmetric consideration gives a flex with parameter value less than t_1, and we will be through.

The crucial points is that the y-coordinate $y(t)$ is increasing near t_0. So see this, introduce the region Ω bounded by the portion of γ where $t \leqslant t_0$ and the portion of the line $x = x(t_0)$ with $y \leqslant y(t_0)$ (the union of the two is a simple closed curve and divides the plane in two connected components). Now, if $y(t)$ did decrease, the point $\phi(t)$ would enter Ω, but $\phi(t)$ cannot reexit Ω without crossing γ because $x(t) < x(t_0)$ when $t \neq t_0$. So since γ has no self-intersection, $\phi(t)$ cannot reach $(0 : 1 : 0)$ through points not in Ω, *e.g.* those with a large positive y-coordinate, and we have a contradiction.

From calculus we know that the tangent to a graph at points sufficiently near a maximum do not cross the graph but possibly far away from the maximum, and we infer for t vicinal to t_0 the third point where the tangent at $\phi(t)$ meets γ, has a parameter value greater than t_0. Hence, in view of $y(t)$ being increasing, $g(t)$ lies in front of $\phi(t)$, and $d(t)$ is positive. ❑

Elliptic curves – the Hesse pencil

Plane cubic curves are among the most studied objects in algebraic geomtry and there arithmetic properties are extremely deep and important (for instance they lie at the ground of Wiles proof of Fermat's last theorem). The steps in their study is to establish so-called norma forms. That is, with a projective equivalence (or a choice of homogenous coordinates) one brings their equations on a particular and simple form. We shall describe the two most frequently met, the Hesse normal form, and the Weierstrass normal form.

12.32 [Hesse normal form The first normal form we describe is named after Otto Hesse and called the Hesse normal form. Basically, one moves the curve so as to "standarize" three of the inflections points; they will be located at the reference points $(0 : 0 : 1)$, $(0 : 1 : 0)$ and $(1 : 0 : 0)$. The normal form depends on one parameter m that governs the inflexionary tangents at the three points. The starting point is the following

LEMMA 12.33 *Let L be a line through three inflexion points of C, and assume that the three corresponding tangents are not concurrent. Then C can be brought on the form*

$$(x + y + z)^3 + 6\kappa xyz = 0. \tag{12.8}$$

PROOF: Chose coordinates so that $z = 0$ is the equation of the line, and that $x = 0$ and $y = 0$ are equations of two of the infelxionary tangents. In the course of the proof of Proposition 12.28, we found an equation $z^3 + axyw = 0$ for C where the linear form $w = ax + by + cz$ defines the tangent at the third flex on the line. As these three tangent are not concurrent by assumption, $c \neq 0$, and we may use $c^{-1}w$ as a new coordinate. Renaming it z, and absorbing the scalars a and b in x and y, gives an equation of the desired shape. ❑

LEMMA 12.34 *In the setting of Lemma 12.8, if C is a real cubic, the coordinate change that brings its equation on the form* (12.8) *may be chosen real, and the parameter κ is uniquely defined.*

PROOF: The first statement is clear from the previous proof. The equation (12.8) is characterized by the three flexes being the points $(0 : -1 : 1)$, $(1 : 0 : -1)$ and

$(-1:1:0)$ and the flexional tangents having equations $x=0$, $y=0$ and $z=0$. So the coordinate system is determine up to a permutation of x, y and z by the curve, and each such permutation leaves the (12.8) invariant, hence also κ. ❑

Among the real C is a real cubics up to projective equivalence just one has concurrent flex-tangents, and in appropriate coordinates, it is given by $z^3 - yx(y-x)$.

PROPOSITION 12.35 (THE HESSE PENCIL) *Assume that k does not have characteristic 3, and let $C \subseteq \mathbb{P}^2$ be a non-singular elliptic curve. In suitable homogeneous coordinates, the equation of C then takes the form*

$$x^3 + y^3 + z^3 + 6mxyz = 0,$$

where m is a scalar.

There are several proofs of this; they all rely on a cubic extension of the field of coefficients of the cubic (*i.e.* the smallest subfield of k containing the coefficients of an equation for C). The one we shall present is well adapted for real cubics, and it follows more or less the proof described in Harold Hilton's classic book "Plane algebraic curves" from 1920. It consists of a sagaciously chosen change of coordinates involving solutions of a cubic equation (which persist being real when the cubic is real) and some trivial but tedious manipulations (not too frightening when well organized). The method applied yields the following (see also the paper[3] by Bonifant and Milnor.)

3

PROPOSITION 12.36 *When $k = \mathbb{C}$, the curve is real if and only if m is real. Among the real curves, the parameter m unambiguously determines the curve. The curve has one oval for $m < -1/2$ and two when $m > -1/2$.*

PROOF: The parameter κ is unambiguously defined by C, and m in its turn is uniquely determined by κ. Refering back to Example 12.30 we checked that the border between mono-curcuited cubics and bicurcuited ones is at $\kappa = -9/2$, which in the relation (12.9) between κ and m below corresponds to $m = -1/2$. ❑

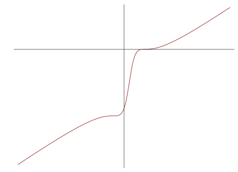
One ways is, clear and the other will follow if we can bring C on Hesse form with a real projective transformation.

PROOF: The starting point is the equation (12.8) for C (the case with three concurrent inflexional tangents is treated apart), in which one performs a change of variables of the form

$$\begin{aligned} x &= -2mu + v + w \\ y &= u - 2mv + w \\ z &= u + v - 2mw \end{aligned}$$

with m being a parameter to be chosen. The determinant of the corresponding matrix equals $-2(2m+1)^2(m-1)$ so the coordinate change is lawful for $m \neq$

[4] One easily checks by studying the function $(x-1)^3/(4x^2-2x+1)$ that there is just one real root



$-\frac{1}{2}, 1$. The parameter m is chosen as the real root[4] of

$$4(m-1)^3 = \kappa(4m^2 - 2m + 1). \quad (12.9)$$

Then $(u+v+w)^3 = 8(1-m)^3(u+v+w)^3$. We shall identify the coefficients in the new equation. Since the substitution is symmetric in u, v an w, symmetric terms have equal coefficient. For pure the cubic terms; *e.g.* u^3, we find a contribution $8(1-m)^3$ from $(x+y+z)^3$ and $-12\kappa m$ from xyz; all together $-2(2m+1)^2$.

For terms consisting of a square and a linear factor; *e.g.* like u^2v, the cube $(x+y+z)^3$ contributes $3 \cdot 8(1-m)^3$ and the contribution from xyz equals $6\kappa(4m^2 - 2m+1)$, and by (12.9), that coefficinet vanishes and no such term appears (which explains the the choice of m in (12.9)).

Finally, the term uvw: the contribution from $(x+y+z)^3$ is $6 \cdot 8(1-m)^3 = -12\kappa(4m^2-2m+1)$ and from xyz it is $-6\kappa(8m^3+6m-2)$ the total is $-6\kappa m(2m+1)^2$. Putting all this together, we findt

$$(x+y+z)^3 + 6\kappa xyz = -2(2m+1)^2(u^3+v^3+w^3+6muvw)$$

It remains to treat the real case with the three inflexional tangent being concurrent. These cubics are all projectively equivalent, and an easy computation shows that the curve with equation $x^3+y^3+z^3+6xyz = 0$ is of the kind.

❑

12.37 Weierstrass normal form A most frequent way of representing a cubic curve is with an equation on what is called the Weierstass normal form. One begins by placing a flex at the point $(0:1:0)$ such that the flex tangent is the line $z = 0$. The equation is then on the form

$$za(x,y,z) + x^3 = 0$$

where a is a quadratic form, or bringing all terms not containig y to the right it takes the form

$$z(y^2 + a_1yx + a_3yz) = x^3 + a_2x^2z + a_4xz^2 + a_6z^3.$$

If the characteristic is not two, we may complete the square by letting y by $y + (a_1y + a_3)/x$ which brings it on the form

$$zy^2 = G(x,z)$$

where G is a cubic form. If furthermore the characteristic different from three, one may eliminat the quadratic term in G, and we end up with the following:

$$zy^2 = x^3 + axz^2 + bz^2$$