Notes on algebraic geometry

Alex Elzenaar

November 18, 2019

One of the most difficult of the geometric concepts invented by the geometric Greeks was the geometric line. A line was said to have no thickness or breadth, and to be generated by a moving point. A point was said to have no length, breadth, or thickness, and to be indivisible. Thus, we are forced to interpret a point as a singleton $\{x\}$; i.e., a universally attracting object in the category of sets. That is simple enough. But how are we to interpret the line? [Lin71, p. 188]

This book can be used as a textbook for an introductory course in algebraic geometry. [Har77, p. vii]

Contents

Preface

Theorem. *There are no good introductory texts on algebraic geometry.*

Proof. Left to the reader. ∎

These notes are based on a course taught at the University of Auckland in Semester 2 of 2019, although a number of the basic proofs and examples differ from those taught. The primary goal of the first half is to draw an essentially straight line from the basic definitions of a first course in abstract algebra to a proof of Hilbert's Nullstellensatz (theorem 2.11.1); the second half (chapter 4 onwards) is a brief introduction to varieties in affine and projective space, and then a *very* fast introduction to Grothendeick's theory of schemes.

We have tried to be rigorous within each section, but when we use results from earlier sections we have allowed ourselves to be more informal than strictly allowed by our definitions: for example, we write $f(P)$ for the result of the evaluation of a polynomial function $f \in K[V]$ even though f is strictly not something which we may evaluate (we evaluate a representative). Our only excuse is laziness, but we hope that this laziness manifests itself as a transparent communcation of difficult ideas to the reader.

Prerequisites

Most large results from algebra which we use are stated, and most are proved. The basic definition of a group, together with the basic theorems like Lagrange's theorem, roughly included in chapter 1 of [Art91], are assumed, but commutative ring theory itself is developed in a self-contained (though extremely fast) manner — Artin is a good reference here too but a more modern and better suited book is [Alu09]. Basic linear algebra is assumed, including knowledge of concepts like dual spaces and their interactions with vector space homomorphisms. An excellent text is [Hal15a].

References and texts

The standard algebraic geometry text is [Har77]; however, the two volumes [Sha94a] and [Sha94b] are a much better introduction. The latter chapters of these notes are based on portions of these three texts. All of these texts require a good understanding of commutative algebra; there are two standard references, [AM69] and [Eis95] and by far the most accessible is the first; chapter 2 of these notes covers basically a set of edited highlights of the results of the former, though motivated and with examples from the objects of algebraic geometry. Of course, the comprehensive [Lan71] is also indispensible.

Notation

In these notes, a ring will always be commutative with unity.

$\subset, \subseteq$	the first will always denote a proper subset, the second a usual subset
$V \simeq W$	isomorphism of algebraic sets (definition 4.2.14)
$V \asymp W$	birational equivalence of varieties (definition 4.3.12)
$f\restriction_V$	the restriction of the function f to the set $V \cap \operatorname{dom} f$
$\sqrt{I}$	the radical of I (definition 2.9.1)
$\mathbb{A}^n_K, \mathbb{A}^n$	affine n-space over a field K (definition 2.4.1)
$\operatorname{codim}_W V$	codimension of W in V (definition 4.4.22)
∂f	degree of the polynomial f (definition 2.2.1)
$\delta_{i,j}$	the Kronecker delta: $\delta_{i,j} = 1$ if $i = j$, and is 0 otherwise
$\dim X, \dim_P X$	dimension of the set X, dimension of X at P (definition 4.4.4)
$D_X f$	algebraic derivative of $f \in R[X]$ (definition 4.5.8)
$\operatorname{eval}_P, \operatorname{eval}_{\phi,P}$	evaluation at P via ϕ (theorem 2.2.4)
$\operatorname{Frac} R, \operatorname{Frac}_S R$	fraction field of R, localisation of R at S (propositions 2.1.18 and 4.6.4)
$\mathrm{GL}_n(K)$	the general linear group over K^n (i.e. the group of invertible elements of $\mathrm{Mat}_{n\times n}(K)$)
ι_V	the identity morphism on V (definition 4.1.1)
$\mathcal{I}(X)$	vanishing ideal of a set $X \subseteq \mathbb{A}^n$ (definition 2.4.1)
$K(l_1, ..., l_n)$	extension field of K generated by $l_1, ..., l_n$ (definition 2.10.6)
$\mathrm{LC}(f)$	leading coefficient of f (definition 2.2.1)
$\mathrm{Mat}_{n\times n}(R)$	the set of $n \times n$ matrices with entries in a ring R
$\mathfrak{m}_{P,X}$	the set of $f \in K[X]$ such that $f(P) = 0$
$\mu_\alpha f, \mu_P(X, Y)$	multiplicity of the root α of f (definition 2.3.10)
$\mu_P(X, Y)$	multiplicity of point P as intersection of X and Y (definition 4.7.1)
$\mathbb{N}$	the natural numbers, $\{1, 2, ...\}$
$\mathbb{N}_0$	the unnatural numbers, $\{0, 1, 2, ...\}$
$[n]$	$\{1, 2, ..., n\}$ (for $n \in \mathbb{N}$)
$\mathfrak{N}(R)$	nilradical of R (definition 2.9.6)
$\pi_n(d)$	set of partitions of $d \in \mathbb{N}_0$ into a sum of n non-negative integers
$\rho(P, Q)$	Euclidean distance between P and Q
R_P	local ring of R at P (proposition 4.6.4)
$R[X_1, ..., X_n]$	set of polynomials over a ring R in n variables (definition 2.2.1)
$R[s_1, ..., s_n]$	extension ring of R generated by $s_1, ..., s_n$ (definition 2.10.6)
$S : R$	the ring S is an extension of R (example 2.10.2)
$\operatorname{Sing} X$	the singular points of X (definition 4.7.21)
$\Theta_{P,X}$	tangent space to X at P (definition 4.7.3)
$\mathrm{trdeg}_{L:K}$	the transcendence degree of L over K (proposition 4.4.11)
π_n	nth projection map from a Cartesian product
$\mathcal{V}(S)$	zero-set of a set $S \subseteq K[X_1, ..., X_n]$ (definition 2.4.1)

Dramatis personae

Emil Artin
(1898–1962)

Max Noether
(1844–1921)

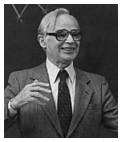
Girard Desargues
(1591–1661)

Blaise Pascal
(1623-1662)

Alexander Grothendeick
(1928–2014)

Jean-Pierre Serre
(1926–)

David Hilbert
(1862–1943)



André van der Waerden
(1903–1996)

Nathan Jacobson
(1910–1999)

André Weil
(1906–1998)

Emmy Noether
(1882–1935)

Oscar Zariski
(1899–1986)

Chapter 1

Philosophy

We shall begin by briefly motivating the study which we will undertake based on applications and relationships to other fields of 'function geometry'. We will then move to a particular set of problems which will illustrate some of the main ideas we will develop in a very concrete setting.

1.1 Motivation for study

Many problems in different fields boil down to solving polynomial equations over some ring or field.

1.1.1 Theorem (Fermat-Wiles)**.** *Let $n \in \mathbb{N}$. If there are non-zero rational numbers x, y such that*

$$x^n + y^n = 1, \tag{1.1}$$

then $n \leq 2$. ∎

This famous theorem, proposed by Fermat in the 1630s (for an account of the history see [Edw77] and [Sin97]), was finally proved by Andrew Wiles in 1994 using powerful techniques from algebraic geometry.

It is likely that another famous problem from number theory, the *abc* conjecture, will be solved using modern algebro-geometric techniques.

In applied mathematics, the study of linkages (e.g. for robot arms) utilises algebraic geometry, although here the goal is numerical calculations to actually solve systems of polynomials in $\mathbb{R}^n$ (see, e.g. [DO07]); and algebraic geometry over finite fields may be used to define and study error-correcting codes for transmissions (see, e.g. [CLO04, chapters 9 and 10]).

Alternatively, algebraic geometry may be motivated by the grounds of naturality. If $U \subset \mathbb{R}^n$ is an open set, we have the following chain of inclusions where each of the sets is a ring of functions $U \to \mathbb{R}$:

$$\begin{aligned} \{\text{functions}\} \supset \{\text{measurable}\} \supset \{\text{continuous}\} \\ \supset \{\text{differentiable}\} \supset C^\infty \supset \{\text{analytic}\} \\ \supset \{\text{polynomial}\} \supset \{\text{constant}\} \end{aligned} \tag{1.2}$$

As we move from general functions downwards, we obtain more and more well-behaved objects. Continuous functions are nicer than the average function, but can still be quite pathological — in a natural sense, almost all continuous functions on $[0, 1]$ are differentiable nowhere [Rud66, p. 114, exercise 14]; differentiable functions are much nicer, but are still only 'locally nice' (e.g. the function $x \mapsto \sin \frac{1}{x}$ on $(0, 1)$ is differentiable but oscillates wildly as one walks towards the origin so that the closure of its graph includes the entire interval $[-1, 1]$ on the y-axis). One is led to consider analytic functions (which in $\mathbb{C}$

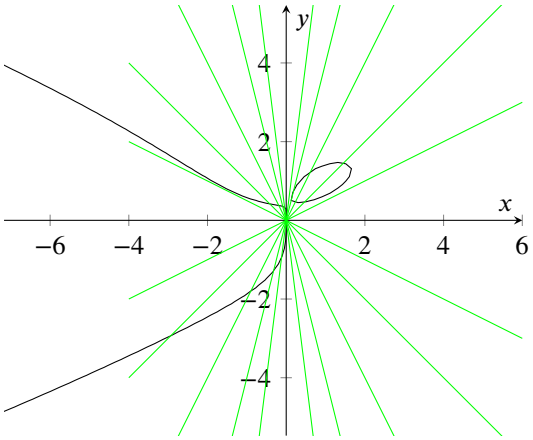


Figure 1.1: A circle and its relationship to a pencil of lines at an exterior point.

turn out to include all differentiable functions), but the problem with these is that moving away from $\mathbb{R}$ and $\mathbb{C}$ to more general number systems like the finite fields or $\mathbb{Z}$ precludes the use of limiting processes; and so naturally one is led to consider the geometry of polynomials. A significant chunk of our work will be to develop analogues of ideas from differential calculus, like tangent spaces and asymptotes, in a way that is intrinsic to polynomials over arbitrary rings.

1.2 Circles and lines

Consider the circle $(x-3)^2+y^2=4$ and the pencil of lines $y=\lambda x$ ($\lambda\in\mathbb{R}$) through the origin of $\mathbb{R}^2$, diagrammed in figure 1.1. For which values of λ do the lines become tangent to the circle?

The basic idea one immediately hits upon is to substitute the equation of the line into the equation of the circle, obtaining

$$0=(x-3)^2+\lambda^2x^2-4=(\lambda^2+1)x^2-6x+5; \tag{1.3}$$

this equation has one real root precisely when $6^2-4\cdot(\lambda^2+1)\cdot 5=0$, or in other words when $\lambda=\pm\frac{2}{\sqrt{5}}$. The coordinates of the intersection points are therefore

$$\left(\frac{5}{3},\pm\frac{2\sqrt{5}}{3}\right), \tag{1.4}$$

and the equation of the line joining them (the **polar line** of the circle with respect to the point) is $x=\frac{5}{3}$.

1.2.1 Remark. There are two main points that the reader should note:

1. The view of a tangency point as being a point where the two curves 'intersect multiply': this is a purely algebraic concept, as it depends solely on the algebra of the polynomials.
2. The operation of 'substitution', which can be more profitably thought of as (for each fixed λ) a *restriction* of the function $f(x,y)=\lambda x-y$ to the set of all points $(x-3)^2-y^2-4=0$ followed by finding the zeroes of f.

Based on the first point, we observe that we will need to have some kind of machinery for counting zeroes. In one dimension over $\mathbb{C}$ we have the fundamental theorem of algebra (theorem 2.3.13), but in

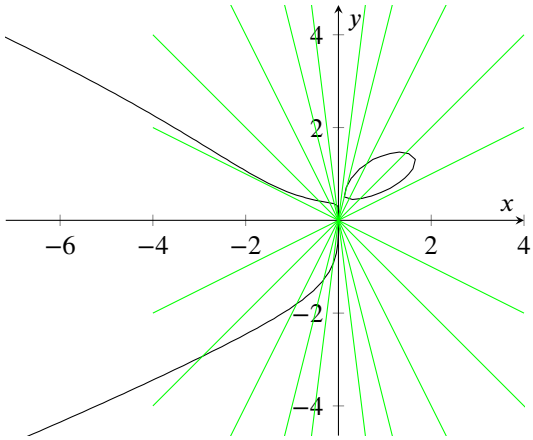


Figure 1.2: A circle and its relationship to a pencil of lines at a point on the circle.

higher dimensional spaces this will take a bit of work since the objects we are working with (polynomials) are incredibly rigid.

Based on the second point, we will need to consider not just polynomials on a field, but polynomials restricted to the zero-sets of *other* polynomials. Again, in one dimension this is easy (here we could just substitute directly), but in higher dimensions we usually cannot solve systems of polynomials by direct substitution and so we will need to use tricks or other such methods.

The fact that the given point was outside the circle was important in the above example. If the point is on the circle, then each line in the pencil of lines intersects the circle at precisely one other point; this gives us the following more general theorem (I will not bother giving the concrete example in terms of the actual equations above, but the picture is in figure 1.2):

1.2.2 Theorem. *The circle* $(x-r)^2+y^2=r^2$*, excluding the point* $(0,0)$*, is in a 1-1 correspondence with* $\mathbb{R}$ *given by*

$$\lambda \mapsto \left(\frac{2r}{\lambda^2+1}, \frac{2\lambda r}{\lambda^2+1}\right). \tag{1.5}$$

Proof. By direct substitution we have that the x-coordinates of intersections between $y=\lambda x$ and $(x-r)^2+y^2=r^2$ are given by

$$(\lambda^2+1)x^2 - 2rx = 0; \tag{1.6}$$

thus the intersection point which is not the origin has coordinates as claimed in the theorem's equation 1.5. Hence for every $\lambda \in \mathbb{R}$, the claimed point is indeed on the circle.

This map is injective, for

$$\begin{aligned}\left(\frac{2r}{\lambda^2+1}, \frac{2\lambda r}{\lambda^2+1}\right) = \left(\frac{2r}{\mu^2+1}, \frac{2\mu r}{\mu^2+1}\right) &\implies \lambda\frac{2r}{\lambda^2+1} = \mu\frac{2r}{\mu^2+1} \\ &\implies \lambda\frac{2r}{\lambda^2+1} = \mu\frac{2r}{\lambda^2+1} \\ &\implies \lambda = \mu\end{aligned}$$

(where the first implication follows by comparison of the second coordinates, and the second implication follows by comparison of the first coordinates).

It is also surjective, for if (x,y) lies on the circle minus the origin then we may try to compute the relevant λ: from 1.5 we may write $\lambda = y/x$, which is defined wherever $x \neq 0$; and the only point on the circle where $x=0$ is the excluded point $(0,0)$. ∎

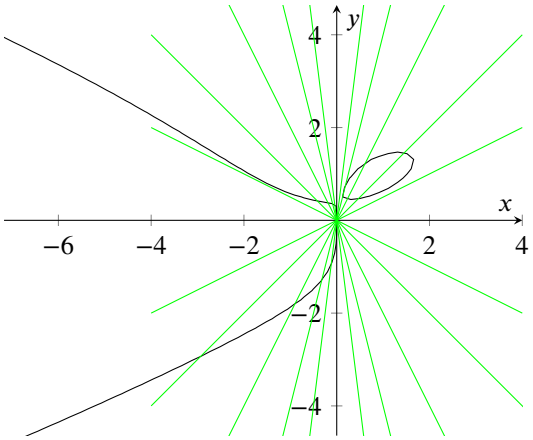


Figure 1.3: A circle and its relationship to a pencil of lines at an interior point.

1.2.3 Remark. In the above proof, the y–coordinate of the intersect between $y = \lambda x$ and $y = 1$ is λ. Thus the correspondence can be viewed as mapping the points on the circle to and from the line $y = 1$ by intersection, and it is easily seen why the origin is missed: the corresponding line is parallel to $y = 1$.

The final case is now where the pencil has centre *inside* the circle (figure 1.3).

1.2.4 Exercise. If $P = (x, y)$ lies within the circle $(x - r)^2 + y^2 = r^2$, that is if $\sqrt{(x - r)^2 + y^2} < r$, then each line in the pencil of lines at P intersects the circle precisely twice and so none of them are tangent to the circle (i.e. the circle has no real polar line with respect to the point as defined above).

The result in the exercise is not so interesting geometrically; what I am interested in is the behaviour of the intersection points over time as we vary the position of the pencil along the x-axis. Indeed, let $\beta \in \mathbb{R}$ and consider the pencil of lines

$$y = \lambda(x - \beta) \tag{1.7}$$

which we will intersect with the unit circle $x^2 + y^2 = 1$. For the sake of variety we shall compute the polar line for various values of β when $-1 \leq \beta \leq 1$ (i.e. the pencil is inside the circle), and hope we obtain something interesting. Because of the symmetry of the problem, the polar line will always be vertical and so we need only compute the x–ordinate of the tangency points.

1.2.5 Exercise. Show that when the point β lies inside the circle then there are two *complex* points of tangency in the sense we have been studying; the complex line joining them *does* intersect the real plane $\mathbb{R}^2$, and so there is still a natural polar line that manifests in the real plane.

Study the location of the *algebraic* polar line that is generated as β varies along the real line.

We can actually check the results of this exercise without doing any computations. The algebraic equations which produce as roots the points of tangency must have real coefficients; thus the two points of tangency must be related by being complex conjugates, say $P = (x_0 + x_1 i, y_0 + y_1 i)$ and $\bar{P}$; hence the line joining them has slope y_1/x_1 and equation

$$y = \frac{y_1}{x_1}(x - x_0 - x_1 i) + y_0 + y_1 i = \frac{y_1}{x_1}x - \frac{x_0 y_1}{x_1} + y_0 \tag{1.8}$$

which clearly manifests itself as a line in the real plane.

Thus study of the algebraic objects associated with the geometric picture gives us more than an easy way to compute things — it reveals symmetries that are invisible from a purely Euclidean point of view.

Chapter 2

The $\mathcal{I}$-$\mathcal{V}$ correspondence

In the previous chapter, we studied the zero-sets of specific polynomials in the plane. In this chapter, we will develop some basic machinery to allow us to discuss polynomial curves in a more formal way that will (in theory, if not in practice) allow us to deal with the classification of curves in a more hands-off manner. At the same time, we will generalise our theories to allow us to work in more general number systems than $\mathbb{C}$.

We will begin by reminding ourselves of some basic algebra.

2.1 Rings and fields

The reader will be familiar with the following large definition. A good reference is [Art91].

2.1.1 Definition. A **ring** is a set R together with two operations $\cdot$ and $+$ such that:

R1. $(R, +)$ is an abelian group; that is,

> **Associativity.** For all $r, s, t \in R$ we have $(r + s) + t = r + (s + t)$;
>
> **Commutativity.** For all $r, s \in R$ we have $r + s = s + r$;
>
> **Zero.** There exists an element $0 \in R$ such that for all $r \in R$, $0 + r = r$;
>
> **Inverses.** If $r \in R$ then there exists some $s \in R$ such that $r + s = 0$. (This element s is unique, and will be notated by $-s$.)

R2. For all $r, s, t \in R$, $(r \cdot s) \cdot t = r \cdot (s \cdot t)$. (We will often simply write rs for $r \cdot s$.)

R3. There exists some element $1 \in R$ such that for all $r \in R$, $1r = r = r1$.

R4. For all $r, s, t \in R$, $r(s + t) = rs + rt$ and $(r + s)t = rt + st$.

A ring is **commutative** if it satisfies the following additional axiom:

RC. For all $r, s \in R$, $rs = sr$.

For the remainder of these notes, the reader should assume that all rings are commutative unless otherwise stated.

In algebra, it is not just objects which pique our interest — functions between those objects are just as fundamental.

2.1.2 Definition. Let R and S be rings. A function $\phi : R \to S$ is called an **homomorphism** if:

RH1. For all $r, r' \in R$, $\phi(r + r') = \phi(r) + \phi(r')$ and $\phi(rr') = \phi(r)\phi(r')$, and

RH2. $\phi(1_R) = 1_S$ (where 1_R is the multiplicitive identity in R, and 1_S is the same in S).

The subset $\phi^{-1}(S)$ of R is called the **kernel** of ϕ and is denoted by $\ker \phi$.

If ϕ is a bijection then ϕ is called an **isomorphism**, and ϕ^{-1} is necessarily an homomorphism. If there is an isomorphism between R and S, we shall write $R \simeq S$ and say the two rings are **isomorphic**.

2.1.3 Remark. Condition RH2 is necessary in the above definition (i.e. it does not follow from RH1, even though for groups we *do* only need to give the equivalent to RH1); indeed, if R, R' are rings then the zero map $0 : R \to R'$ satisfies RH1 but not RH2 (unless R' is the zero ring).

The notion of a **subring** is not very useful; many natural subsets of a ring, like kernels of homomorphisms, are not usually subrings.

2.1.4 Definition. An **ideal** of a ring R is a subset $I \subseteq R$ such that for all $i, j \in I$, $i + j \in I$; and for all $i \in I$ and $r \in R$, $ri \in I$.

Ideals are important for the same reason that normal subgroups are. We have the following familiar result:

2.1.5 Lemma. *If $\phi : R \to S$ is a ring homomorphism, then $\ker \phi$ is an ideal of R.* ∎

Let I be a subgroup of the additive group of R, and let $\sim$ be the relation on R defined by $a \sim b \iff a - b \in I$. This is an equivalence relation; recall that the set $R/\sim$ is the set of equivalence classes of $\sim$, that is

$$R/\sim = \{[x] = \{y \sim x : y \in R\} : x \in R\}. \tag{2.1}$$

We shall also denote $[x]$ by $x + I$, and we shall usually denote $R/\sim$ by R/I.

2.1.6 Theorem. *There is a ring structure on R/I such that the canonical projection $\pi : R \ni x \mapsto x + I \in R/I$ is a ring homomorphism if and only if I is an ideal. We call the ring R/I the **quotient** of R by I.*

Proof. For π to be a ring homomorphism, we must have $\pi(rs) = rs + I \overset{\text{def}}{=} (r + I)(s + I) = \pi(r)\pi(s)$. The additive structure, $(r + I) + (s + I) = (r + s) + I$, is forced by the related theorem for groups.

The reader may show as an easy exercise that this operation is well-defined if I is an ideal; that is, if $r + I = r' + I$ and $s + I = s' + I$ then $(r + I)(s + I) = (r' + I)(s' + I)$ and so forth.

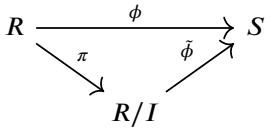
Conversely, suppose the projection is a ring homomorphism. Then $I = \ker \pi$, so I is an ideal by the previous lemma 2.1.5. ∎

We have therefore shown that $I \subseteq R$ is the kernel of a homomorphism if and only if I is an ideal in R: kernel $\implies$ ideal follows from the earlier lemma 2.1.5, and ideal $\implies$ kernel is shown by the existence of the canonical projection.

The following theorem (or rather, its generalisation) would have great claim to be called the 'fundamental theorem of algebra', were it not for the fact that the title is given to a small theorem in complex analysis (theorem 2.3.13).

2.1.7 Theorem (Homomorphism lemma for rings)**.**

1. *Let I be an ideal in a ring R. Then for every ring homomorphism $\phi : R \to S$ such that $I \subseteq \ker \phi$ there exists a unique ring homomorphism $\tilde{\phi} : R/I \to S$ such that the following diagram commutes:*



(2.2)

2. *Every ring homomorphism $\phi : R \to S$ may be decomposed as a product of a surjection $\twoheadrightarrow$, an isomorphism $\tilde{\phi}$, and an injection $\hookrightarrow$:*

$$\begin{array}{ccccccc} & & & \overset{\phi}{\frown} & & & \\ R & \twoheadrightarrow & R/\ker\phi & \overset{\tilde{\phi}}{\hookrightarrow\!\!\twoheadrightarrow} & \operatorname{im}\phi & \hookrightarrow & S \end{array} \tag{2.3}$$

Proof. Part 2 follows immediately from part 1; for part 1, take $\tilde{\phi}(r+I) = \phi(r)$ (one must check that it is well-defined and a ring homomorphism); for uniqueness of $\tilde{\phi}$, suppose $\underset{\sim}{\phi} : R/I \to S$ also makes the diagram commute, but is distinct from $\tilde{\phi}$; then there is some $x + I \in R/I$ such that $\underset{\sim}{\phi}(x+I) \neq \tilde{\phi}(x+I)$; hence $\phi(x) = (\tilde{\phi}\circ\pi)(x) = \tilde{\phi}(x+I) \neq \underset{\sim}{\phi}(x+I) = (\underset{\sim}{\phi}\circ\pi)(x) = \phi(x)$ which is absurd. ∎

2.1.8 Corollary (First homomorphism theorem)**.** *If $\phi : R \to S$ is a surjective ring homomorphism, then*

$$S \simeq \frac{R}{\ker\phi}. \tag{2.4}$$

∎

2.1.9 Corollary (Third homomorphism theorem)**.** *If I and J are ideals of R such that $I \subseteq J \subseteq R$, then J/I is an ideal of R/I and*

$$\frac{R/I}{J/I} \simeq \frac{R}{J}. \tag{2.5}$$

Proof. Define $\Psi : R/I \to J/I$ by $\Psi(r+I) = r+J$; then Ψ is a well-defined function which is a surjective ring homomorphism; it has kernel J/I; hence J/I is an ideal of R/I by lemma 2.1.5, and the claimed isomorphism holds by the previous corollary. ∎

2.1.10 Corollary (Correspondence theorem)**.** *If I is an ideal of R, there is a bijective inclusion-preserving correspondence between ideals of R containing I and ideals of R/I.*

Proof. By the previous corollary, if $I \subseteq J \subseteq R$ is a chain of ideals then J/I is an ideal of R/I. Conversely, if K is an ideal of R/I then write $J = \{j \in R : j + I \in K\}$. Then J is an ideal: $i, j \in J \implies i + I \in K$ and $j + I \in K$, so $(i+j)+I \in K$ and thus $i+j \in J$; furthermore, if $r \in R$ then $K \ni (r+I)(i+I) = (ri)+I$ so $ri \in J$.

Now if $I \subseteq J \subseteq R$ is a chain of ideals and we pick J/I inside R/I, consider $J_0 = \{j \in R : j + I \in J/I\}$. Then $j \in J \iff j + I \in J/I$, so $J_0 = J$. Conversely, suppose K is an ideal in R/I; then $\frac{\{j\in R : j+I\in K\}}{I} = K$ by definition. ∎

Some rings are badly behaved. The standard example is $\mathbb{Z}/6\mathbb{Z}$, in which the rule $ab = 0 \iff a = 0$ or $b = 0$ that we are used to in the integers fails: $(3 + 6\mathbb{Z})(2 + 6\mathbb{Z}) = (0 + 6\mathbb{Z})$.

2.1.11 Definition. A non-zero element $z \in R$ is called a **zero-divisor** if there exists some non-zero $s \in R$ such that $zs = 0$. A ring with no zero-divisors is called an **integral domain**; if every non-zero element has a multiplicitive inverse, a commutative ring is called a **field**.

If $I \subseteq R$ is an ideal, then I is called a **prime ideal** if R/I is an integral domain, and a **maximal ideal** if R/I is a field.

If $S \subseteq R$ is a subset, then the **ideal generated by** S is the set of all finite 'R–linear combinations'

$$s_1 r_1 + \cdots + s_n r_n \tag{2.6}$$

for $s_1, ..., s_n \in S$ and $r_1, ..., r_n \in R$.

If $S = \{s_1, ..., s_n\}$ is finite, we write $(s_1, ..., s_n) := (S)$.

2.1.12 Exercise.

1. A ring R is an integral domain iff for every three elements $r, s, t \in R$ such that $r \neq 0$, $rs = rt \iff s = t$.
2. $\mathbb{Z}/n\mathbb{Z}$ is an integral domain iff n is prime.
3. $I \cap J$ is always an ideal but $I \cup J$ is not.

2.1.13 Remark. If $u \in R$ has an inverse (such elements are **units**), then if I is an ideal of R we have $u \in I \iff I = R$. Thus a ring R is a field iff it has precisely two ideals, (0) and (1). The ideal $R = (1)$ is called the **unit ideal**, and the ideal (0) is called the **zero ideal**.

2.1.14 Proposition.

1. If $I \subseteq R$ is an ideal, then I is prime iff it satisfies the following condition:

$$\text{If } ab \in I, \text{ then } a \in I \text{ or } b \in I. \tag{PC}$$

2. If $I \subseteq R$ is an ideal, then I is maximal iff it satisfies the following condition:

$$\text{If } J \text{ is an ideal such that } I \subseteq J \subseteq R, \text{ then either } J = I \text{ or } J = R. \tag{PM}$$

Proof.

1. Suppose I is not prime; then R/I is not an integral domain, and thus there exist non-zero elements $r + I, s + I \in R/I$ such that $(r + I)(s + I) = 0 + I = I$. But $(r + I)(s + I) = (rs) + I$. Hence $r \notin I$, $s \notin I$, but $rs \in I$; so I does not satisfy the condition in the proposition.

   Conversely, suppose I does not satisfy (PC); so there exist $a, b \in R$ such that $ab \in I$ but neither $a \in I$ nor $b \in I$. Then $(a + I)(b + I) = (ab) + I = I$ is an exhibition of zero-divisors.

2. Suppose I is not maximal; then R/I is not a field, and thus there exists a non-zero element $b + I$ that has no inverse. Consider the ideal (I, b) of R. Clearly it contains I but is not equal to I. Further, if 1 was in (I, b) then there would exist elements $i \in I$ and $r \in R$ such that $1 = i + rb$; in other words, we would have $1 + I = rb + I = (r + I)(b + I)$. Hence $1 \notin (I, b)$ and thus proper inclusions hold in the chain $I \subset (I, b) \subset R$ — so I does not satisfy (MC).

   Conversely, suppose I does not satisfy (MC); then there is a proper chain of ideals $I \subset J \subset R$. Pick $b \in J \setminus I$, so $I \subset (I, b) \subset J \subset R$. In particular, $b + I$ cannot have an inverse in R/I; for if it did, then there would exist $r \in R$ such that $(rb) + I = 1 + I$, or equivalently $1 = rb + i$ for some $i \in I$; hence $1 \in (I, b)$ violating the inclusion condition. So R/I is forbidden to be a field. ∎

2.1.15 Exercise. Let R be a ring and let $I \subseteq R$ be an ideal. If u is a unit in R, then $u + I$ is a unit in R/I.

2.1.16 Proposition (Prime correspondence theorem)**.** *Prime ideals of R/I are in a bijective inclusion-preserving correspondence with prime ideals of R containing I.*

Proof. The inclusion-preserving bijection is given by the correspondence theorem (corollary 2.1.10).

Let $I \subseteq J \subseteq R$ be such that J is an ideal of R, then J is prime iff R/J is an integral domain; by the third homomorphism theorem, $R/J \simeq (R/I)/(J/I)$, so R/J is an integral domain iff $(R/I)/(J/I)$ is an integral domain; i.e. iff J/I is a prime ideal of R/I. ∎

2.1.17 Proposition. *Let $P_1, ..., P_n$ be prime ideals of R, and let $I \subseteq \cup_{i=1}^{n} P_i$ be an ideal of R. Then $I \subseteq P_i$ for some i.*

Proof. We shall show that if $\forall_i(I \not\subseteq P_i)$ then $I \not\subseteq \cup_{i=1}^n P_i$. If $n = 1$ then the result is trivial; we proceed by induction on n. Suppose we have proved the result for all sets of $n-1$ prime ideals of R. For each i, we have (applying the inductive hypothesis to the set of prime ideals $P_1, ..., P_n$ excluding P_i)

$$I \not\subseteq \bigcup_{j\in[n]\setminus\{i\}} P_j; \tag{2.7}$$

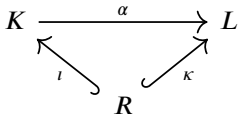
i.e. for each i there exists some $r_i \in I$ such that $r_i \notin P_j$ for all $j \neq i$. If there exists an i such that $r_i \notin P_i$ then $r_i \notin \cup_{i=1}^n P_i$ and we are done; hence we may assume that $r_i \in P_i$ for all i. Construct the element

$$s = \sum_{i=1}^{n} \sum_{j\in[n]\setminus\{i\}} r_j; \tag{2.8}$$

then $s \in I$ since it is a sum of products of elements of I. Further, for each i the residue of s modulo P_i is $r_1 \cdots r_{i-1} r_i \cdots r_n$ which cannot lie in P_i since P_i is prime and none of the multiplicands lie in P_i. Thus $s \in I$ and $\forall_i(s \notin P_i)$, so $I \not\subseteq \cup_{i=1}^n P_i$. ∎

Fields are very convenient, and so it can be useful to extend rings to form fields.

2.1.18 Proposition. *Let R be an integral domain. Then there exists a unique (up to isomorphism) field K and a unique injective homomorphism $\iota : R \hookrightarrow K$ such that for every field L and every injective homomorphism $\kappa : R \hookrightarrow L$, there exists a unique homomorphism of fields $\alpha : K \to L$ such that the following diagram commutes:*



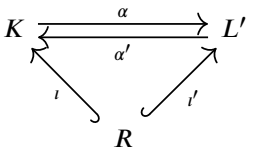
(2.9)

*The field K is called the **field of fractions** of R and denoted by* Frac R.

Since every homomorphism of fields is injective (its kernel must be an ideal of the domain), α gives us an 'inclusion' from K into L; so K is the 'smallest' field containing R.

Proof.

Uniqueness. Suppose (K, ι) and (K', ι') both satisfy the conditions above. Then there exist isomorphisms α and α' such that the following diagram commutes:



(2.10)

Hence $\iota = \alpha' \circ \iota'$ and $\iota' = \alpha \circ \iota$; so $\iota = \alpha' \circ \alpha \circ \iota$ and by injectivity we have $\alpha' = \alpha^{-1}$, so K and L are isomorphic.

Existence. Consider the set $R \times R^*$ (where R^* denotes the nonzero elements of R), and define the equivalence relation $\sim$ by $(a, b) \sim (c, d) \iff ad = bc$. Then define K to be the quotient $(R \times R^*)/\sim$ with operations defined by $[(a, b)] + [(c, d)] = [(ad + bc, bd)]$ and $[(a, b)][(c, d)] = [(ac, bd)]$ (one may check easily that these are well-defined). Further, we have a natural injection $R \hookrightarrow K$ given by $r \mapsto [(r, 1)]$. For obvious reasons, we will write $[(a, b)] = \frac{a}{b}$.

Now let L be a field such that $\kappa : R \to L$ is an injection. We wish to define an α as given in diagram 4.31. But this map is given uniquely by the need to be a homomorphism:

$$\alpha\left(\frac{a}{b}\right) = \alpha\left(\frac{a}{1}\right)\alpha\left(\frac{b}{1}\right)^{-1} = (\alpha\circ\iota)(a)\,(\alpha\circ\iota)(b)^{-1} = \kappa(a)\kappa(b)^{-1}, \tag{2.11}$$

and one can check that it is indeed a well-defined function $K \to L$ that preserves addition. ∎

2.1.19 Example. Frac $\mathbb{Z} = \mathbb{Q}$; Frac $K = K$ for all fields K.

2.1.20 Remark. The requirement in proposition 2.1.18 that the base ring R be an integral domain is important: since every embedding $R \hookrightarrow K$ will be injective, if R has zero divisors then S will also have zero divisors and thus can never be a field.

2.2 Polynomial rings

We know what polynomials are: they are things 'of the form' $37x^3 + 7xy - 4xy^3 + 2$. If we want to examine the geometry of polynomials, we will need a formal definition.

2.2.1 Definition. If R is a ring, and X is a symbol distinct from those in R, then the **polynomial ring** in X over R is the ring $R[X]$ such that:

1. Every $r \in R[X]$ is of the form $\sum_{i=0}^{d} r_i X^i$, such that each r_i lies in R, and such that $r_d = 0$ only if $d = 0$.
2. If $r, s \in R[X]$ then
$$r + s = \sum_{i=0}^{d} r_i X^i + \sum_{i=0}^{d} s_i X^i =: \sum_{i=0}^{d} (r_i + s_i) X^i. \tag{2.12}$$
3. If $r, s \in R[X]$ then
$$rs = \left(\sum_{i=0}^{d} r_i X^i\right)\left(\sum_{i=0}^{d} s_i X^i\right) =: \sum_{k=0}^{d} \sum_{i+j=k} \left(r_i s_j\right) X^k. \tag{2.13}$$

These definitions make $R[X]$ into a ring, which contains an isomorphic copy of R as a subfield via the injective homomorphism $R \ni r \to rX^0 \in R[X]$.

If $r = \sum_{i=0}^{d} r_i X^i$ with $r_d \neq 0$ then d is called the **degree** of r (we write ∂r for the degree); if $r = 0$, then $\partial r = -\infty$. The term r_d is called the **leading coefficient**, LC(r). If LC(r) is a unit, then r is called **monic**.

The polynomial ring over R in n variables, $X_1, ..., X_n$, is defined inductively by

$$R[X_1, ..., X_n] := \left(R[X_1, ..., X_{n-1}]\right)[X_n]. \tag{2.14}$$

If $r \in R[X_1, ..., X_n]$ then we may write

$$r = \sum_{\alpha_1 + \cdots + \alpha_n \leq d} r_{\alpha_1, ..., \alpha_n} X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n} \tag{2.15}$$

where there exists some n–tuple $\alpha_1, ..., \alpha_n$ such that their sum is exactly d; we then define the degree of r to be $\partial r := d$.

2.2.2 Remark. Note that :-

1. We have $R[X,Y] = R[X][Y] \simeq R[Y][X] = R[Y,X]$; this is a formalisation of the idea of 'collecting variables'.

2. If R is an integral domain then $\partial fg = \partial f + \partial g$.

3. Further, $\partial(f+g) \leq \max\{\partial f, \partial g\}$.

2.2.3 Lemma. *If R is an integral domain, then $R[X]$ is an integral domain.*

Proof. Suppose R is an integral domain; then if r and s are non-zero polynomials in $R[X]$ we have

$$rs = \left(\sum_{i=0}^{m} r_i X^i\right)\left(\sum_{i=0}^{n} s_i X^i\right) =: \sum_{k=0}^{m+n} \sum_{i+j=k} \left(r_i s_j\right) X^k \tag{2.16}$$

where $r_m \neq 0$ and $s_n \neq 0$. If the right-hand side were zero, this would imply $r_m s_n = 0$ which is not possible, so $rs \neq 0$. ∎

We want to study the zero-sets of polynomials, which forces us to treat them a little like functions. We have formalised what polynomials *are*, so we just need to talk about what it means to *substitute* values into them.

2.2.4 Theorem. *Let R and S be rings, and let $\phi : R \to S$ be a homomorphism. For every point $P = (r_1, ..., r_n) \in S^n$, there exists a unique homomorphism $\mathrm{eval}_{\phi,P} : R[X_1, ..., X_n] \to S$ which agrees with ϕ on R and sends $X_i \mapsto r_i$ for all i.*

Proof. The definition of $\mathrm{eval}_{\phi,P}$ must clearly be

$$\mathrm{eval}_P\left(\sum_{k=0}^{d} r_k X^k\right) = \sum_{k=0}^{d} \phi(r_k) r_i^k, \tag{2.17}$$

and in fact this definition is forced onto us by the requirements of the theorem and the definition of a homomorphism. ∎

If ϕ is the identity map, or is otherwise obvious, we shall simply write eval_P for the evaluation map. We shall often denote the image of $\mathrm{eval}_P\, f$ by $f(P)$ for obvious reasons.

2.2.5 Proposition. *Let $\phi : R[X] \to S$ be a ring homomorphism. Then $\phi = \mathrm{eval}_{\psi,x}$ for some ring homomorphism $\psi : R \to S$ and some $x \in S$. Further, if ϕ is injective then the induced homomorphism ψ is also injective.*

Proof. Let $\phi : R[X] \to S$ be a homomorphism, and set $x = \phi(X)$. Then for every polynomial f in $R[X]$ we are forced to send

$$\sum_{i=0}^{n} f_i X^i \mapsto \sum_{i=0}^{n} \phi(f_i) x^i \tag{2.18}$$

which is simply $\mathrm{eval}_{\psi,x}\, f$ where ψ is the restriction of ϕ to R.

Now suppose $\psi(r) = \psi(r')$ for some elements $r, r' \in R$. Then $\phi(r) = \mathrm{eval}\,\psi, xr = \mathrm{eval}\,\psi, xr' = \phi(r')$, so $r = r'$ by injectivity of ϕ. ∎

2.3 Single-variable polynomials

In one variable we have a rather nice result, as long as we make a couple of 'niceness' assumptions.

2.3.1 Theorem (Strong division algorithm)**.** *Let R be a ring; if $f, g \in R[X]$ $(g \neq 0)$ and g is monic then there exist $q, r \in K[X]$ such that $\partial r < \partial g$ and $f = gq + r$.*

Further, if R is an integral domain then q and r are unique.

Proof.

Existence. We shall use a standard trick from number theory. Let $S = \{f - gq : q \in R[X]\}$ (this is non-empty since $f \in S$), and let $r \in S$ be a polynomial of minimal degree. Clearly $f = gq + r$; suppose for the sake of contradiction that $m = \partial r \geq \partial g = n$. Write $r = \sum_{i=0}^{m} r_i X^i$ and $g = \sum_{i=0}^{n} g_i X^i$; consider $q' = q + \frac{r_m}{g_n} x^{m-n}$. Then:-

$$\begin{aligned} f - gq' &= f - gq - g\frac{r_m}{g_n}x^{m-n} \\ &= r - \left(\sum_{i=0}^{n} g_i X^i\right)\frac{r_m}{g_n}x^{m-n} \\ &= \sum_{i=0}^{m} r_i X^i - \sum_{i=0}^{n-1}\frac{g_i r_m}{g_n}X^i - r_m x^m \\ &= \sum_{i=0}^{m-1} r_i X^i - \sum_{i=0}^{n-1}\frac{g_i r_m}{g_n}X^i \end{aligned}$$

is a polynomial in S with degree at most $\min\{m-1, n-1\} = m - 1 < m$, contradicting minimality of degree of r.

Uniqueness in an integral domain. Now suppose $f = gq_1 + r_1 = gq_2 + r_2$ where both pairs (q_1, r_1) and (q_2, r_2) satisfy the conditions of the theorem. We therefore have $0 = g(q_1 - q_2) + (r_1 - r_2)$. Since $\partial(r_1 - r_2) < \partial g$, and we have coefficients in an integral domain (i.e. no zero divisors), the leading term of the right-hand side comes entirely from $g(q_1 - q_2)$. Thus the product of the leading coefficients of g and $(q_1 - q_2)$ is zero; since the LC of g is non-zero by assumption, the leading coefficient of $q_1 - q_2$ is zero; by definition, the leading coefficient is non-zero unless $q_1 - q_2 = 0$. Immediately we have $r_1 = r_2$ as well. ∎

2.3.2 Corollary (Weak division algorithm)**.** *Let K be a field; if $f, g \in K[X]$ $(g \neq 0)$ then there exist unique $q, r \in K[X]$ such that $\partial r < \partial g$ and $f = gq + r$.* ∎

The polynomial r is known as the **remainder** of f upon division by g. We say g **divides** f and write $g \mid f$ if the remainder is the zero polynomial.

More generally, if R is any ring we say a **divides** b and write $a \mid b$ if there exists some $q \in R$ such that $aq = b$.

2.3.3 Corollary (Remainder theorem)**.** *Let R be a ring, let $f \in R[X]$, and let $\alpha \in R$. Then $f(\alpha)$ is simply the remainder of f upon division by $(X - \alpha)$.* ∎

2.3.4 Corollary. *Let R be a ring, let $f \in R[X]$, and let $\alpha \in R$. Then $f(\alpha) = 0$ iff $(X - \alpha) \mid f$.* ∎

Up until this point we have significant parallels between the theory of divisibility in $\mathbb{Z}$ and in $K[X]$. There is one final (and significant) similarity.

2.3.5 Corollary. *Every ideal in $K[X]$ is generated by a single element: that is, if $I \subseteq K[X]$ then $I = (f)$ for some $f \in K[X]$.*

Proof. For hopefully obvious reasons we may assume $I \neq (0)$.

Pick $f \in K[X]$ such that ∂f is minimal and non-negative (i.e. $f \neq 0$). It is trivial that $(f) \subseteq I$, so we need to check the other direction. Suppose $g \in I$. Then we may write $f = gq + r$ for some $q, r \in K[X]$ and $\partial r < \partial g$. But $r = f - gr \in I$; this contradicts minimality of ∂f unless $r = 0$. ∎

2.3.6 Definition. An ideal generated by a single element is called a **principal ideal**, and a ring in which every ring is principal is a **principal ideal domain** or **PID**.

2.3.7 Example. Corollary 2.3.5 told us that $K[X]$ is a PID for any field K. In addition, $\mathbb{Z}$ is a PID (see [Art91, proposition 3.10.18]).

Note also, every field is trivially a PID: by remark 2.1.13, there are only two cases to check and the two ideals are generated by 0 and 1.

A pleasant discussion of the properties of PIDs may be found in the first chapter of [IR82].

2.3.8 Exercise. In a PID, prime ideals are maximal.

Unfortunately, $K[X_1, ..., X_n]$ for $n > 1$ is not necessarily a PID, even though K is a field.

2.3.9 Example. Consider $(X, Y) \subseteq \mathbb{C}[X, Y]$; then there is no $f \in \mathbb{C}[X, Y]$ such that $(X, Y) = (f)$. Indeed, suppose there was; then there would exist $\alpha, \beta \in \mathbb{C}[X, Y]$ so that $X = f\alpha$ and $Y = f\beta$. Then $1 = \partial X = \partial f + \partial\alpha$ and (likewise) $1 = \partial f + \partial\beta$. We know that $\partial f \geq 1$, since otherwise $(f) = (X, Y)$ would contain a unit and then there would exist $a, b \in \mathbb{C}[X, Y]$ such that $aX + bY = 1$; but $(aX + bY)(0, 0) = 0$ and $1(0, 0) = 1 \neq 0$. Further, if $\partial f = 1$ then α, β are non-zero constants, so $\alpha^{-1}X = f = \beta^{-1}Y$ which is clearly nonsense (if the reader really wishes for a reason, $(\alpha^{-1}X)(0, 1) = 0 \neq \beta^{-1} = (\beta^{-1}Y)(0, 1)$)). So $\partial f > 1$, which means that $0 > 1 - \partial f = \partial\alpha$; hence $\alpha = 0$; so $X \equiv 0$ (where $\equiv$ is used to emphasise that this is the statement that the *polynomials* are equal, not the evaluations), which is nonsense since $X(1, 1) = 1 \neq 0 = 0(1, 1)$.

This example is instructive as it shows the power of evaluating polynomials at zero in order to study their algebraic behaviour. We will return to this theme again and again.

2.3.10 Definition. A polynomial $f \in R[X_1, ..., X_n]$ is said to have a **zero** at $P \in R^n$ if $f(P) = 0$.

If R is an integral domain and $f \in R[X]$ has a zero at $\alpha \in R$, then by corollary 2.3.4), $(x - \alpha) \mid f$. The **multiplicity** of α is the largest $n \in \mathbb{N}$ such that $(x - \alpha)^n \mid f$; if f does not have a zero at α we shall say that the multiplicity of α is zero, and we shall denote the multiplicity by $\mu_\alpha f$. We will generalise it to multi-variable polynomials later on.

2.3.11 Definition. A field K is called **algebraically closed** if every polynomial $f \in K[X]$ has a zero.

2.3.12 Proposition. *If K is algebraically closed, then every non-constant $f \in K[X_1, ..., X_n]$ has a zero.*

Proof. Consider $\text{eval}_{(X,0,...,0)} : K[X_1, ..., X_n] \to K[X]$; then $\text{eval}_{(X,0,...,0)} f$ has a zero in K, say α. In other words, $0 = (\text{eval}_\alpha \circ \text{eval}_{(X,0,...,0)})(f) = (\text{eval}_{(\alpha,0,...,0)})(f)$. ∎

2.3.13 Theorem (So-called fundamental theorem of algebra)**.** *The field $\mathbb{C}$ is algebraically closed.*

Proof. [Art91, §13.9] ∎

2.3.14 Corollary. *If K is algebraically closed, and $f \in K[X]$ is non-constant, then*

$$\sum_{\alpha \in K} \mu_\alpha f = \partial f. \tag{2.19}$$

Proof. Repeated application of division, corollary 2.3.4, and remark 2.2.2 (to ensure the degree reduces by 1 upon each division). ∎

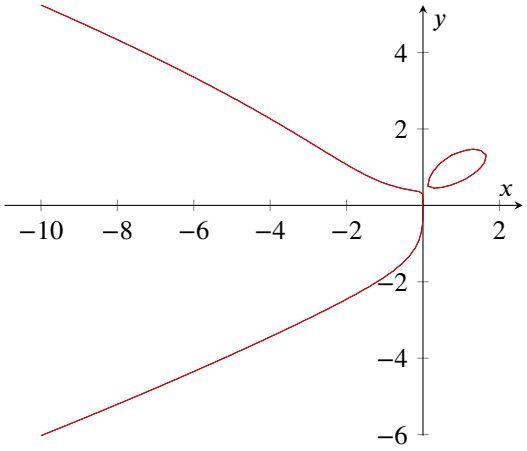


Figure 2.1: Two curves, which are not graphs of functions.

2.4 Translating between geometry and algebra

We now come to the two main definitions:

2.4.1 Definition. Let K be a field.

1. Let $S \subset k[X_1, ..., X_n]$ be a set of polynomials. Then the **zero-set** of S is the set
$$\mathcal{V}(S) := \{P \in k^n : \forall_{f \in S}, f(P) = 0\}. \tag{2.20}$$
2. Let $X \subset k^n$ be a set of points. Then the **vanishing ideal** of X is the set
$$\mathcal{I}(X) := \{f \in k[X_1, ..., X_n] : \forall_{P \in X}, f(P) = 0\}. \tag{2.21}$$

A set which is the zero-set of some set is called an **algebraic set**; a set which is the zero-set of a single polynomial is called a **hypersurface**, and a hypersurface in K^2 is called a **plane algebraic curve** if its defining polynomial is non-constant.

It is traditional to write $\mathbb{A}^n_K$ for K^n when we are doing algebraic geometry over it; the main advantage of this notation is that it encourages us to forget about the special algebraic points of K^n like the origin. If the field K is apparent, we will write $\mathbb{A}^n$ for brevity, and refer to it as **affine n-space**.

2.4.2 Example. Figure 2.1 shows $\mathcal{V}(X^3 + Y^4 - 5XY + 2X)$ and $\mathcal{V}(X^2 + Y^2 - 25)$.

2.4.3 Lemma. *If $X \subseteq \mathbb{A}^n$, then $\mathcal{I}(X)$ is an ideal in $K[X_1, ..., X_n]$.*

Proof. Suppose $f, g \in \mathcal{I}(X)$ and $k \in K[X_1, ..., X_n]$. Then for all $P \in X$, $(f + g)(P) = f(P) + g(P) = 0 + 0 = 0$; and $(kf)(P) = k(P)f(P) = k(P)0 = 0$. ∎

Our goal is to set up a nice correspondence between sets of points in $\mathbb{A}^n$ and sets of polynomials in $K[X_1, ..., X_n]$; our point-sets will always be algebraic sets, and our sets of polynomials will always be ideals. Unfortunately this correspondence does not work immediately: the function $\mathcal{V}$ is not injective, even in $\mathbb{A}^1$, since $\mathcal{V}((X)) = \{0\} = \mathcal{V}((X^2))$ but $(X) \neq (X^2)$.

In order to fix this problem, we will study the properties of combinations of algebraic sets and polynomial ideals, and we shall map such combinations through $\mathcal{I}$ and $\mathcal{V}$. We will first check that $\mathcal{V}$ and $\mathcal{I}$ form a

bijection if and only if we work only with algebraic sets and vanishing ideals; the problem is then reduced to working out exactly which ideals *are* vanishing ideals, which will follow from Hilbert's Nullstellensatz (theorem 2.11.1).

2.4.4 Proposition. *Let $S \subseteq K[X_1, ..., X_n]$ and $X \subseteq \mathbb{A}^n$. Then:-*

1. *$\mathcal{I}(\mathcal{V}(S)) \supseteq S$ and $\mathcal{V}(\mathcal{I}(X)) \supseteq X$.*
2. *$\mathcal{I}(\mathcal{V}(S)) = S$ iff S is a vanishing ideal, and $\mathcal{V}(\mathcal{I}(X)) = X$ iff X is algebraic.*

Proof.

1. Suppose $f \in S$. Then $f(P) = 0$ for all $P \in \mathcal{V}(S)$; so $f \in \mathcal{I}(\mathcal{V}(S))$. Similarly, suppose $P \in X$. Then $f(P) = 0$ for all $f \in \mathcal{I}(X)$; so $P \in \mathcal{V}(\mathcal{I}(X))$.
2. If $S = \mathcal{I}(\mathcal{V}(S))$ then clearly S is a vanishing ideal; similarly, if $X = \mathcal{V}(\mathcal{I}(X))$ then X is obviously algebraic.

   Suppose S is a vanishing ideal; so $S = \mathcal{I}(X)$ for some set $X \subseteq \mathbb{A}^n$. We therefore need to prove that $\mathcal{I}(\mathcal{V}(\mathcal{I}(X))) = \mathcal{I}(X)$. Part 1 tells us that $\mathcal{I}(\mathcal{V}(\mathcal{I}(X))) \supseteq \mathcal{I}(X)$; on the other hand, if $f \in \mathcal{I}(\mathcal{V}(\mathcal{I}(X)))$ then $f(P) = 0$ for all $P \in \mathcal{V}(\mathcal{I}(X))$. But by part 1, $\mathcal{V}(\mathcal{I}(X)) \supseteq X$; so $f(P) = 0$ for all $P \in X$, and thus $f \in \mathcal{I}(X)$.

   Suppose X is an algebraic set; so $X = \mathcal{V}(S)$ for some set $S \subseteq K[X_1, ..., X_n]$. We therefore need to prove that $\mathcal{V}(\mathcal{I}(\mathcal{V}(S))) = \mathcal{V}(S)$. Part 1 tells us that $\mathcal{V}(\mathcal{I}(\mathcal{V}(S))) \supseteq \mathcal{V}(S)$; on the other hand, if $P \in \mathcal{V}(\mathcal{I}(\mathcal{V}(S)))$ then $f(P) = 0$ for all $f \in \mathcal{I}(\mathcal{V}(S))$. But by part 1, $\mathcal{I}(\mathcal{V}(S)) \supseteq S$; so $f(P) = 0$ for all $f \in S$, and thus $P \in \mathcal{V}(S)$. ∎

Consider the union of two algebraic sets, $\mathcal{V}(f) \cup \mathcal{V}(g)$. It is clear that $P \in \mathcal{V}(f) \cup \mathcal{V}(g) \iff f(P) = 0$ or $g(P) = 0 \iff (fg)(P) = 0$; so $\mathcal{V}(f) \cup \mathcal{V}(g) = \mathcal{V}(fg)$. In general, we want to be able to do this for ideals of polynomials: we would like to say that $P \in \mathcal{V}(I) \cup \mathcal{V}(J) \iff P \in \mathcal{V}(IJ)$, where $IJ = \{fg : f \in I, g \in J\}$. Unfortunately, this is not an ideal in general (consider (X) and (Y): xy and x^2y are products of elements of (X) and (Y), but their sum is not). We will take the smallest ideal containing all these products. We will also need a sum of ideals, which is better behaved.

2.4.5 Definition. Let I, J be ideals of R. The **product** of I and J is the ideal $IJ = (\{ij : i \in I, j \in J\})$; the **sum** of I and J is the ideal $I + J = \{r + s : (r, s) \in I \times J\}$.

2.4.6 Exercise.

1. In $\mathbb{Z}$, the product of two ideals *is* just the set of products of elements from the two ideals.
2. $IJ \subseteq I \cap J$, $I + J \supseteq I \cup J$, and $I + J \subseteq (I \cup J)$.

2.4.7 Proposition (Calculus of algebraic sets and vanishing ideals)**.** *Let $X, Y \subseteq \mathbb{A}^n$ (not necessarily algebraic), $\{X_\beta\}_{\beta \in B}$ a family of subsets of $\mathbb{A}^n$ (not necessarily algebraic), $I, J, S, T \subseteq K[X_1, ..., X_n]$ (I, J ideals), and $\{I_\alpha\}_{\alpha \in A}$ a family of ideals in $K[X_1, ..., X_n]$. Then:-*

1. *$\mathcal{V}(\emptyset) = \mathbb{A}^n$ and $\mathcal{I}(\emptyset) = K[X_1, ..., X_n]$;*
2. *$\mathcal{V}(K[X_1, ..., X_n]) = \emptyset$; $\mathcal{I}(\mathbb{A}^n)$ is $((X_1^q - X_1), ..., (X_n^q - X_n))$ if K is finite of order q, and is (0) otherwise;*
3. *$\mathcal{V}(S) = \mathcal{V}((S))$;*
4. *$\mathcal{V}(\cup_{\alpha \in A} I_\alpha) = \cap_{\alpha \in A} \mathcal{V}(I_\alpha)$, $\mathcal{I}(\cup_{\beta \in B} X_\beta) = \cap_{\beta \in B} \mathcal{I}(X_\beta)$, and $\mathcal{I}(X) \cap \mathcal{I}(Y) = \mathcal{I}(X)\mathcal{I}(Y)$;*

5. $\mathcal{V}(I \cap J) = \mathcal{V}(IJ) = \mathcal{V}(I) \cup \mathcal{V}(J)$ *("smaller ideals have more shared zeroes");*

6. $\mathcal{V}(I \cup J) = \mathcal{V}(I + J) = \mathcal{V}(I) \cap \mathcal{V}(J)$ *("if there are more functions, having shared zeroes becomes harder");*

7. $S \subseteq T \implies \mathcal{V}(S) \supseteq \mathcal{V}(T)$ *and* $X \subseteq Y \implies \mathcal{I}(X) \supseteq \mathcal{I}(Y)$.

Proof. These are not proved in the order stated above.

1. $P \in \mathcal{V}(\emptyset) \iff \forall_{f \in \emptyset}, f(P) = 0$. But I shall pay \$100 to the first person who can exhibit a function in the empty set which does not vanish at P. The second equality is an exercise.

4. For the algebraic set property,

$$\begin{aligned} P \in \mathcal{V}(\cup_\alpha I_\alpha) &\iff \forall_{\alpha \in A}(f \in I_\alpha \implies f(P) = 0) \\ &\iff \forall_{\alpha \in A}, P \in \mathcal{V}(I_\alpha) \\ &\iff P \in \cap_{\alpha \in A} \mathcal{V}(I_\alpha). \end{aligned} \tag{2.22}$$

   For the vanishing ideal equality chain,

$$\begin{aligned} f \in \mathcal{I}(\cup_\beta X_\beta) &\iff \forall_{\beta \in B}(P \in X_\beta \implies f(P) = 0) \\ &\iff \forall_{\beta \in B} f \in \mathcal{I}(X_\beta) \\ &\iff f \cap_{\beta \in B} \mathcal{I}(X_\beta) \end{aligned} \tag{2.23}$$

   so $\mathcal{I}(X \cup Y) = \mathcal{I}(X) \cap \mathcal{I}(Y)$.

   For the final equality, note that by exercise 2.4.6 we have that $\mathcal{I}(X) \cap \mathcal{I}(Y) \subseteq \mathcal{I}(X)\mathcal{I}(Y)$. Conversely, $f \in \mathcal{I}(X)\mathcal{I}(Y)$ implies $f = \sum f_i g_i h_i$ for some $g_i \in \mathcal{I}(X)$, $h_i \in \mathcal{I}(Y)$, and $f_i \in K[X_1, ..., X_n]$. Thus for all $P \in X$ we have $f(P) = \sum f_i(P) g_i(P) h_i(P) = \sum f_i(P) \cdot 0 \cdot h_i(P) = 0$; so $f \in \mathcal{I}(X)$. Similarly, $f \in \mathcal{I}(Y)$ and so $\mathcal{I}(X)\mathcal{I}(Y) \subseteq \mathcal{I}(X) \cap \mathcal{I}(Y)$.

7. Suppose $P \in \mathcal{V}(T)$; then $\forall_{f \in T}, f(P) = 0$. But $f \in S \implies f \in T$, so $\forall_{f \in S}, f(P) = 0$; hence $p \in \mathcal{V}(S)$. The second implication is an exercise.

3. Note that $S \subseteq (S)$, so (7) implies that $\mathcal{V}((S)) \subseteq \mathcal{V}(S)$. Conversely, suppose $P \in \mathcal{V}(S)$; then for all $f \in S$, $f(P) = 0$. If $g \in (S)$, then $g = \sum r_i f_i$ for some $r_i \in K[X_1, ..., X_n]$ and $f_i \in S$; thus $\mathrm{eval}_P\, g = \sum \mathrm{eval}_P(r_i)\, \mathrm{eval}_P(f_i) = 0$; so $P \in \mathcal{V}((S))$ and $\mathcal{V}((S)) \supseteq \mathcal{V}(S)$.

2. By (3), $\mathcal{V}(K[X_1, ..., X_n]) = \mathcal{V}((1)) = \mathcal{V}(1)$. But $\mathrm{eval}_P\, 1 = 1 \neq 0$ for all P (since K is a field). The statement about vanishing ideals is theorem 2.8.1; it can be easily checked that the proof of that theorem, despite occuring later in these notes, does not depend on this proposition.

5. Note that $I \cap J \subseteq IJ$ (exercise 2.4.6). Thus by (7), $\mathcal{V}(I \cap J) \supseteq \mathcal{V}(IJ)$.

   Now let $P \in \mathcal{V}(I) \cup \mathcal{V}(J)$; so either $P \in \mathcal{V}(I)$ or $P \in \mathcal{V}(J)$. Thus for all pairs $(f, g) \in I \times J$ we have $f(P)g(P) = 0$ since at least one of the factors vanishes. Thus $P \in \mathcal{V}(\{fg : f \in I, g \in J\}) = \mathcal{V}(IJ)$ (where the second equality comes from (3)), and so $\mathcal{V}(IJ) \supseteq \mathcal{V}(I) \cup \mathcal{V}(J)$.

   Finally, note that $P \notin \mathcal{V}(I) \cup \mathcal{V}(J) \implies \exists_{f \in \mathcal{V}(I)}$ and $\exists_{g \in \mathcal{V}(J)}$ such that $f(P) \neq 0$ and $g(P) \neq 0$; but $fg \in IJ \subseteq I \cap J$, and since K is an integral domain $f(P)g(P) \neq 0$; so $f \notin \mathcal{V}(I \cap J)$. Thus $\mathcal{V}(I \cap J) \supseteq \mathcal{V}(I) \cup \mathcal{V}(J)$.

6. By exercise 2.4.6, $I+J \subseteq (I \cup J)$. Hence $\mathcal{V}(I+J) \supseteq \mathcal{V}((I \cup J)) = \mathcal{V}(I \cup J)$. By the same exercise, $I+J \supseteq I \cup J$ and so $\mathcal{V}(I \cup J) \subseteq \mathcal{V}(I+J)$. Hence $\mathcal{V}(I+J) = \mathcal{V}(I \cup J)$.

   Finally, $P \in \mathcal{V}(I) \cap \mathcal{V}(J) \implies \forall_{f \in I}, \forall_{g \in J}, f(P) = g(P) = 0$. Thus $(f+g)(P) = 0$ and $P \in \mathcal{V}(F+G)$. Conversely, $I \subseteq I+J$ so $\mathcal{V}(I) \supseteq \mathcal{V}(I+J)$; similarly, $\mathcal{V}(J) \supseteq \mathcal{V}(I+J)$; so $\mathcal{V}(I) \cap \mathcal{V}(J) \supseteq \mathcal{V}(I+J)$. Thus $\mathcal{V}(I) \cap \mathcal{V}(J) \subseteq \mathcal{V}(I+J) \subseteq \mathcal{V}(I) \cap \mathcal{V}(J)$. ∎

We shall spend the next few sections studying ways to "cut up" and "glue together" algebraic sets.

2.5 Introduction to topology

It will be useful to use a few notions from topology. A readable introductory text is [Mun00].

2.5.1 Definition. If X is a set, a **topology** on X is a set $\mathcal{T}$ of subsets of X such that:

T1. $\emptyset$ and X are open;

T2. $U \cap V$ open for all U, V open; and

T3. $\cup_{\alpha \in I} U_\alpha$ for every family $\{U_\alpha\}_{\alpha \in I}$ of open sets.

The elements of $\mathcal{T}$ are called **open sets** of X; the complements of open sets in X are called **closed sets**. A set which is both closed and open is called **clopen**, and a set which is neither closed nor open is called **cleither**.

2.5.2 Remark. One might also define a topology as a set of closed sets, and one can easily show that the closed sets satisfy the following axioms:

T1′. $\emptyset \in X$ and $X \in \mathcal{T}$;

T2′. $F \cup G \in \mathcal{T}$ for all $F, G \in \mathcal{T}$; and

T3′. $\cap_{\alpha \in I} F_\alpha$ for every family $\{F_\alpha\}_{\alpha \in I} \subseteq \mathcal{T}$.

It is easily seen that the two definitions are equivalent.

2.5.3 Example.

1. If X is any set, the **discrete topology** on X is the set of all subsets of X. The **indiscrete topology** is the set $\{\emptyset, X\}$.

2. The **usual topology** on $\mathbb{R}^n$ has, as open sets, all unions of balls $B_r(x) = \{y \in \mathbb{R}^n : |x-y| < r\}$.

3. More generally, if (X, ρ) is a metric space, the **usual topology** on X has, as open sets, all unions of balls $B_r(x) = \{y \in X : \rho(x,y) < r\}$.

4. If $Y \subseteq X$ is a subset, there is a natural **subspace topology** on Y: namely, a subset $U \subseteq Y$ is open iff there is an open set U' in X such that $U' = U \cap Y$.

5. If X is any set, the set
$$\mathcal{T}_\infty = \{Y \subseteq X : X \setminus Y \text{is finite}\} \cup \{\emptyset, X\} \tag{2.24}$$
   is the **cofinite topology** on X.

Recall that in $\mathbb{R}^n$ we have a definition of 'continuity'. This is generalised in topological terms as follows.

2.5.4 Definition. Let X and Y be topological spaces; then a function $f : X \to Y$ is **continuous** if, whenever U is open in Y, $f^{-1}(U)$ is open in X. The function f is a **homeomorphism** if f is bijective and f^{-1} is continuous.

2.5.5 Remark. In $\mathbb{R}^n$, or more generally for every pair of metric spaces (X, ρ), (Y, σ), this is indeed equivalent (it is easy to check) to the famous classical definition of continuity; i.e. $f : X \to Y$ is continuous at $x \in X$ iff

$$\forall_{\varepsilon>0}\exists_{\delta>0}\forall_{x'\in X}\left(\rho(x-x') < \delta \implies \sigma(f(x)-f(x')) < \varepsilon\right) \tag{2.25}$$

and f is continuous if it is continuous at all $x \in X$.

2.5.6 Remark. If X and Y are topological spaces, $f : X \to Y$ is continuous iff whenever F is closed in Y, $f^{-1}(F)$ is closed in X.

We care about polynomials, so we want polynomials to be 'continuous' in some sense (though they're not functions). For our purposes, we only care about *zeroes* of polynomials; and so in a useful topology that captures the geometry of polynomials we should have that inverse images of the singleton $\{0\}$ (which should, intuitively, be closed) are closed. It turns out that this is indeed a topology.

2.5.7 Lemma. *Let $\mathscr{A}$ be the set of all algebraic subsets of $\mathbb{A}^n$; let $\mathcal{T} = \{\mathbb{A}^n \setminus X : X \in \mathscr{A}\}$. Then $\mathcal{T}$ is a topology on $\mathbb{A}^n$ (known as the **Zariski topology**). If V is an algebraic subset of $\mathbb{A}^n$, then the subspace topology inherited from the Zariski topology on $\mathbb{A}^n$ will be called the Zariski topology on V.*

When we discuss $\mathbb{A}^n$ we will almost always endow it with the Zariski topology, while when we discuss $\mathbb{C}^n$ or $\mathbb{R}^n$ we will almost always endow them with the usual topology.

Proof. We verify the closed set conditions given in remark 2.5.2.

1. By parts 2 and 1 respectively of proposition 2.4.7 (the calculus of the $\mathcal{I}$–$\mathcal{V}$ correspondence), $\emptyset$ and $\mathbb{A}^n$ are algebraic.
2. By part 5 of the calculus, $\mathcal{V}(I) \cup \mathcal{V}(J) = \mathcal{V}(J \cap I)$ (where I and J are ideals of polynomials); thus, by induction, finite unions of algebraic sets are algebraic.
3. By part 4 of the calculus, $\cap_{\alpha\in A}\mathcal{V}(I_\alpha) = \mathcal{V}(\cup_\alpha I_\alpha)$ (where $\{I_\alpha\}_{\alpha\in A}$ is a family of sets of polynomials); thus arbitrary intersections of algebraic sets are algebraic. ∎

2.5.8 Exercise. The Zariski topology on $\mathbb{A}^1_K$ is the cofinite topology if K is infinite. If K is finite, the Zariski topology on $\mathbb{A}^n_K$ is the discrete topology.

2.5.9 Exercise. If $\mathbb{A}^2$ is identified with $\mathbb{A}^1 \times \mathbb{A}^1$ in the natural way, show that the Zariski topology on $\mathbb{A}^2$ is not the product topology of the Zariski topologies on the two copies of $\mathbb{A}^1$.

2.5.10 Definition. A subset Y of a topological space X is called **dense in** X if for every nonempty $U \subseteq X$ open, $Y \cap U \neq \emptyset$.

2.5.11 Example. The rational numbers $\mathbb{Q}$ are dense in $\mathbb{R}$ with the usual topology. (Proof: every ball contains a rational number.)

2.5.12 Proposition. *If Y is dense in X, then the smallest closed set containing Y (the **closure** of Y) is X itself.*

Proof. Suppose $Y \subset X$ and there exists some closed set F such that $Y \subseteq F \subset X$. Then $X \setminus F$ is a non-empty open set such that $Y \cap F = \emptyset$. ∎

This concept allows us to formalise the idea that 'an algebraic set covers almost none of $\mathbb{A}^n$'.

2.5.13 Proposition. *Let K be an infinite field. If $\mathbb{A}^n_K$ is endowed with the Zariski topology, then every non-empty open set is dense in $\mathbb{A}^n_K$.*

Proof. It suffices to show that every pair of non-empty open sets intersects.

Suppose $U, V \subseteq \mathbb{A}^n$ are open such that $U \cap V = \emptyset$. By openness, there exist vanishing ideals I, J such that $U = \mathbb{A}^n \setminus \mathcal{V}(I)$ and $V = \mathbb{A}^n \setminus \mathcal{V}(J)$. Thus

$$\emptyset = \mathbb{A}^n \setminus (\mathcal{V}(I) \cup \mathcal{V}(J)) = \mathbb{A}^n \setminus \mathcal{V}(I \cap J). \tag{2.26}$$

Hence $\mathcal{V}(I \cap J) = \mathbb{A}^n$; since the intersection of vanishing ideals is a vanishing ideal (part 4 of the calculus), we may write $I \cap J = \mathcal{I}(\mathcal{V}(I \cap J)) = \mathcal{I}(\mathbb{A}^n) = (0)$ ∎

In $\mathbb{R}^n$ (with the usual topology), we have a variety of separation results: individual points are in some sense far apart from each other, since we can separate them by open sets. The mathematical content of this may be generalised to more abstract spaces as follows.

2.5.14 Definition. A topological space X is said to satisfy the T_1 **axiom** and is called a **Fréchet space** if

$$\text{If } x, y \in X \text{ distinct there exist open sets } U_x, U_y \text{ such that } x \in U_x,\ y \in U_y, \text{ and } x \notin U_y,\ y \notin U_x. \tag{T_1}$$

A topological space X is said to satisfy the T_2 **axiom** and is called a **Hausdorff space** if

$$\text{If } x, y \in X \text{ distinct there exist disjoint open sets } U_x, U_y \text{ such that } x \in U_x,\ y \in U_y. \tag{T_2}$$

(There are six separation axioms; see [SS95, p. 11].)

2.5.15 Proposition. *A topological space X is T_1 iff singletons are closed sets.*

Proof. Suppose singletons are closed sets; then $x, y \in X$ implies $U_y = X \setminus \{x\}$ and $U_x = Y \setminus \{y\}$ are both open sets satisfying the axiom condition. Conversely, suppose $x \in X$. Then for all $y \in X$ distinct from x pick open U_y such that $y \in U_y$ but $x \notin U_y$. Then $\cup_{y \in X \setminus \{x\}} U_y$ is an open set; and by construction it is equal to $X \setminus \{x\}$. ∎

Spaces which are Hausdorff are well-behaved: for example, limit points are unique. It would be nice if $\mathbb{A}^n$ was Hausdorff.

2.5.16 Proposition. *$\mathbb{A}^n_K$ is Hausdorff iff K is finite.*

Proof. Let $x, y \in \mathbb{A}^n$ be distinct and arbitrary. (Note this is possible since fields always have at least two points.) Then if K is infinite, every open set containing x intersects every open set containing y by proposition 2.5.13. If K is finite, it has the discrete topology by exercise 2.5.8; in the discrete topology, points are open and so $\{x\}$ and $\{y\}$ are non-intersecting open neighbourhoods of x and y. ∎

A small consolation is the following, which verifies that in particular $\{0\}$ is indeed closed like we wanted:

2.5.17 Proposition. *Points are closed in $\mathbb{A}^n$, so it is T_1.*

Proof. Let $P \in \mathbb{A}^n$; then $P = (x_1, ..., x_n)$ and we may form the ideal $I = ((X_1 - x_1), ..., (X_n - x_n))$. Suppose $f \in I$, so $f = \sum_{i=0}^n f_i(X_i - x_i)$. First note that $f(P) = 0$, since

$$\operatorname{eval}_P f = \sum_{i=0}^{n} \left((\operatorname{eval}_P f_i)(\operatorname{eval}_P (X_i - x_i)) \right) = 0. \tag{2.27}$$

Thus $P \in \mathcal{V}(I)$. Conversely, suppose $(y_1, ..., y_n) = Q \in \mathcal{V}(I)$. For every i $(1 \leq i \leq n)$, note that the following diagram commutes:

$$\begin{array}{ccc} K[X_1, ..., X_n] & \xrightarrow{\text{eval}_{(y_1,...,y_{i-1},X,y_{i+1},...,y_n)}} & K[X] \\ & \searrow^{\text{eval}_Q} & \downarrow{\scriptstyle \text{eval}_{y_i}} \\ & & K \end{array} \tag{2.28}$$

Hence we have

$$0 = \text{eval}_Q(X_i - x_i) = \text{eval}_{y_i}\left(\text{eval}_{y_1,...,X,...,y_n}(X_i - x_i)\right) = \text{eval}_{y_i}(X - x_i) = y_i - x_i \tag{2.29}$$

so for each i we have $y_i = x_i$ and thus $P = Q$.

We have therefore shown that $\mathcal{V}(I) = \{P\}$, hence $\{P\}$ is algebraic. ∎

We actually made a few important observations (or at least came close to doing so) in that proof, which we collect in the following rather non-trivial theorem.

2.5.18 Theorem. *Fix $P = (x_1, ..., x_n) \in \mathbb{A}^n$ and $I = ((X_1 - x_1), ..., (X_n - x_n))$. Then:-*

1. $\mathcal{V}(I) = \{P\}$.
2. $I \subseteq \ker \text{eval}_P$.
3. $\ker \text{eval}_P$ *is a maximal ideal in* $K[X_1, .., X_n]$.
4. $I = \ker \text{eval}_P$.
5. $\mathcal{I}(\{P\}) = I$.

Proof.

1. Was shown in the proof of the above proposition.
2. This is equation 2.27 and the preceeding discussion in the proof of the above proposition.
3. Note that $\text{eval}_P : K[X_1, ..., X_n] \to K$ is surjective: if $x \in K$ then $\text{eval}_P\, x = x$. Hence by the first homomorphism theorem (corollary 2.1.8) we have
$$K \simeq \frac{K[X_1, ..., X_n]}{\ker \text{eval}_P} \tag{2.30}$$
and hence by definition we have $\ker \text{eval}_P$ is maximal.
4. By part 2, $I \subseteq \ker \text{eval}_P$; thus it suffices to show that $f \in \ker \text{eval}_P \implies f \in I$. Proceed by induction on n: If $n = 1$ then $\ker \text{eval}_{x_1} f = (X_1 - x_1)$ by corollary 2.3.4.

   If $n > 1$, assume that for all $m < n$ we have for $(x_1, ..., x_m) = Q \in \mathbb{A}^n$ that $\ker \text{eval}_Q = \big((X_1 - x_1), ..., (X_m - x_m)\big)$. Suppose $f \notin I$. We shall show that $f \notin \ker \text{eval}_P$.

   Since $f \notin I$, $f = g + g_1(X_1 - x_1) + \cdots + g_n(X_n - x_n)$ where for all i we have $(X_i - x_i) \nmid g$. Fix i and consider the following commutative diagram (c.f. equation 2.28):

$$\begin{array}{ccc} K[X_1, ..., X_n] & \xrightarrow{\text{eval}_{(x_1,...,x_{i-1},X,x_{i+1},...,x_n)=\varphi}} & K[X] \\ & \searrow^{\text{eval}_P} & \downarrow{\scriptstyle \text{eval}_{x_i}} \\ & & K \end{array} \tag{2.31}$$

Note that $K[X_1, ..., X_n] \simeq (K[X_i])[X_1, ..., X_{i-1}, X_{i+1}, ..., X_n]$. Thus the diagram becomes:

$$\begin{array}{ccc} (K[X])[X_1, ..., X_{i-1}, X_{i+1}, ..., X_n] & \xrightarrow{\text{eval}_{(x_1,...,x_{i-1},x_{i+1},...,x_n)}=\varphi} & K[X] \\ & \searrow^{\text{eval}_P} & \downarrow \text{eval}_{x_i} \\ & & K \end{array} \tag{2.32}$$

We claim that $(X - x_i) \nmid \varphi(g)$. (Indeed, suppose it did; then we may write $\varphi(g) = (X - x_i)g'$ for some $g' \in K[X]$; but $K[X] \simeq \frac{(K[X])[X_1,...,X_{i-1},X_{i+1},...,X_n]}{\ker \varphi}$ by the inductive hypthesis (noting as in part 3 that evaluation is surjective) and so we might write

$$g = (X_i - x_i)\varphi^{-1}(g') + \sum_{j \in \{1,...,i-1,i+1,...,n\}} r_i(X_j - x_j) \tag{2.33}$$

(where $\varphi^{-1}(g')$ stands for some particular arbitrary inverse image of g') which lies in the ideal I, a contradiction.)

But $\varphi(g) \in K[X]$, so by corollary 2.3.4 we must have $\text{eval}_{x_i} \varphi(g) \neq 0$. In particular,

$$0 \neq \text{eval}_{x_i} \varphi(g) = \text{eval}_{x_i} \left(\text{eval}_{(x_1,...,x_{i-1},X,x_{i+1},...,x_n)} g \right) = \text{eval}_P g \tag{2.34}$$

so $f \notin \ker \text{eval}_P$.

5. By part 4 we may show that $\mathcal{I}(\{P\}) = \ker \text{eval}_P$. But this is trivial: $f \in \mathcal{I}(\{P\}) \iff f(P) = 0 \iff f \in \ker \text{eval}_P$. ∎

2.5.19 Remark. Note that the proofs for parts 2, 3, and 4 are valid in arbitrary rings since that those are the environment for which we proved corollary 2.3.4. This remark will allow us to take advantage of the first isomorphism theorem in order to prove various rings of interest are isomorphic.

2.5.20 Remark. Part 4 of the theorem concerns evaluation within a single integral domain. If $\phi : R \to S$ is an injective homomorphism of IDs, we may write that (for $P = (x_1, ..., x_n) \in S^n$)

$$\ker \text{eval}_{\phi,P} = ((X_1 - \phi^{-1}(x_1)), ..., (X_n - \phi^{-1}(x_n))). \tag{2.35}$$

Indeed, we may decompose the evaluation as follows -

$$\begin{array}{ccc} R[X_1, ..., X_n] & \xrightarrow{\tilde{\phi}} & S[X_1, ..., X_n] \\ & \searrow^{\text{eval}_{\phi,P}} & \downarrow \text{eval}_P \\ & & S \end{array} \tag{2.36}$$

Here $\tilde{\phi}$ is the extension of ϕ sending $X_i \mapsto X_i$ for all i.

By the theorem, points of $\mathbb{A}^n$ correspond to maximal ideals in $K[X_1, ..., X_n]$. The converse is not necessarily true in general (though it is true if the field is algebraically closed — see corollary 2.11.4):

2.5.21 Example. In $\mathbb{R}[x]$, $(x^2 + 1)$ is maximal. Indeed, consider the homomorphism $\text{eval}_i : \mathbb{R}[x] \to \mathbb{C}$; this is clearly surjective (since $a + bi = \text{eval}_i(a + bx)$), and has kernel $(x^2 + 1)$: it is obvious that $\ker \text{eval}_i \supseteq (x^2 + 1)$, so suppose $f \notin (x^2 + 1)$; then write $f = (x^2 + 1)q + r$ with $\deg r < 2$ by the division algorithm, so $\text{eval}_i f = \text{eval}_i r$. By assumption, $r \neq 0$. Hence r is of the form $r = ax + b$ for real a, b (at most one of which is zero); so $f(i) = ai + b \neq 0$ and $f \notin \ker \text{eval}_i$. Thus $\mathbb{C} \simeq \frac{\mathbb{R}[x]}{\ker \text{eval}_i} = \frac{\mathbb{R}[x]}{(x^2+1)}$ (by the first homomorphism theorem, corollary 2.1.8) and so $(x^2 + 1)$ is maximal.

This argument corresponds to an argument about 'irreducibility' of $(x^2 + 1)$ in $\mathbb{R}[x]$, a concept which we shall develop in the next section in great detail as it has a nice geometric interpretation.

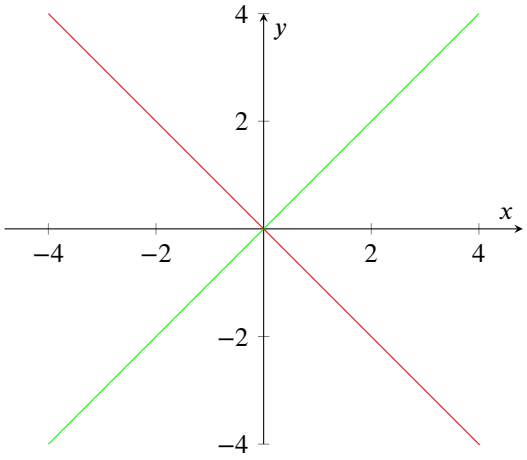


Figure 2.2: $\mathcal{V}(X^2 - Y^2)$.

2.6 Irreducibility and factorisation

Consider the polynomial $f(X,Y) = X^2 - Y^2 \in \mathbb{R}[X,Y]$. The zero-set of f, pictured in 2.2, is clearly the union of two smaller zero-sets: $\mathcal{V}(X^2 - Y^2) = \mathcal{V}(X - Y) \cup \mathcal{V}(X + Y)$ (this is either by direct calculation, or by use of the calculus, proposition 2.4.7).

2.6.1 Definition. An algebraic set V is **reducible** if there exist two non-empty algebraic sets V_1 and V_2 such that $V = V_1 \cup V_2$.

More generally, a topological space X is **reducible** if there exist two non-empty closed sets X_1 and X_2 such that $X = X_1 \cup X_2$.

An algebraic set or topological space is **irreducible** if it is not reducible.

In some sense, irreducible sets should be connected. Let us recall what 'connectedness' is in the topological setting.

2.6.2 Theorem. *Let X be a topological space. The following are equivalent:*

1. *There exist disjoint non-empty proper open subsets $U, V \subset X$, such that $U \cup V = X$.*
2. *There exist disjoint non-empty proper closed subsets $F, G \subset X$, such that $F \cup G = X$.*
3. *There exists a clopen set K such that $\emptyset \subset K \subset X$.*

*If any (and thus all) hold, X is called **disconnected**; if X is not disconnected, it is called **connected**.*

Proof.

1 $\implies$ 2**.** Suppose U, V are disjoint non-empty proper open subsets of X whose union is X. Then $X \setminus U$ and $X \setminus V$ are are disjoint non-empty proper closed subsets of X whose union is X.

2 $\implies$ 3**.** Suppose F, G are disjoint non-empty proper closed subsets of X whose union is X. Then $X \setminus G = F$, so F is a non-empty clopen proper subset of X.

1 $\implies$ 2**.** Suppose K is a non-empty clopen proper subset of X; then K is open, and $X \setminus K$ is an open non-empty proper subset of X, and $(X \setminus K) \cup K = X$. ∎

2.6.3 Proposition. *Irreducible topological spaces are connected.*

Proof. Suppose X is a disconnected space; then $X = F \cup G$ for F, G disjoint non-empty proper closed subsets of X; so X is reducible. ∎

2.6.4 Example. It is not necessarily true that reducible sets are disconnected. For example, the set we started with — $\mathcal{V}(X^2 - Y^2)$ — is connected. A pure topological proof is easy — there are a couple of ways of doing it, e.g. by showing it is path-connected, or by noting that it is a union of homeomorphic images of $\mathbb{R}$ that share a point (there are some details to fill in in either case).

2.6.5 Example. Note that 'connectedness' in the Zariski topology is a little counterintuitive: $\mathcal{V}(XY - 1)$ is irreducible and hence connected in the Zariski topology, but over $\mathbb{R}$ its graph is disconnected under the usual topology.

We can begin by strengthening an earlier result, proposition 2.5.13.

2.6.6 Proposition. *If $X \subseteq \mathbb{A}^n$ is irreducible and endowed with the Zariski topology, then every non-empty open set is dense in X.*

Proof. It suffices to show that every pair of non-empty open sets intersects.

Suppose $U, V \subseteq X$ are open such that $U \cap V = \emptyset$. By openness, there exist vanishing ideals I, J such that $U = X \setminus \mathcal{V}(I)$ and $V = X \setminus \mathcal{V}(J)$. Thus

$$\emptyset = X \setminus (\mathcal{V}(I) \cup \mathcal{V}(J)) = X \setminus \mathcal{V}(I \cap J). \tag{2.37}$$

Hence $\mathcal{V}(I \cap J) \supseteq X$; i.e. $\mathcal{V}(I) \cup \mathcal{V}(J) \supset X$. By the irreducibility condition, either $\mathcal{V}(I) \supseteq X$ (so $U = \emptyset$) or $\mathcal{V}(J) \supseteq X$ (so $V = \emptyset$). ∎

Irreducibility is a condition on zero-sets, so we would like to find a condition on vanishing ideals that corresponds to it. Note that in the calculus (proposition 2.4.7) we have that $\mathcal{V}(I) \cup \mathcal{V}(J) = \mathcal{V}(IJ)$. Thus our condition should be something like 'not factoring into a product of larger[1] ideals'.

2.6.7 Definition. An ideal I is **reducible** if there exist two ideals I_1, I_2 such that $I \subset I_1, I_2 \subset R$ and $I = I_1 I_2$. An ideal is **irreducible** if it is not reducible.

Let us now prove the standard 'our definition which we pulled out of thin air actually does what we want it to' theorem.

2.6.8 Theorem. *An algebraic set V is irreducible iff $\mathcal{I}(V)$ is irreducible.*

Proof. Suppose V is reducible, and let $\mathcal{I}(V) = I$ be its vanishing ideal. Then $V = V_1 \cup V_2$ for two non-empty proper algebraic subsets V_1 and V_2. Thus $\mathcal{I}(V) = \mathcal{I}(V_1 \cup V_2) = \mathcal{I}(V_1)\mathcal{I}(V_2)$ is decomposable as a product of larger ideals and is therefore reducible.

Conversely, suppose $\mathcal{I}(V) = IJ$ for two larger non-unit ideals I, J. Then $\mathcal{V}(\mathcal{I}(V)) = \mathcal{V}(IJ) = \mathcal{V}(I) \cup \mathcal{V}(J)$; and since V is algebraic, by proposition 2.4.4 we are done. ∎

2.6.9 Remark. We will not be able to prove the dual result — that $\mathcal{V}(I)$ is irreducible if I is irreducible — for a while. It will emerge as a corollary to the Nullstellensatz (corollary 2.11.11).

2.6.10 Proposition. *Let I be an ideal of R. If I is a prime ideal then I is an irreducible ideal.*

Proof. Suppose I is reducible. Then there are ideals I_1, I_2 in R such that $I \subset I_1, I_2 \subset R$. Pick $a \in I_1 \setminus I$ and $b \in I_2 \setminus I$; then $ab \in I_1 I_2 = I$, so I is not prime. ∎

[1] Since taking ideal products produces a smaller ideal.

There is a partial converse to this:

2.6.11 Proposition. *Let I be a vanishing ideal in $\mathbb{A}^n_K$. If I is irreducible then I is prime.*

Proof. Suppose I is not prime. Then there exist $f, g \in K[X_1, ..., X_n]$ such that $fg \in I$ but neither $f \in I$ nor $g \in I$ (and in particular neither f nor g is the zero polynomial). Let $V = \mathcal{V}(I)$. Then for all $P \in V$, $(fg)(P) = 0$. But K is an integral domain, so either $f(P) = 0$ or $g(P) = 0$. Hence $V \subseteq \mathcal{V}(f) \cup \mathcal{V}(g)$, with V not a subset of either of the vanishing sets on the right; so V is reducible and thus $\mathcal{I}(V) = I$ is reducible. ∎

2.6.12 Exercise. Check that the above proposition is true if the field K is replaced by an arbitrary integral domain. (We shall not use this more general result.)

We are often interested in ideals generated by single polynomials. At the risk of overloading one word too often, we make the following

2.6.13 Definition. Let R be an integral domain and let $r \in R$. Then r is **irreducible** if (r) is an irreducible ideal.

2.6.14 Proposition. *An element r is irreducible in an integral domain R iff whenever $a, b \in R$ such that $ab = r$, one of a or b is a unit.*

Proof. Suppose r is reducible; then $(r) = IJ$ for two ideals I and J such that neither I nor J is (r), and neither is the full ring. In particular we may choose $a \in I$ and $b \in J$ such that $ab = r$, and neither a nor b is a unit since neither I nor J is the full ring.

Conversely, suppose that r is irreducible; then if $a, b \in R$ such that $ab = r$, either (a) is a unit ideal or (b) is a unit ideal. Suppose it is (a); then $ak = 1$ for some $k \in R$, and a is a unit. Similarly, if (b) is a unit ideal then b is a unit. ∎

There are a number of results which allow us to prove that various polynomials in one variable are irreducible. We quote them without proof:

2.6.15 Theorem.

1. *If $f \in \mathbb{R}[X]$, and $\partial f > 2$, then f is reducible.*
2. *If $f \in \mathbb{C}[X]$ is non-constant, f is irreducible iff $\partial f = 1$.*
3. *Let $f \in \mathbb{Z}[X]$ has entirely coprime coefficients (i.e. the ideal of $\mathbb{Z}$ generated by the set of coefficients is (1)), and let p be a prime integer. Assume that the polynomial f_p obtained by the natural map $\mathbb{Z}[X] \to (\mathbb{Z}/p\mathbb{Z})[X]$ has the same degree as f, and is irreducible in $\mathbb{Z}/p\mathbb{Z}$. Then f is irreducible in $\mathbb{Z}[X]$.*
4. *(Eisenstein) Let R be a ring, and let P be a prime ideal of R. Let $f = \sum_{i=0}^{N} f_i X^i \in R[X]$, and suppose that*
   - *$a_N \notin P$;*
   - *$a_i \in P$ for $i \in \{0\} \cup [n-1]$;*
   - *$a_0 \notin P^2$.*

   Then f is not the product of polynomials of lower degree in $R[X]$.

Proof.

1. [Alu09, proposition V.5.13]

2. [Alu09, proposition V.5.13]

3. [Alu09, proposition V.5.15]

4. [Alu09, proposition V.5.17] ∎

2.6.16 Exercise. Is $Y^5 + X^2Y^3 + X^3Y^2 + X$ irreducible in $\mathbb{C}[X, Y]$? [Alu09, exercise V.5.23]

These results motivate us to take a closer look at the notion of factorisation in a ring. It will be convenient to work only in integral domains, because nasty things happen in more general rings: for example, $2 \cdot 2 = 4 = 2 \cdot 5$ in $\mathbb{Z}/6\mathbb{Z}$, and so we don't have factorisations in any unique sense. The general objects we are interested in will be defined as follows.

2.6.17 Definition. Let R be an integral domain. An element $r \in R$ has a **factorisation into irreducibles** if there exist irreducible elements $r_1, ..., r_n$ such that $r = r_1 \cdots r_n$. The factorisation is a **unique factorisation** if whenever $r_1 \cdots r_n = r'_1 \cdots r'_n$ are two factorisations into irreducibles, there exists a permutation $\pi : [n] \to [n]$ and units $u_1 \cdots u_n$ such that $r'_i = u_i r_{\pi(i)}$ for all i.

The integral domain R is called a **factorisation domain** if every non-unit, non-zero element has a factorisation into irreducibles; the ring is called a **unique factorisation domain** or **UFD** if it is a factorisation domain and every factorisation is unique.

Our goal, which we will achieve in the next section, is to prove that $R[X_1, ..., X_n]$ is a UFD if and only if R is a UFD; this result will enable us to split any algebraic set up into finitely many irreducible sets in a unique way.

2.6.18 Theorem. *In a ring R, the following are equivalent:*

Ascending chain condition (ACC). *If $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ is an ascending chain of ideals, then there is some $n \in N$ such that $I_n = I_{n+1} = \cdots$. We say the chain terminates.*

Noetherian property. *If I is an ideal of R, then there exist finitely many elements $r_1, ..., r_n$ such that $I = (r_1, ..., r_n)$; that is, every ideal is finitely generated.*

Weak Zorn condition. *If X is a non-empty set of ideals of R, then X has a maximal element under inclusion.*

*A ring satisfying one (and thus all) of these properties is called **Noetherian**.*[2]

Proof.

$1 \implies 2$**.** Let I be an ideal of R. Every finite ideal of R is clearly finitely generated, so we may assume I has infinitely many elements. If I is not finitely generated, we may pick a sequence of elements $r_i \in I$ such that $(r_1) \subset (r_1, r_2) \subset \cdots$; this contradicts the ascending chain condition.

$2 \implies 3$**.** Let X be a non-empty set of ideals of R. Let $I_1 \subseteq I_2 \subseteq \cdots$ be a chain of ideals in X; the union $I = \cup_{i \in \mathbb{N}} I_i$ is an ideal of R, and is finitely generated by the Noetherian property. Since the ideals form an inclusion chain, there must be some ideal I_N which contains all the generators of I; then $I_N = I$, and $I_i = I$ for all $i \geq N$; so I_N is a maximal element of X.

$3 \implies 1$**.** Let $I_1 \subseteq I_2 \subseteq \cdots$ be an ascending chain of ideals in R; then $X = \{I_i\}_{i \in \mathbb{N}}$ is a non-empty set of ideals and hence has a maximal element I_N under inclusion; hence for all $i \geq N$, $I_i = I_N$ and so the chain terminates. ∎

2.6.19 Example.

[2] Noetherian here, and in the topological space definition below, refers to Emmy Noether; see the *dramatis personae*.

1. Every field is Noetherian. Indeed, every field has only two ideals (remark 2.1.13), so every set of ideals — being finite — must have a maximal element.

2. More generally, every PID is Noetherian. This gives us a range of other examples; see example 2.3.7.

Before we start doing anything difficult, the following is a nice property of Noetherian rings and provides an example of the proof techniques they enable.

2.6.20 Proposition. *If R is a (non-zero!) Noetherian ring, then R has a maximal ideal.*

Proof. If R is finite, then the result is clearly trivial; it therefore suffices to consider infinite rings.

Suppose R is an infinite ring with no maximal ideals. Pick some $r_1 \neq 0$ in R, and let $I_1 = (r_1)$. Since I_1 is not maximal, there is some non-unit ideal I_2 such that I_1 is strictly contained in I_2. Repeating this process produces a strictly ascending chain $I_1 \subset I_2 \subset \cdots$, and so R is not Noetherian. ∎

The next two theorems will give us a wide range of examples of Noetherian rings.

2.6.21 Theorem. *If R is Noetherian, then R/I is Noetherian for all ideals $I \subseteq R$.*

Proof. Let $I_1 \subseteq I_2 \subseteq \cdots$ be an ascending chain in R/I. By the correspondence theorem, corollary 2.1.10, this corresponds to an ascending chain of ideals $J_1 \subseteq \cdots$ in R; but R is Noetherian, so this chain terminates and hence the chain in R/I terminates. ∎

2.6.22 Theorem (Hilbert basis theorem)**.** *Let R be a ring. Then R is Noetherian iff $R[X]$ is Noetherian.*

Proof. Suppose $R[X]$ is Noetherian; then any ideal I in R is an ideal I' in the isomorphic copy of R that lies in $R[X]$; this ideal is finitely generated; so I is finitely generated.

The converse is the hard direction. Suppose R is Noetherian, and let I be an ideal in $R[X]$; we will show that I is finitely generated. Form the sets $I_n \subseteq R[X]$ and $J_n \subseteq R$ for $n \in \mathbb{N}_0$, defined as follows:

$$\begin{aligned} I_n &:= \{I \in I : \partial f = n\} \\ J_n &:= \{r \in R : \exists_{f \in I_n} \text{ such that } \mathrm{LC}(f) = r\} \end{aligned} \tag{2.38}$$

(recall that $\mathrm{LC}(f)$ is the leading coefficent of f, definition 2.2.1).

▶ **Claim A.** *Each J_n is an ideal of R.*

Firstly, let $j \in J_n$ and $r \in R$. By definition of J_n there exists some $f \in I$ such that $j = \mathrm{LC}(f)$ and $\partial f \leq n$. Since I is an ideal, $rf \in I$, and multiplying by a constant does not change the degree; then $rj = r\,\mathrm{LC}(f) = \mathrm{LC}(rf) \in J_n$. Now suppose $j, k \in J_n$, such that $j = \mathrm{LC}(f)$ and $k = \mathrm{LC}\, g$ for some polynomials $f, g \in I$. Without loss of generality assume $\partial f \leq \partial g$. Hence $j + k = \mathrm{LC}(f) + \mathrm{LC}(g) = \mathrm{LC}(h + g) \in J_n$ (noting that the sum of polynomials has degree at most n, and we may multiply by powers of X to increase the degree to n exactly if necessary). ◀

Hence the $J_1 \subseteq J_2 \subseteq \cdots$ form an ascending chain of ideals in R; since R is Noetherian there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $J_n = J_N$. For each J_n such that $n \leq N$, there is a finite set of elements generating J_n; call this set K_n, and set $K = \cup J_n$. Since the chain of J_n ideals terminates, K is finite.

▶ **Claim B.** *The ideal I is generated by K.*

Suppose $g \in I$ is of minimal degree such that g is not generated by K. Then $\mathrm{LC}(g)$ is contained within one of the ideals $J_1, ..., J_N$; let k be the smallest integer such that $\mathrm{LC}(g) \in J_k$. Then there are elements $h_1, ..., h_\ell \in I_k \subseteq K$ such that $\mathrm{LC}(g) = \mathrm{LC}(h_1 + \cdots + h_\ell)$. Let

$$h = X^{\partial g - \partial(h_1 + \cdots + h_\ell)}(h_1 + \cdots + h_k); \tag{2.39}$$

then $h \in I$ so $g - h \in I$; and the leading terms of g and h cancel, so $\partial(g-h) < \partial g$. Further, $g - h$ cannot be generated by K, since h is generated by k and so $g = (g-h) + h$ would be generated by K. But this contradicts minimality of ∂g. ◀

Thus I is finitely generated as an ideal of $R[X]$. ∎

The main example is clearly the following:

2.6.23 Example. The polynomial ring $K[X_1, ..., X_n]$ is Noetherian.

In particular, we can say something about irreducible sets in affine space.

2.6.24 Proposition. *If X is a topological space, the following are equivalent:*

Descending chain condition (DCC). *If $Y_1 \supseteq Y_2 \supseteq I_3 \supseteq \cdots$ is an decending chain of closed sets, then there is some $n \in N$ such that $Y_n = Y_{n+1} = \cdots$. We say the chain terminates.*

Weak Zorn condition. *If $\mathscr{A}$ is a non-empty set of closed sets of X, then $\mathscr{A}$ has a minimal element under inclusion.*

*A space satisfying either (and thus both) of these properties is called **Noetherian**.*

The proof is analogous to theorem 2.6.18.

Proof.

$1 \implies 2$**.** Suppose $\mathscr{A}$ is a non-empty set of closed sets of X without a minimal element under inclusion. Then one can inductively construct a descending chain of closed sets in $\mathscr{A}$.

$2 \implies 1$**.** If $Y_1 \supseteq Y_2 \supseteq \cdots$ is a descending chain, then the set $\{Y_i\}_{i\in\mathbb{N}}$ has a minimal element Y_N; in particular, for all $i \geq N$, $Y_i = Y_N$ and the chain terminates. ∎

2.6.25 Example. The affine space $\mathbb{A}^n$ is Noetherian: if $Y_1 \supseteq Y_2 \supseteq I_3 \supseteq \cdots$ is an decending chain of closed sets, then the set of ideals $\mathcal{I}(Y_1) \subseteq \mathcal{I}(Y_2) \subseteq \mathcal{I}(Y_3) \subseteq \cdots$ is an ascending chain of ideals, which terminates by example 2.6.23; thus by the calculus of the $\mathcal{I}$–$\mathcal{V}$ correspondence, the closed set chain terminates.

2.6.26 Proposition. *Every algebraic subset $X \subseteq \mathbb{A}^n$ is the intersection of finitely many hypersurfaces.*

Proof. By example 2.6.23, $\mathcal{I}(X)$ is generated by finitely many elements $f_1, ..., f_k$; but by the calculus, we have $X = \mathcal{V}(\mathcal{I}(X)) = \mathcal{V}((f_1, ..., f_k)) = \mathcal{V}(f_1, ..., f_k) = \cap_{i=1}^{k} \mathcal{V}(f_i)$. ∎

2.6.27 Theorem. *Let X be a Noetherian topological space and let $Y \subseteq X$ be closed. Then there is a decomposition of Y into irreducible sets,*

$$Y = Y_1 \cup \cdots \cup Y_n, \tag{2.40}$$

*where the Y_k are irredundant (i.e. $Y_i \subseteq Y_j \implies i = j$) and are uniquely determined up to order. The Y_k are called the **components** of Y.*

Proof. Let $\mathscr{B}$ be the set of subsets that do not have a decomposition into irreducibles; suppose for the sake of contradiction that $\mathscr{B}$ is non-empty. Then, by the weak Zorn condition, $\mathscr{B}$ has a minimal element Y. But Y cannot be reducible; hence $Y = V \cup W$ for two closed sets V and W strictly included in Y. Because Y is minimal, neither V nor W lie in $\mathscr{B}$; in particular, they have decompositions into irreducibles, which is a contradiction since the union of the decompositions forms Y.

We will now prove uniqueness of this decomposition. Suppose $Y = \cup_{i\in I} V_i = \cup_{j\in J} W_j$ can be written as a product of irredundant irreducibles in two different ways. Then there exists some $i \in I$ such that for all $j \in J$, $V_i \neq W_j$. For each $j \in J$, set $\tilde{W}_j := V_i \cap W_j$. Then $V_i = \cup_{j\in J} \tilde{W}_j$. There are two cases:

$V_i \neq \tilde{W}_j$ **for all** j**.** Then V_i is a finite union of closed proper subsets, contradicting its irreduciblity.

$V_i = \tilde{W}_j$ **for some** j**.** So $V_i \subset W_j$ (note the inclusion is proper). For each $i \in I$, set $\tilde{V}_i = V_i \cap W_j$. Then we may write $W_j = \cup_{i \in I} \tilde{V}_i$, where each $\tilde{V}_i$ is a proper closed subset of W_j, contradicting its irreduciblity. ∎

Let X be a Noetherian topological space.

2.6.28 Exercise. 1. Show that if $Y \subseteq X$ is given the induced topology then Y is Noetherian.

2. Show that X is **compact**: that is, for every family $\{U_\alpha\}_{\alpha \in I}$ of open sets of X such that $\cup_\alpha U_\alpha = X$, there exist finitely many $\alpha_1, ..., \alpha_n$ such that $\cup_{i=1}^n U_{\alpha_i} = X$.

   (It is traditional in algebraic geometry to use the term **quasi-compact** to describe this notion, and to reserve the term 'compact' for spaces which are also Hausdorff.)

3. Show that every subspace of X is compact.

4. State and prove Tychonoff's theorem for Noetherian topological spaces.

2.6.29 Exercise. If X is an irreducible non-empty topological space, then every open subset $U \subseteq X$ is dense and irreducible.

2.7 Characterisation of unique factorisation domains

2.7.1 Definition. Let R be an integral domain. An element $r \in R$ is **prime** if (r) is a prime ideal.

2.7.2 Exercise. An element $r \in R$ is prime iff whenever $r \mid ab$, either $r \mid a$ or $r \mid b$.

2.7.3 Lemma. *Let R be an integral domain, and let $r \in R$ be prime. Then r is irreducible.*

Proof. Suppose $r \in R$ is reducible; that is, $(r) = IJ$ for two proper ideals I, J of R and such that $I \neq (r)$, $J \neq R$. Pick $a \in I \setminus (r)$, $b \in J \setminus (r)$. Then $ab \in (r)$; i.e. $r \mid ab$. But $a \notin (r)$ and $b \notin (r)$, so r is not prime. ∎

2.7.4 Lemma. *Let R be a UFD, and let $r \in R$ be irreducible. Then r is prime.*

Proof. Suppose $r \in R$ is irreducible. Then r is not a unit. Assume $ab \in (r)$ for some $a, b \in R$; so $(ab) \subseteq (r)$. Factorise a and b into irreducibles as $a = a_1 \cdots a_N$, $b = b_1 \cdots b_M$; then $a_1 \cdots a_N b_1 \cdots b_M = rk$ for some $k \in R$. Hence by uniqueness of the factorisation of rk, there exists some factor c on the left hand side such that $r = uc$ for some unit c. If c is a factor of a then $r \mid a$ and $a \in (r)$; if c is a factor of b then $r \mid b$ and $b \in (r)$. ∎

We may now prove another large classification theorem.

2.7.5 Theorem. *An integral domain R is a UFD if and only if the following conditions hold:*

1. *If $a_1, a_2, ...$ are elements of R, then the ascending chain $(a_1) \subseteq (a_2) \subseteq \cdots$ terminates.*

2. *Every irreducible element of R is prime.*

Proof. Suppose R is a UFD. Consider any ascending strict chain $(a_1) \subset \cdots$. Then $a_1 \in (a_k)$ for all $k \geq 1$; i.e. for all $k \geq 1$, $a_k \mid \cdots \mid a_1$. Since R is a UFD, this implies that a_k must eventually be irreducible or a unit. In either case, $a_{k+1} = ua_k$ for some unit u, and thus $(a_{k+1}) = (a_k)$ (contradicting the existence of a strict ascending chain). The second condition, every irreducible element is prime, is just the previous lemma.

Conversely, suppose the two conditions hold in an integral domain R.

▶ **Claim A.** *Every element $r \in R$ has a factorisation into irreducibles.*

Let $r \in R$; assume it does not have a factorisation. Then r is not irreducible, so there exist $r_1, s_1 \in R$ such that neither r_1 nor s_1 is a unit and $r_1 s_1 = r$. If both r_1 and s_1 have factorisations, then so does r; so r_1 does not have a factorisation and $(r) \subset (r_1)$ (since r_1 strictly divides r). Iterating this argument produces a strictly ascending chain of principal ideals of r, contradiction. ◀

▶ **Claim B.** *Factorisations in R are unique.*

Let $r \in R$ be factorised into irreducibles as $p_1 \cdots p_n$ and $q_1 \cdots q_m$. Then $q_1 \cdots q_m = r \in (p_1)$; and by condition 2, (p_1) is prime; so one of the q_i — we may assume $i = 1$ — lies in (p_1). Hence $q_1 = up_1$ for some $u \in R$, and thus (since q_1 is irreducible) we have that u is a unit. We may therefore cancel q_1 from both sides of the equality $p_1 \cdots p_n = q_1 \cdots q_m$; iterating this argument finishes the proof. ◀ ∎

2.7.6 Corollary. *If R is a PID, then R is a UFD.*

Proof. Every PID is Noetherian (example 2.6.19) and hence criteria 1 is met in the theorem. We just need to show that criteria 2 is met: that irreducibles are prime. Suppose r is irreducible in R. If $(r) \subset (a) \subseteq R$, then a is a unit; indeed, we have a strict divisibility $a \mid r$. Hence (r) is maximal among the principal ideals of R, and thus is maximal in R. But maximal ideals are prime almost by definition, so we are done. ∎

2.7.7 Example. The converse is not true: not all UFDs are PIDs. Once we have proved that $R[X]$ is a UFD whenever R is a UFD, example 2.3.9 will furnish us with a counterexample.

2.7.8 Lemma (Gauss)**.** *Let R be a UFD. An element $h \in R[X]$ is **primitive** if no prime element of R divides all the coefficients of h.*

The product of primitive polynomials is primitive.

Proof. Let $f, g \in R[X]$ be primitive, and let $p \in R$ be prime. The ring $R/(p)$ is an integral domain and so $R/(p)[X]$ is an integral domain (lemma 2.2.3); since p does not divide all the coefficients of f and g in R, they do not all lie in the kernel of the residue map $R \to R/(p)$ and so the polynomials do not lie in the kernel of the natural map $\pi : R[X] \to R/(p)[X]$. Thus the product $\pi(f)\pi(g) = \pi(fg)$ is non-zero, and fg does not lie in the kernel of π — it is not in the ideal (p). ∎

The following theorem allows us to clear denominators.

2.7.9 Theorem. *Let R be a UFD, and let $K = \operatorname{Frac} R$.*

1. *If $f \in K[X]$, then $f = \lambda f'$ for some $\lambda \in K$ and some primitive $f' \in R[X]$.*
2. *If $g, f \in R[X]$ such that g is primitive and $g \mid f$ in $K[X]$, then $g \mid f$ in $R[X]$.*
3. *If $f \in R[X]$ is nonconstant and reducible in $K[X]$, then f is reducible in $R[X]$.*

Proof.

1. Suppose $f \in K[X]$, and suppose that f may be written as

$$f = \sum_{i=0}^{n} \frac{a_i}{b_i} X^i \tag{2.41}$$

for $a_i, b_i \in R$ such that $b_i \neq 0$ for all i. We first clear denominators:

$$f = (b_1 \cdots b_n) \sum_{i=0}^{n} (a_i \prod_{j \in [n] \setminus \{i\}} b_j) X^i := (b_1 \cdots b_n) \sum_{i=0}^{n} c_i X^i \tag{2.42}$$

Let $p_1, ..., p_m$ be the set of all irreducible elements dividing any of $c_1, ..., c_n$, with no repetitions. By uniqueness of factorisations, there exists some maximal subset $U_1 \subseteq [m]$ such that $\prod_{j \in U_1} p_j \mid c_i$ for all j; set $c_i^1 = \frac{c_i}{\prod_{j \in U_1} p_j}$. Since factorisations are finite, we may continue factoring out products of irreducibles in this way until we obtain some sequence c_i^k of coefficients with no common irreducible factors; by lemma 2.7.3 and lemma 2.7.4, in a UFD primality and irreducibility are equivalent; so the coefficients have no common prime factors and we obtain the required demonstration,

$$f = (b_1 \cdots b_n) \left(\prod_{j \in U_1} p_j \right) \cdots \left(\prod_{j \in U_k} p_j \right) \sum_{i=0}^{n} c_i^k X^i \tag{2.43}$$

2. We may write $f = gh$ for some $h \in K[X]$. By part 1, factorise h as $\lambda h'$ for $\lambda = \frac{a}{b} \in K$ and $h' \in R[X]$ primitive. Thus $bf = agh'$, and gh' is primitive by Gauss' lemma. Thus $b \mid a$ (since otherwise it would divide each coefficient of gh'), and $\lambda \in R$, so $f = (\lambda h')g$ is a factorisation in $R[X]$.

3. Suppose $f = gh$ in $K[X]$. We may write $g = \lambda g'$ for g' primitive by part 1, so $f = \lambda h g'$; then by part 2, since g' is a primitive divisor of f in $K[X]$, g' divides f in $R[X]$; so $\lambda h \in R[X]$ by unique factorisation and thus f is reducible in $K[X]$. ∎

Lo!

2.7.10 Theorem. *Let R be an integral domain. Then R is a UFD iff $R[X]$ is a UFD.*

Proof. If $R[X]$ is a UFD then R is trivially a UFD by inclusion.

Conversely, suppose R is a UFD. We will verify the two conditions of theorem 2.7.5.

Firstly, suppose $f \in R[X]$ is irreducible. Suppose $f \mid gh$ in $R[X]$; by parts 1 and 2 of the previous theorem, we may write $g = rg'$, $h = sh'$ for elements $r, s \in R$ and primitive polynomials $g', h' \in R[X]$. If f is constant, it does not divide $g'h'$ since the latter is primitive; thus it divides rs and hence either $f \mid r$ or $f \mid s$ since in R irreducible elements are prime (lemma 2.7.4). On the other hand, if f is non-constant then it is irreducible in $K[X]$ by taking the contrapositive of part 3 of the previous theorem. In particular, since by corollary 2.7.6 $K[X]$ is a UFD and irreducible elements are prime, $f \mid g'$ or $f \mid h'$; and by part 3 of the previous theorem again, these divisibility relations hold in $R[X]$.

Secondly, suppose $(f_1) \subset (f_2) \subset \cdots$ is an ascending chain in $R[X]$. Then for all $k \geq 1$, $f_k \mid \cdots \mid f_1$. In particular, f_1 is reducible; if it is constant, all its divisors are constant and so the chain terminates since R is a UFD. On the other hand, if f_1 is non-constant then we might write $f_1 = f_2 g_2$, $f_2 = f_3 g_3$, and in general $f_k = f_{k+1} g_{k+1}$ for some $g_i \in R[X]$. By part 1, rewrite each of these divisibilities in $K[X]$ as a product of $\lambda_i g_i$ for primitive g_i. Write f_1 as a product of irreducibles in $K[X]$; then by part 2 of the preceeding theorem, the g_i all strictly divide f in $K[X]$ which is a contradiction. ∎

2.7.11 Corollary. *If K is a field, $K[X_1, ..., X_n]$ has infinitely may irreducible elements.*

Proof. Suppose $\pi_1, ..., \pi_n$ are irreducible; then there $\pi_1 \cdots \pi_n + 1$ is an element of $K[X_1, ..., X_n]$ which is a UFD, so for some irreducible π, $\pi \mid \pi_1 \cdots \pi_n + 1$; but if $\pi = \pi_k$ for some k, then $\pi \mid 1$, hence π is constant and not irreducible; thus $\pi \neq \pi_k$ for all k. ∎

2.7.12 Exercise. Show that if K is algebraically closed, then K is infinite.

Here is an application of the fact that $\mathbb{R}[X, Y]$ is a UFD:-

2.7.13 Example. In $\mathbb{R}[X, Y]$, $f = X^2(X-1)^2 + Y^2$ is irreducible. Indeed, if f is reducible then we may write $f = \lambda gh$ for $\lambda \in \mathbb{R}$ and g and h of the form $g = g' + Y$ and $h = h' + Y$ for $g', h' \in \mathbb{R}[X]$. Now consider

$$X^2(X-1)^2 + Y^2 = f = \lambda(g' + Y)(h' + Y) = \lambda(g'h' + Y(g' + h') + Y^2); \tag{2.44}$$

matching coefficients we have that $g' = -h'$ and $\lambda g'h' = X^2(X-1)^2$, so $-\lambda g'^2 = X^2(X-1)^2$ and $-\lambda h'^2 = X^2(X-1)^2$; by uniqueness of factorisation, $\sqrt{-\lambda}g' = X(X-1)$ and $\sqrt{-\lambda}h' = X(X-1)$; thus $g' = h'$, contradiction.

2.8 Reducibility of affine space

Now we take a short break from commutative algebra in order to verify that one very important set is irreducible.

2.8.1 Theorem. *Let K be a field, and let $f \in K[X_1, ..., X_n]$ be a polynomial such that* $\mathrm{eval}_P f = 0$ *for all $P \in K^n$. Then if K is infinite, f is the zero polynomial; if K is finite with q elements, then $f \in ((X_1^q - X_1), ..., (X_n^q - X_n))$.*

Proof. In the infinite case we will proceed by induction on n, the number of variables.

Suppose $f \in K[X]$. If $f(P) = 0$ for every $P \in K$, then $(X - P)$ divides f for each P; since $k[X]$ is a unique factorisation domain and K is infinite, the only possibility is that $f = 0$.

Now suppose that for all $m < n$, if $f \in K[X_1, ..., X_m]$ is nonzero then there exists $P \in K^m$ such that $f(P) \neq 0$. Let $f \in K[X_1, ..., X_n] = (K[X_1, ..., X_{n-1}])[X_n]$; set $L = K[X_1, ..., X_{n-1}]$ so $f \in L[x_n]$. Suppose f is such that $f(P) = 0$ for all $P \in K^n$. Then, treating f as a single-variable polynomial over L, $f(P) = 0$ for all $P \in K$. Thus $(X_n - P) \mid f(x_n)$ in $L[x_n]$ for all $P \in K$; thus since each $(X_n - P)$ is an irreducible polynomial, f is divisible by infinitely many distinct irreducibles in $L[Y]$; but this means that $f = 0$ since $L[Y]$ is a UFD.

The finite case proceeds in much the same way as the infinite case:

Suppose $f \in K[X]$. If $f(P) = 0$ for every $P \in K$, then $(X - P)$ divides f for each P; since $k[X]$ is a unique factorisation domain, we must have that $(X^q - X) \mid f$.

Now suppose that for all $m < n$, if $f \in K[X_1, ..., X_m]$ is nonzero then there exists $P \in K^m$ such that $f(P) \neq 0$. Let $f \in K[X_1, ..., X_n] = (K[X_1, ..., X_{n-1}])[X_n]$; set $L = K[X_1, ..., X_{n-1}]$ so $f \in L[x_n]$. Suppose f is such that $f(P) = 0$ for all $P \in K^n$. Then, treating f as a single-variable polynomial over L, $f(P) = 0$ for all $P \in K$. Thus $(X_n - P) \mid f(X_n)$ in $L[X_n]$ for all $P \in K$; thus since each $(X_n - P)$ is an irreducible polynomial, f is divisible by $(X_n^q - X)$ in $L[Y]$. ∎

2.8.2 Remark. Note that we have actually proved something a little stronger: if $f \in K[X_1, ..., X_n]$, and there exist infinitely many points $P \in K^n$ such that $f(P) = 0$, then $f = 0$.

2.8.3 Example. $\mathbb{A}^n_K$ is reducible when K is finite, since it is a union of finitely many singleton sets (c.f. proposition 2.5.17).

When K is infinite, $\mathbb{A}^n_K$ is irreducible. Indeed, suppose $\mathbb{A}^n_k = \mathcal{V}(I) \cup \mathcal{V}(J)$ for vanishing ideals I and J; then $(0) = \mathcal{I}(\mathbb{A}^n_k) = \mathcal{I}(\mathcal{V}(I) \cup \mathcal{V}(J)) = IJ$. Suppose $f \in I$ is non-zero. Then for all $j \in J$ we have that for all $P \in \mathbb{A}^n_K$, $(fj)(P) = 0$. In particular, since f has only finitely many zero points, and since K is an integral domain, j has infinitely many zeroes and is the zero polynomial; so $J = (0)$. Similarly, if $J \neq (0)$ then $I = (0)$. Hence $\mathbb{A}^n_k = \mathcal{V}(I) \cup \mathcal{V}(J)$ implies that one of $\mathcal{V}(I)$ or $\mathcal{V}(J)$ is $\mathcal{V}(0) = \mathbb{A}^n_K$; so $\mathbb{A}^n_K$ is irreducible.

2.9 Radical ideals

We would like to complete our discussion of vanishing ideals of sets by characterising them.

2.9.1 Definition. An ideal I of a ring R is called **radical** if it satisfies the following property:

$$\text{for all } x \in R, \text{ if } x^n \in I \text{ for some } n \in \mathbb{N}, \text{ then } x \in I. \tag{RAD}$$

The **radical of an ideal** I, denoted $\sqrt{I}$, is the intersection of the radical ideals containing I.

2.9.2 Proposition. *Let I be an ideal of R; then $\sqrt{I}$ is indeed an ideal of R, and*

$$\sqrt{I} = \{x \in R : \exists_{n\in\mathbb{N}} x^n \in I\}. \tag{2.45}$$

Proof. Let $J = \{x \in R : \exists_{n\in\mathbb{N}} x^n \in I\}$. We will first show that J is a radical ideal containing I. Suppose $a, b \in J$. Then there exist $m, n \in \mathbb{N}$ such that $a^m \in I$ and $b^n \in I$; we compute

$$\begin{aligned}(a+b)^{m+n} &= \sum_{i=0}^{m+n} \binom{m+n}{i} a^{m+n-i} + b^i \\ &= \sum_{i=0}^{n} \binom{m+n}{i} a^{m+n-i} b^i + \sum_{i=n+1}^{m+n} \binom{m+n}{i} a^{m+n-i} b^i \in I\end{aligned} \tag{2.46}$$

so $(a+b) \in \sqrt{I}$; similarly, let $r \in R$ and $a \in J$. Then $j^n \in I$ for some n, thus $(rj)^n \in I$ and $rj \in J$. Hence J is an ideal.

Clearly $I \subseteq J$ since $x \in I \implies x^1 \in I \implies x \in J$; and J is radical since if $x^n \in J$ then $(x^n)^m \in I$ for some $m \in \mathbb{N}$ and hence $x^{nm} \in I$; so $x \in J$.

We need only show that J is the intersection of the radical ideals containing I. Suppose there is another radical ideal K containing I. Let $x \in J$. Then $x^m \in I$ for some $m \in \mathbb{N}$; but then $x^m \in K$; so $x \in K$ by radicalness, and $J \subseteq K$. ∎

2.9.3 Proposition. *If $X \subseteq \mathbb{A}^n$, then $\mathcal{I}(X)$ is a radical ideal.*

Proof. Suppose $f^n \in \mathcal{I}(X)$. Then for all $P \in X$, $f^n(P) = 0$. In particular, since K is an integral domain, $f(P) = 0$; so $f \in \mathcal{I}(X)$. ∎

The main result of this chapter, Hilbert's famous Nullstellensatz (theorem 2.11.1), is the converse: vanishing ideals are precisely the radical ideals of $K[X_1, ..., X_n]$. The next few theorems belong strictly to commutative algebra; a nice book for light reading is [AM69].

2.9.4 Exercise. Let I and J be ideals of R. Then:-

1. $\sqrt{\sqrt{I}} = \sqrt{I}$

2. $\sqrt{IJ} = \sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}$

3. $\sqrt{I} = R$ iff $I = R$

4. $\sqrt{I + J} = \sqrt{\sqrt{I} + \sqrt{J}}$

2.9.5 Exercise. Check that in $\mathbb{Z}$, $\sqrt{4\mathbb{Z}} = 2\mathbb{Z}$ and $\sqrt{12\mathbb{Z}} = 6\mathbb{Z}$. What is $\sqrt{n\mathbb{Z}}$ in general?

2.9.6 Definition. Let R be a ring.

1. An element $x \in R$ is called **nilpotent** if $x^n = 0$ for some $n \in \mathbb{N}$.

2. The **nilradical** of R, denoted $\mathfrak{N}(R)$, is the set of all nilpotent elements of R.

2.9.7 Lemma. *If R is a ring, then $\mathfrak{N}(R)$ is an ideal of R and $R/\mathfrak{N}(R)$ has no nilpotent elements except 0.*

Proof. Clearly if $a \in \mathfrak{N}(R)$ and $r \in R$ then $(ra)^n = r^n a^n = 0$ for some n and thus $an \in \mathfrak{N}(R)$. Suppose $a, b \in \mathfrak{N}(R)$ such that $a^m = 0$ and $b^n = 0$. Then

$$\begin{aligned}(a+b)^{m+n} &= \sum_{i=0}^{m+n} \binom{m+n}{i} a^{m+n-i} + b^i \\ &= \sum_{i=0}^{n} \binom{m+n}{i} a^{m+n-i} b^i + \sum_{i=n+1}^{m+n} \binom{m+n}{i} a^{m+n-i} b^i = 0\end{aligned} \tag{2.47}$$

so $(a+b)^{m+n} \in \mathfrak{N}(R)$.

Let $[r] \in R/\mathfrak{N}(R)$ be such that $[0] = [r]^n$ for some n. Then $[0] = [r^n]$, so $r^n \in \mathfrak{N}(R)$; hence there exists $m \in \mathbb{N}$ such that $(r^n)^m = 0$; i.e. $r^{n+m} = 0$ and so $r \in \mathfrak{N}(R)$: $[r] = [0]$. ∎

2.9.8 Lemma. *If R is a Noetherian ring, then $\mathfrak{N}(R)$ is the intersection of all the prime ideals of R.*

Proof. Let $r \in \mathfrak{N}(R)$; then $r^n = 0$ for some n. Let P be a prime ideal of R; then $0 \in P$, so $r^n \in P$ and thus $r \in P$ by primality. Hence r lies in every prime ideal of R.

Conversely, suppose r is not nilpotent. We wish to find some prime ideal of R that does not contain r. Let Σ be the set of ideals $I \subseteq R$ such that $n > 0 \implies r^n \notin I$. Then $(0) \in \Sigma$, so Σ contains a maximal element under inclusion, P. Let $x, y \in R$ be such that $x \notin P$ and $y \notin P$. Then the both the sum ideals $(x) + P$ and $(y) + P$ strictly contain P, and hence do not lie in Σ; so $r^n \in (x) + P$ and $r^m \in (y) + P$ for some $n, m > 0$. Thus $r^{n+m} \in (xy) + P$, so $P + (xy) \notin \Sigma$ and $xy \notin P$. Hence P is prime, and does not contain r. ∎

2.9.9 Proposition. *Let R be a Noetherian ring. If I is an ideal of R, then $\sqrt{I}$ is the intersection of the prime ideals containing I.*

Proof. Form the ring R/I. Then by the previous lemma, $\mathfrak{N}(R/I)$ is the intersection of the prime ideals of R/I. By the prime correspondence theorem, proposition 2.1.16, there is an inclusion-preserving bijection between the ideals of R/I and the ideals of R containing I given by $J \leftrightarrow J/I$ which preserves primality. Passing back through the correspondence we must have that $\mathfrak{N}(R/I)$ corresponds to the ideal J which is the intersection of the prime ideals of R containing I.

But $x \in J$ iff $[x] \in \mathfrak{N}(R/I)$ iff $\exists_{n \in N}([x]^n = [0])$ iff $\exists_{n \in N}([x^n] \in R/I)$ iff $\exists_{n \in N}(x^n \in I)$; thus $J = \sqrt{I}$. ∎

2.9.10 Corollary. *If $I \subseteq R$ is prime, then I is radical.* ∎

2.9.11 Exercise. Here are some fun problems, from [AM69, chapter 1]. Let R be an *arbitrary* ring; .

1. The nilradical of R is the intersection of all the prime ideals of R. (Remark: you will need Zorn's lemma.)
2. If I is an ideal of I, $\sqrt{I}$ is the intersection of the prime ideals containing I. (Hint: try to use 1.)
3. Define $\mathfrak{J}$ (the **Jacobson radical** of R) to be the intersection of all the maximal ideals of R. Then $x \in \mathfrak{J} \iff 1 - xy$ is a unit in R for all $y \in R$.
4. In a polynomial ring $R[X]$, the Jacobson radical is equal to the nilradical.

2.10 Ring extensions

In order to prove the Nullstellensatz (theorem 2.11.1), we will need to discuss rradicalings which are generated by adding elements to other rings (similarly to how we produce the polynomial rings by adjoining new symbols).

2.10.1 Definition. Let R be a ring; an **R-algebra** is a ring homomorphism $\rho : R \to S$. We will normally talk of an 'R-algebra S', forgetting about the homomorphism.

More explicitly, S is an R-algebra if it is equipped with a 'multiplication' $\cdot : R \times S \to S$ (which is 'secretly' the homomorphism — we write $r \cdot s$ for $\rho(r)s$) that satisfies:-

A1. $(r_1 + r_2) \cdot s = r_1 \cdot s + r_2 \cdot s$ (i.e. $\rho(r_1 + r_2)s = \rho(r_1)s + \rho(r_2)s$);

A2. $r \cdot (s_1 + s_2) = r \cdot s_1 + r \cdot s_2$ (i.e. $\rho(r)(s_1 + s_2) = \rho(r)s_1 + \rho(r)s_2$);

A3. $(r_1 r_2) \cdot s = r_1 \cdot (r_2 \cdot s)$ (i.e. $\rho(r_1 r_2)s = \rho(r_1)[\rho(r_2)s]$);

A4. $1 \cdot s = s$ (i.e. $\rho(1)s = 1s = s$).

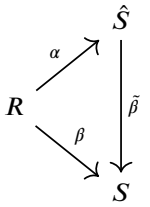
2.10.2 Example.

1. Let $\beta : R \to S$ be an injective ring homomorphism. Then β makes S into an R-algebra, called a **ring extension** of R. If S is an extension of R then we often write $S : R$.
   (a) $\mathbb{Q}$ is an extension of $\mathbb{Z}$.
   (b) More generally, Frac R is an extension of R.
   (c) $R[X_1, ..., X_n]$ is an extension of R.
   (d) If K is a field, then every K-algebra is an extension (since all field homomorphisms are injective). In fact, this K-algebra is just a vector space over K (though it has some additional structure: instead of a field action on an abelian group we have a field action on a ring).
2. If I is an ideal of R, then R/I is an R-algebra when we equip it with the standard projection $\pi : R \to R/I$.
3. The notions of $\mathbb{Z}$-algebra and ring are identical. (If R is any ring, there exists a unique ring homomorphism $\varphi : \mathbb{Z} \to R$; we say that $\mathbb{Z}$ is 'initial in the category Ring' [Alu09, §III.2].)
4. The set $\mathrm{Mat}_{n\times n}(R)$ of $n \times n$ matrices with entries in R is an R-algebra.

2.10.3 Exercise. Fix a ring R.

1. Define the notion of an R-algebra homomorphism.

2. Does there exist an R-algebra $\alpha : R \to \hat{S}$ such that, for all R-algebras $\beta : R \to S$, there is a unique homomorphism $\tilde{\beta}$ of R-algebras so that the following diagram commutes?



(2.48)

We shall be primarily interested in extensions of rings which come from adding only a small number of elements. We will make our definitions for more R-algebras in general, though a 'proper' definition may be found in [Alu09, §III.6.5].

2.10.4 Definition. A ring S is **finitely generated as an R-algebra** (or **of finite type**) if there is a surjective homomorphism (of R-algebras) $R[X_1, ..., X_n] \twoheadrightarrow S$ for some $n \in \mathbb{N}$.

The ring S is **linearly generated as an R-algebra** (or **finite over R**, though we will try to avoid this terminology) if it is an R-algebra and there exist $s_1, ..., s_n \in S$ such that

$$S = \{r_1 s_1 + \cdots + r_n s_n : r_1, ..., r_n \in R\}; \tag{2.49}$$

i.e. if S is the set of R-linear combinations of finitely many elements of S.

2.10.5 Remark. By proposition 2.2.5, a ring homomorphism $R[X_1, ..., X_n] \to S$ is the evaluation map at some point $\hat{s} \in S^n$. In particular, if $\phi : R[X_1, ..., X_n] \to S$ is a surjective R-algebra homomorphism, then it is a surjective ring homomorphism, so $S = \operatorname{im} \operatorname{eval}_{\psi,\hat{s}}$ for some $\hat{s} \in S$ and where ψ is the restriction of ϕ to R.

Consider the ring homomorphism which actually forms the R-algebra S, $\alpha : R \to S$. I claim that $\psi = \alpha$. But note that $\psi : R \to S$ is an algebra homomorphism; so in particular, multiplication by the ring factors through: $\psi(r1) = \alpha(r)\psi(1) = \alpha(r)$.[3]

What we have showed is that if S is finitely generated over R, then S is exactly (and in a very natural way) the set

$$\left\{ \operatorname{eval}_{\alpha,(s_1,...,s_n)} f : f \in R[X_1, ..., X_n] \right\} = \left\{ \sum \alpha(r_i) s_1^{i_1} s_2^{i_2} \cdots s_n^{i_n} : \forall_i, r_i \in R \right\} \tag{2.50}$$

for some fixed $s_1, ..., s_n \in S$. The converse is just the definition of a linear generated algebra: if the equality holds, then evaluation provides the needed surjection.

2.10.6 Definition. If $S : R$ is a finitely generated extension, and if the extension homomorphism is $\operatorname{eval}_{(s_1,...,s_n)}$, we will call S the **ring extension over R generated by** $s_1, ..., s_n$ (it is clear that changing the order of the s_i produces an isomorphic extension), and we shall write $S = R[s_1, ..., s_n]$.

If $L : K$ is an ring extension such that L and K are fields, then the **field extension over K generated by** $l_1, ..., l_n$ for $l_i \in L$ is defined to be

$$L(l_1, ..., l_n) := \operatorname{Frac} K[l_1, ..., l_n] \tag{2.51}$$

which is clearly an intermediate field between L and K.

It is not true that if $L : K$ is finitely generated as a field extension then $L : K$ is finitely generated as a ring extension. Indeed, $K(X_1, ..., X_m)$ is finitely generated as a field extension. However:

[3] I additionally claim a new record: we are 47 pages into some algebra notes, and this is the first 'proof by multiplying by 1'!

2.10.7 Lemma. *If K is a field, $K(X_1, ..., X_m)$ is not a finitely generated K-algebra.*

Proof. Let $A = K[f_1, ..., f_n] \subseteq K(X_1, ..., X_m)$ be a finitely generated K-algebra. Then we may write each f_i in the form p_i/q_i for $p_i, q_i \in K[X_1, ..., X_m]$; further, by replacing f_i with $p_i(\prod_{j\neq i} q_j)/(\prod q_j)$ we may arrange for all the q_i to be equal. Hence $A = K[p_1/q, ..., p_n/q]$.

Now, $K[X_1, ..., X_m]$ has infinitely many irreducible elements by corollary 2.7.11. Since q has only finitely many irreducible divisors since $K[X_1, ..., X_m]$ is a UFD, we may pick some element irreducible $\pi \in K(X_1, ..., X_m)$ such that $\pi \nmid q$. But every combination over K of the p_i/q will be of the form

$$\sum c_j \frac{p_1^{j_1} \cdots p_n^{j_n}}{q^{(j_1+\cdots+j_n)}}, \tag{2.52}$$

and thus if $1/\pi$ is in A we may write

$$\frac{1}{\pi} = \sum c_j \frac{p_1^{j_1} \cdots p_n^{j_n}}{q^{(j_1+\cdots+j_n)}} \implies q^{(j_1+\cdots+j_n)} = \pi \sum c_j p_1^{j_1} \cdots p_n^{j_n} \tag{2.53}$$

so $\pi \mid q^{(j_1+\cdots+j_n)}$ and (since irreducibles are prime in a UFD) $\pi \mid q$, contradiction.

In particular, the non-zero element $\pi \in K(X_1, ..., X_m)$ does not have an inverse in A; so A is not the whole of $K(X_1, ..., X_m)$. ∎

However, we do have some relations within towers of ring extensions.

2.10.8 Proposition. *If $C : B$ and $B : A$ are finitely generated, then $C : A$ is finitely generated. Further, if $C : B$ is linearly generated and $B : A$ is linearly generated, then $C : A$ is linearly generated.*

Proof. Let $C = B[c_1, ..., c_m]$ and $B = A[b_1, ..., a_n]$. Then if $\phi = \mathrm{eval}_{(c_1,...,c_m)} : B[Y_1, ..., Y_m] \to C$ we have $C \simeq \mathrm{im}\,\phi$; and if $\psi = \mathrm{eval}_{(b_1,...,b_n)} : A[X_1, ..., X_n] \to B$ then $B \simeq \mathrm{im}\,\psi$; thus we have a chain:

$$A[X_1, ..., X_n, Y_1, ..., Y_m] \xrightarrow{\tilde{\psi}} B[Y_1, ..., Y_m] \xrightarrow{\phi} C \tag{2.54}$$

The composition is surjective, so we have a surjective homomorphism $A[X_1, ..., X_n, Y_1, ..., Y_m] \to C$; hence C is finitely generated over A.

If $B : A$ is linearly generated, there exist $a_1, ..., a_n \in A$ such that every $b \in B$ is a B-linear combination of the a_i. If $C : B$ is linearly generated, there exist $b_1, ..., b_n \in A$ such that every $c \in C$ is a C-linear combination of the b_i. In particular, the b_i are B-linear combinations (and thus A-linear combinations) of the a_i; thus all $c \in C$ are A-linear combinations of the a_i. ∎

2.10.9 Proposition. *If $S : R$ is a linearly generated extension, then $S : R$ is finitely generated.*

Proof. Let $\iota : R \to S$ be the embedding of R in S, and suppose S is the set of R-linear combinations of $s_1, ..., s_n$. The ring homomorphism $\phi := \mathrm{eval}_{\iota,(s_1,...,s_n)} : R[X_1, ..., X_n] \to S$ is surjective: note that $s \in S$ implies the existence of $r_1, ..., r_n$ such that $s = r_1 s_1 + \cdots + r_n s_n = \phi(r_1 X_1 + \cdots + r_n X_n)$. Since S and R are algebras by virtue of being rings, an R-algebra homomorphism $R \to S$ is in particular a ring homomorphism. ∎

The converse is not true, and there is a simple example.

2.10.10 Example. Let R be a ring; $R[X]$ is finitely generated but not linearly generated. Indeed, it is obvious that $R[X]$ is finitely generated by X: the surjective homomorphism $R[X] \to R[X]$ that we need is just the identity!

It is also obvious that $R[X]$ is not linearly generated over R: let $f_1, ..., f_n$ be any finite subset of $R[X]$, and let g be a polynomial of degree $\partial g > \max\{\partial f_i : i \in [n]\}$. Then g cannot be written as a linear combination of the f_i with coefficients in R.

In a ring extension $S : R$, it is possible for the elements to be related in the larger ring to elements in the smaller ring.

2.10.11 Example.

1. Consider $\mathbb{R} \supseteq \mathbb{Q}[\sqrt{2}] : \mathbb{Q}$; then the adjoined element $\sqrt{2}$ satisfies a relation in the smaller ring $\mathbb{Q}$, namely that it is a zero of the polynomial $X^2 - 2 \in \mathbb{Q}[X]$.
2. Now consider $\mathbb{Z}[1/2] : \mathbb{Z}$; this is the ring of fractions with denominators that are powers of two. Then $1/2$ satisfies a relation in $\mathbb{Z}$, being a zero of $2X = 1$. Note that the polynomial is no longer monic.
3. On the other hand, consider $\mathbb{Q}[\pi] : \mathbb{Q}$. Then π is not a zero of any polynomial in $\mathbb{Q}[X]$. The number e satisfies the same property. A proof of the latter was first given by Hermite in 1873; the former was proved by Lindemann in 1882. Fairly simple proofs of the two results may be found in [Ste15, chapter 24].

2.10.12 Definition. Let $S : R$ be a ring extension. If $s \in S$, then s is called:

1. **algebraic** over R if there exists some nonzero $f \in R[X]$ such that $f(s) = 0$;
2. **integral** over R if there exists some *monic* nonzero $f \in R[X]$ such that $f(s) = 0$;
3. **transcendental** over R if it is not algebraic.

The entire extension is called algebraic over R if every element $s \in S$ is algebraic over R; integral if every element is integral; and transcendental if there exists an element which is not algebraic.

For convenience, we shall define the ring extension $0 : 0$ (i.e. the zero ring over itself) to be integral and algebraic.

We have seen that if $K : L$ is a finitely generated field extension then it is not necessarily finitely generated as a ring extension (lemma 2.10.7). However, we can say something useful: if it *is* finitely generated as a ring extension, then it is linearly generated as a K-algebra. (Recall that for general rings, linearly generated implies finitely generated but not conversely — example 2.10.10.) The next large theorem will enable us to show this.

2.10.13 Theorem (Zariski's lemma)**.** *Let $L : K$ be a field extension. If L is finitely generated as a K-algebra then L is linearly generated and algebraic over K.*

The remainder of this section will be devoted to proving Zariski's lemma. We shall follow the treatment given in [AM69, chapter 7], and will only prove precisely the lemmata we need in order to do what we want to do. However, the subject of algebraic and transcendental ring extensions is quite elegant in its own right; see [Lan71, chapters VII, IX, and X].

Technical results for Zariski's lemma

2.10.14 Lemma. *If $K(\alpha) : K$ is an extension of field where α is transcendental, then $K[\alpha] \simeq K[X]$ and $K(\alpha) \simeq K(X)$.*

Proof. The evaluation homomorphism $\phi = \mathrm{eval}_\alpha : K[X] \to K[\alpha]$ is surjective and so $K[\alpha] \simeq \frac{K[X]}{\ker \phi}$. Since α is transcendental over K, $\ker \phi = (0)$ and so $K[\alpha] \simeq K[X]$. But if two rings are isomorphic, then their fields of fractions are isomorphic; so we are done. ∎

2.10.15 Lemma. *Let $S : R$ be a ring extension.*

1. *If $s \in S$ is integral over R, then $R[s] : R$ is linearly generated.*
2. *If $S : R$ is finitely generated and integral then it is linearly generated.*
3. *If $S : R$ is linearly generated then it is integral.*

Proof. (See [Lan71, §IX.1].)

1. Let g be a monic nonzero polynomial with degree at least 1 in $R[X]$ such that $g(s) = 0$, and let $f \in R[X]$ be arbitrary. By the strong division algorithm (theorem 2.3.1), we may write $f = gq + r$ for polynomials $q, r \in R[X]$ and $\partial r < \partial g$; and $f(s) = r(s)$.

   In particular, if t is any element of $R[s]$ then $t = f(s)$ for some $f \in R[X]$ (remark 2.10.5) and so $t = r(s)$. Thus t is an R-linear combination of the elements $1, s, ..., s^{\partial g - 1}$, and so the latter ∂g elements generate $R[s]$ linearly.

2. Let $S = R[s_1, ..., s_n]$. By proposition 2.10.8, we may use induction on the chain $R : R[s_1] : \cdots : R[s_1, ..., s_n]$ since at each step the R-algebra $R[s_1, ..., s_i]$ is finitely generated and integral.

   So suppose $T' : T$ is integral and generated by a single element, s: so $T' = T[s]$ and s is integral over T; thus $T' : T$ is linearly generated by part 1.

3. Suppose S is linearly generated by $s_1, ..., s_n$, and let $s \in S$. For all i, write ss_i as a linear combination of the s_i:

$$\begin{aligned} ss_1 &= \alpha_{1,1}s_1 + \cdots + \alpha_{1,n}s_n \\ &\vdots \\ ss_n &= \alpha_{n,1}s_1 + \cdots + \alpha_{n,n}s_n. \end{aligned} \tag{2.55}$$

   Moving the right hand side to the left, this set of equations may be rewritten as the matrix equation

$$\begin{bmatrix} s-\alpha_{1,1} & & & & & \\ & s-\alpha_{2,2} & & -\alpha_{i,j} & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & -\alpha_{i,j} & & & \ddots & \\ & & & & & s-\alpha_{n,n} \end{bmatrix} \begin{bmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ s_n \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \tag{2.56}$$

   Recall from elementary linear algebra that for every matrix $A \in \mathrm{Mat}_{n\times n}(K)$ there exists a matrix $\tilde{A} \in \mathrm{Mat}_{n\times n}(K)$ (called the **adjoint** of A) such that

$$\tilde{A}A = (\det A)I_n; \tag{2.57}$$

   if $A = (a_{i,j})$ then the adjoint is given by the matrix $(\tilde{a}_{i,j})$ such that

$$\tilde{a_{i,j}} = (-1)^{i+j} \det(A_{ji}) \tag{2.58}$$

   (where A_{ji} is the matrix formed by deleting the jth row and ith column of A).

   If the determinant is defined by the Laplace expansion then one can show that determinants exist and are well-defined (i.e. are independent of the row or column that the expansion is taken with respect to) over an arbitrary ring, and that the adjoint of a matrix as defined above still satisfies equation 2.57. (This is true because the proofs of the Laplace expansion and of the adjoint equation do not use properties special to integral domains or fields; this is checked explicitly in [Lan71, p. XIII.4].)

In particular, let A be the large matrix in equation 2.56 above; we may then write

$$\begin{bmatrix} (\det A)s_1 \\ \vdots \\ (\det A)s_n \end{bmatrix} = (\det A)I_n \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} = \tilde{A}A \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.59}$$

. Since $(\det A)s_i = 0$ for all the generators s_i, we must have $(\det A)t = 0$ for all $t \in S$. In particular, since $1 \in S$ we have $\det A = 0$.

By the Laplace expansion definition of $\det A$, $\det A$ is polynomial in the entries of A. In particular, the equation

$$\det(\delta_{i,j}X - \alpha_{i,j}) \tag{2.60}$$

is a polynomial in $R[X]$ that has a zero at $s \in S$; further, one can check explicitly that the highest power of X will be n, and that the coefficient of X^n is 1. Thus the polynomial 2.60 is monic and s is integral over R as desired.

∎

2.10.16 Lemma. *Let $S : R$ be a ring extension. Then the set of elements of S which are integral over R is a subring of S (which clearly contains R).*

Proof. Let T be the set of elements of S which are integral over R. By the famed One Step Test®, we need to check that if $a, b \in T$ then $a + b, a - b, ab \in T$.

Clearly b is integral over $R[a]$ since $R[a]$ contains R. Then by part 1 of the lemma 2.10.15, the extension $R[a, b] : R[a]$ is linearly generated; since $R[a] : R$ is integral and finitely generated, by part 2 of that lemma it is linearly generated; hence $R[a, b] : R[a] : R$ is a chain of linearly generated extensions and so $R[a, b] : R$ is linearly generated (by proposition 2.10.8) and thus integral by part 3 of lemma 2.10.15. Since $a \pm b$ and ab lie in $R[a, b]$, we are done. ∎

2.10.17 Corollary. *Let $S : R$ be a finitely generated ring extension. If all the generators $s_1, ..., s_n$ of S over R are integral over R, then $S : R$ is integral.*

Proof. The set of elements of S which are integral over R is a subring of S by the lemma. The generators $s_1, ..., s_n$ lie in this subring; hence all the combinations of the generators lie in the subring; thus the subring must be the entire ring. ∎

2.10.18 Lemma. *Let $S : R$ be a finitely generated extension of rings. If R is a Noetherian ring, then so is S. In particular, every finitely generated algebra over a field is a Noetherian ring.*

Proof. By remark 2.10.5, there is a surjective homomorphism $\phi : R[X_1, ..., X_n] \to S$ for some $n \in \mathbb{N}$. In particular, $S \simeq \frac{R[X_1,...,X_n]}{\ker \phi}$. By the Hilbert basis theorem, more specifically example 2.6.23, $R[X_1, ..., X_n]$ is Noetherian; hence (by theorem 2.6.21) S is Noetherian. ∎

2.10.19 Lemma. *If $S : R$ is linearly generated and generated by n elements of S, there is an ideal $I \subseteq R^n$ such that $S \simeq R^n/I$.*

Proof. Let $x_1, ..., x_n$ be a generating set for S over R, and define $\phi : R^n \to S$ by $(r_1, ..., r_n) \mapsto r_1x_1 + \cdots + r_nx_n$. Then ϕ is an injective homomorphism and thus $S \simeq R^n/\ker \phi$. ∎

2.10.20 Lemma (Noetherian property for extensions)**.** *Let $S : R$ be a linearly generated extension of rings. If R is a Noetherian ring, then S has the following property:*

If Σ is a non-empty set of intermediate rings σ between S and R,
then Σ has a maximal element under inclusion. (NM)

Proof. We shall do induction on the number of generators n of $S : R$.

If $n = 1$, $S \simeq R/I$ (lemma 2.10.19). If $\sigma_1 \subset \cdots$ is a chain of intermediate rings $R \subset \sigma_1 \subset \cdots \subset S \simeq R/I$, for all i let $J_i = \ker \pi_i$ where $\pi_i : \sigma_i \to \sigma_i/I$ is the canonical projection; then J_i is a strictly ascending chain of ideals in S/I, which corresponds to a strictly ascending chain of ideals in S (by corollary 2.1.10, the correspondence theorem), which is absurd by the ascending chain condition of S. In particular, there is no strictly ascending chain of intermediate rings in Σ and thus Σ must have a maximal element under inclusion.

If $n > 1$, say $S = R[x_1, ..., x_n]$ for $x_1, ..., x_n \in S$, then $S \simeq (R[x_1, ..., x_{n-1}])[x_n]$; by lemma 2.10.18, $R[x_1, ..., x_{n-1}]$ is Noetherian; and clearly $(R[x_1, ..., x_{n-1}])[x_n]$ is linearly generated over $R[x_1, ..., x_{n-1}]$; hence any non-empty set of intermediate rings between $R[x_1, ..., x_{n-1}]$ and $(R[x_1, ..., x_{n-1}])[x_n]$ has a maximal element, by the case $n = 1$; and any non-empty set of intermediate rings between R and $R[x_1, ..., x_{n-1}]$ has a maximal element, by induction; and every set of intermediate rings between S and R is the union of two such sets, so it must have the property (NM), for suppose it did not — then there must be an infinite ascending inclusion chain somewhere, and by the pigeonhole principle at least one of the sets would have infinitely many elements from the chain. ∎

2.10.21 Lemma. *If $S : R$ is a ring extension such that S satisfies property (NM), then every intermediate ring S_1 such that $S \supseteq S_1 \supseteq R$ is linearly generated over R.*

Proof. Fix such an intermediate ring S_1, and let Σ be the set

$$\Sigma := \{N \subseteq S_1 : N \text{ is a linearly generated extension of } R\}. \tag{2.61}$$

Then $R \in \Sigma$, so it is non-empty. By property (NM), Σ has a maximal element S_0 under inclusion. If $S_0 \neq S_1$, pick $s \in S_1 \setminus S_0$; then the ring $S_0[s]$ is finitely generated over R and strictly contains S_0, contradiction. Thus $S_0 = S_1$, i.e. $S_1 \in \Sigma$ and $S_1 : R$ is linearly generated. ∎

The crux of our proof of Zariski's lemma will be the following theorem proved by Emil Artin and John Tate in 1951 (this is [AM69, proposition 7.8]):-

2.10.22 Theorem (Artin-Tate)**.** *Let $R \subseteq S \subseteq T$ be a chain of rings. Suppose R is Noetherian, that $T : R$ is finitely generated, and that $T : S$ is linearly generated. Then $S : R$ is finitely generated.*

Proof. Let $x_1, ..., x_m$ generate T as an extension of R; let $y_1, ..., y_n$ generate T as a set of S-linear combinations. Then there exist expressions of the form

$$x_i = \sum_j b_{i,j} y_j \tag{A}$$

$$y_i y_j = \sum_k b_{i,j,k} y_k \tag{B}$$

for $b_{i,j}$ and $b_{i,j,k}$ in S; both follow by finiteness of $T : S$.

Let S_0 be the extension of R generated by the set of all the $b_{i,j}$ and $b_{i,j,k}$ (note that S_0 is finitely generated). By lemma 2.10.18, S_0 is Noetherian; and we have $R \subseteq S_0 \subseteq S$.

By remark 2.10.5 we may view each element $t \in T$ as the evaluation of a polynomial in $R[X_1, ..., X_m]$ at the point $(x_1, ..., x_m) \in T^m$. Substituting equation (A) above allows us to write t as the evaluation of a polynomial in $S_0[X_1, ..., X_m]$ at the point $(y_1, ..., y_m) \in T^m$; and repeated application of (B) allows us to replace all products of the y_i with S_0-linear combinations, so t is the evaluation of a linear polynomial with coefficients in S_0 at $(y_1, ..., y_m)$. In particular, T is linearly generated over S_0.

By proposition 2.10.9, $T : S_0$ is finitely generated. Hence by lemma 2.10.20, since S_0 is Noetherian we have that T satisfies property (NM). By lemma 2.10.21 applied to the chain $T : S : S_0$ we have that S is linearly generated over S_0.

But this means that $S : S_0$ is finitely generated (proposition 2.10.9), and $S_0 : R$ is finitely generated, so the composition $S : R$ is finitely generated by proposition 2.10.8. ∎

Proof of Zariski's lemma, theorem 2.10.13.

Let $L : K$ be a field extension which is finitely generated as a K-algebra; assume $L = K[l_1, ..., l_n]$. Our goal is to show that $L : K$ is algebraic; by corollary 2.10.17, it suffices to show that the generators of L are algebraic over K.

Suppose that not all the generators of L are algebraic over K. Renumber them so that $l_1, ..., l_r$ $(1 \leq r)$ are transcendental and $l_{r+1}, ..., l_n$ are algebraic over K. In particular, $l_{r+1}, ..., l_n$ are algebraic over the larger field $M := K(l_1, ..., l_r)$. Hence L is a finitely generated algebraic extension of K and hence by part 2 of lemma 2.10.15 $L : M$ is a linearly generated extension.

Apply the Artin-Tate theorem (theorem 2.10.22) to the chain $K \subseteq M \subseteq L$ — since K is a field it is Noetherian, and we just proved that $L : M$ is linearly generated and $L : K$ is finitely generated. We may therefore conclude that M is finitely generated as a K-algebra, say $M = K[m_1, ..., m_s]$. But $M \simeq N(X_1, ..., X_n)$ by lemma 2.10.14; and this is a contradiction based on lemma 2.10.7.

We have shown that $L : K$ is algebraic, and it remains to prove the finiteness conclusion. For each l_i there is some $f_i \in K[X]$ such that $f_i(l_i) = 0$. Define the sets

$$S_i = \{l_i^d : d < \partial f_i\}, \text{ and } G = \{\prod_{i=0}^{n} s_i : s_i \in S_i\}. \tag{2.62}$$

I claim that every $x \in L$ is a K-linear combination of elements in G.

Indeed, we may write $x = \sum k_i l_1^{i_1} \cdots l_n^{i_n}$ for some elements k_i (since the l_is generate L). But for each term in this sum, we may write $l_k^{i_k}$ as some K-linear combination of powers of l_k with degree less than ∂f_k (since $f_k(l_k) = 0$ gives us a K-linear dependence between powers of l^k with degree ∂f_k and those of lower degree). Expanding out, we have written x as a K-linear combination of products of the l_i such that each power of l_k appearing is at most ∂f_k; but these products are precisely the members of G. ∎

2.11 Hilbert's Nullstellensatz

The previous section has finally furnished us with strong enough machinery to be able to prove the Nullstellensatz.

2.11.1 Theorem (Hilbert's Nullstellensatz)**.** *Let K be an algebraically closed field. If $I \subseteq K[X_1, ..., X_n]$ is an ideal, then*

$$\mathcal{I}(\mathcal{V}(I)) = \sqrt{I}. \tag{2.63}$$

Thus the correspondences $I \mapsto \mathcal{V}(I)$ and $V \mapsto \mathcal{I}(V)$ induce a bijection between the algebraic subsets of $\mathbb{A}^n_K$ and the radical ideals of $K[X_1, ..., X_n]$.

The intuition behind this theorem is very geometric: if a function vanishes on the whole zero set of some ideal (i.e. it lies in $\mathcal{I}(\mathcal{V}(I))$) then we can find some power of it in the ideal. Even less precisely, 'if a function vanishes on the whole set, then all its roots (in the nth root sense) vanish there too'.

Proof of Hilbert's Nullstellensatz, theorem 2.11.1.

2.11.2 Remark. Our proof will follow [AM69, corollary 7.10 and exercise 7.14]. Further details including five(!) proofs of the Nullstellensatz may also be found in [Eis95]; see pages 33 to 34.

The reader might also find the remarks immediately following the proof to be enlightening while reading the proof itself.

2.11.3 Lemma (Weak Nullstellensatz)**.** *Let K be a field, and let R be a finitely generated extension of K. Let M be a maximal ideal of R. Then the field R/M is an algebraic extension of K. In particular, if K is algebraically closed then $R/M \simeq K$.*

Proof. Suppose R has n generators over K, with surjective homomorphism $\phi : K[X_1, ..., X_n] \to R$. Then the canonical projection $\pi : R \to R/M$ produces a surjective homomorphism $\pi \circ \phi : K[X_1, ..., X_n] \to R/M$. Thus R/M is finitely generated over K. Further, the map $\pi \circ \phi$ is injective when restricted to K, since it is a homomorphism out of a field.

Based on these observations we may apply Zariski's lemma, theorem 2.10.13, with the field R/M playing the part of L, to conclude that R/M is algebraic over K. ∎

The following corollary wraps up our discussion of example 2.5.21:

2.11.4 Corollary (Characterisation of maximal ideals)**.** *Let K be an algebraically closed field. Then M is a maximal ideal of $K[X_1, ..., X_n]$ if and only if it is of the form $((X_1 - x_1), ..., (X_n - x_n))$ for some $(x_1, ..., x_n) \in K^n$.*

Proof. Every point of the given form is maximal even when the field is not algebraically closed — this was parts 3 and 4 of theorem 2.5.18.

Conversely, let M be a maximal ideal; then by the weak Nullstellensatz, $K[X_1, ...X_n]/M \simeq K$. Let $\phi : K[X_1, ..., X_n]/M \to K$ be the isomorphism, and set $P = (x_1, ..., x_n) = (\phi(X_1), ..., \phi(X_n))$. I claim that $\mathcal{I}(P) = M$; indeed, $f \in M \iff 0 = \bar{f} \iff 0 = \bar{f}(X_1 + M, ..., X_n + M) \iff 0 = f(P) \iff f \in \mathcal{I}(P)$ (where $\bar{f}$ is the residue of f in $K[X_1, ..., X_n]/M$). Since $\mathcal{I}(P) = ((X_1 - x_1), ..., (X_n - x_n))$ by theorem 2.5.18, we are done. ∎

We will now proceed to prove the Nullstellensatz proper.

We know from proposition 2.4.4 that $I \subseteq \mathcal{I}(\mathcal{V}(I))$, and we know that $\mathcal{I}(\mathcal{V}(I))$ is a radical ideal by proposition 2.9.3. Hence $\sqrt{I} \subseteq \mathcal{I}(\mathcal{V}(I))$.

Conversely, suppose $f \notin \sqrt{I}$. Recall that $\sqrt{I}$ is the intersection of the prime ideals containing I (this was proposition 2.9.9); hence there exists some prime ideal P containing I such that $f \notin P$. Let $\bar{f}$ be the (non-zero) image of f in the integral domain $R = K[X_1, ..., X_n]/P$, and consider the ring extension $R[1/\bar{f}]$; this extension is finitely generated over K since there is a surjective homomorphism $K[X_1, ..., X_{n+1}] \to R[1/\bar{f}]$ given by

$$K[X_1, ..., X_{n+1}] \xrightarrow{\text{eval}_{(X_1,...,X_n,1/f)}} K[X_1, ..., X_n][1/f] \xrightarrow{\text{canonical proj.}} (K[X_1, ..., X_n]/P)[1/\bar{f}]. \tag{2.64}$$

Let M be a maximal ideal of $R[1/\bar{f}]$ (we know one exists by proposition 2.6.20 since $R[1/\bar{f}]$ is Noetherian by a combination of theorem 2.6.22 and theorem 2.6.21). We may apply lemma 2.11.3 to conclude that $R[1/\bar{f}]/M \simeq K$.

Let $\phi : K[X_1, ..., X_{n+1}] \to K$ be the map we create as follows:

$$\begin{array}{c} K[X_1, ..., X_n] \\ \downarrow \\ K[X_1, ..., X_n]/I \\ \downarrow \\ K[X_1, ..., X_n]/P = R \\ \downarrow \\ R[1/\bar{f}] \\ \hookrightarrow \\ R[1/\bar{f}]/M \\ \xrightarrow{\sim} \\ K \end{array} \tag{2.65}$$

Now note that ϕ is a surjective map onto a field, and thus its kernel is a maximal ideal (apply the first homomorphism theorem). By corollary 2.11.4, $\ker \phi = ((X_1 - x_1), ..., (X_n - x_n))$ for some $(x_1, ..., x_n) \in K^n$; so $\phi(X_i) = x_i$ for each i, and by proposition 2.2.5, $\phi = \text{eval}_{(x_1,...,x_n)}$. In particular, for all $g \in I$, $\text{eval}_{(x_1,...,x_n)}\, g = \phi(g) = 0$ (by the first quotient map) and thus $(x_1, ..., x_n) \in \mathcal{V}(I)$.

On the other hand, $\bar{f}$ is a unit in $R[1/\bar{f}]$ and therefore remains a unit in K (exercise 2.1.15). In particular, $f(x) = \phi(\bar{f}) \neq 0$. Hence f fails to vanish at a point of $\mathcal{V}(I)$ and therefore does not lie in $\mathcal{I}(\mathcal{V}(I))$. This establishes the inclusion $\mathcal{I}(\mathcal{V}(I)) \subseteq \sqrt{I}$, and the Nullstellensatz is proved. ∎

Remarks and philosophy surrounding the Nullstellensatz' proof

In the proof, we used the following fact:

Proposition (2.9.9)**.** *Let R be a Noetherian ring. If I is an ideal of R, then $\sqrt{I}$ is the intersection of the prime ideals containing I.*

Now that we know that $\sqrt{I}$ is simply the vanishing ideal corresponding to I, this makes sense geometrically: each prime ideal containing I corresponds to an irreducible curve contained within $\mathcal{V}(I)$ (since the $\mathcal{I}$-$\mathcal{V}$ correspondence reverses inclusions), and so since the algebraic curve containing $\mathcal{V}(I)$ contains all these irreducibles (and only these irreducibles), the vanishing ideal corresponding to it must be included by the vanishing ideals of all the irreducible curves (and only included by these irreducibles).

The idea behind the proof is very simple: given a polynomial in $\sqrt{I}$, we construct a point which lies in K^n but at which the polynomial does not vanish. The way we do this is as follows:-

1. We find some irreducible curve (this is the prime ideal P) which is contained within I and which f does not vanish on. (If f vanished on all the prime ideals contined within I, it would vanish on the whole thing, since $\mathcal{V}(I)$ is a union of vanishing sets of prime ideals by the proposition 2.9.9 quoted above).

2. We consider the restriction of f to this prime ideal. Recall that the intuition behind dividing polynomial rings by ideals is to do with introducing relations: dividing $\mathbb{R}[X]$ by $(X^2 - 1)$ produces a new field in which $X^2 - 1$ is killed (set to zero); this introduction of relations is also efficient, in that

only as many relations as required are produced. For example, $\mathbb{R}[X]/(X^2+1)$ is exactly a field $\mathbb{R}$ in which some elements have been adjoined — namely, $\mathbb{C}$ where $i \leftrightarrow X$. Thus considering the residue of f in $K[X_1, ..., X_n]/P$ is equivalent intuitively to 'substituting' the relations defined by P into f and 'simplifying'. (For a more concrete discussion of this, see remark 1.2.1 and its surrounding example.) For convenience of reference we set $R := K[X_1, ..., X_n]/P$.

3. Since $f \notin P$, f is not identically killed by this process. Thus its residue $\bar{f}$ is not zero; in particular, there is an element $1/\bar{f}$ which lives in Frac R; and we may adjoin it to R with no problems.

4. Now we take a maximal ideal M of this new ring $R[1/\bar{f}]$. This maximal ideal corresponds to a point x of the original space, and thus quotienting out by it simply 'substitutes' (in the same relation-introduction sense) the point x into the functions of $R[1/\bar{f}]$.

5. Now if g was in the original ideal I, then g lies in P (since P contains I). In particular, when we killed the ideal P, the function g was sent to zero. Thus for the remainder of the quotienting process, g is zero. In particular, $\bar{g}(x) = 0$ in the final step (where $\bar{g}$ corresponds to the residue of g); but x is isomorphic to a point of K^n (by the application of the weak Nullstellensatz, lemma 2.11.3) and so passing back through the isomorphism allows us to conclude that $g(x) = 0$ in the original ring.

6. On the other hand, $\bar{f} \neq 0$ in the quotient ring $R[1/\bar{f}]$. When we quotient out again by a maximal ideal M, we cannot send $\bar{f}$ to zero (because quotienting by an ideal does not send units to non-units, exercise 2.1.15), and so (equivalently) the substitution of x does not send it to zero. Hence we have constructed a point $x \in K^n$ such that $f(x) \neq 0$ but $x \in \mathcal{V}(I)$; so $f \notin \mathcal{I}(\mathcal{V}(I))$.

Note also that the only point where algebraic closedness of K was used is the application of Zariski's lemma; more precisely, it is in using the result of the weak Nullstellensatz, lemma 2.11.3. If K is not algebraically closed, all that we produce is a point in the affine space corresponding to some algebraic extension of K at which the polynomial $f \notin \sqrt{I}$ does not vanish — the polynomial may still vanish at all the points of $\mathcal{V}(I)$ which lie in the affine space of the base field K.

Examples of the Nullstellensatz' proof

In this section we will actually work through the proof for several specific examples. The calculations made will be necessarily informal; we leave it to the bored reader to actually work through the details, as the idea is to display the intuition behind the calculations.

In one dimension, for example $\mathbb{A}^1_{\mathbb{C}}$, the Nullstellensatz does not say anything particularly interesting since all primes in $\mathbb{C}[X]$ are maximal. When one works through a particular example, half the proof collapses:

2.11.5 Example. Let $K = \mathbb{C}$, and $I = (X^2)$. It is clear that $\sqrt{I} = (X)$, and that if $f \in \sqrt{I}$ then f must vanish at least as often as anything which vanishes wherever I vanishes (i.e. $\sqrt{I} \subseteq \mathcal{I}(\mathcal{V}(I))$).

Suppose $f \notin \sqrt{I}$; for example, $f = X^2 - 1$. Then there exists a prime ideal P such that $P \not\ni f$ but $P \supseteq I$; e.g. (X). In $\mathbb{C}[X]/P$, we have the relation $X = 0$; f becomes $\bar{f} = 0^2 - 1 = -1$. Note also that in $\mathbb{C}[X]$, all primes are maximal and so $\mathbb{C}[X]/P \simeq \mathbb{C}$. In particular, -1 already has a unit and we may skip the modulo by a maximal ideal M in the proof.

In this case we have $\phi : \mathbb{C}[X] \to \mathbb{C}$ given by $\phi : \sum \alpha_i X^i \mapsto x_0$; hence $\phi = \mathrm{eval}_0$. Note that $\phi(X) = 0$.

It is clear that if $g \in I$ then $g(0) = \mathrm{eval}_0\, g = \phi(g) = 0$, since $X \mid g$ and so g has no constant term; hence $0 \in \mathcal{V}(I)$. On the other hand, $\phi(f) = -1$. In particular, there is a point of $\mathcal{V}(I)$ which f does not vanish at; so $f \notin \mathcal{I}(\mathcal{V}(I))$.

Let us consider now an example in two dimensions over $\mathbb{C}$; we will continue to work with 'simple' ideals.

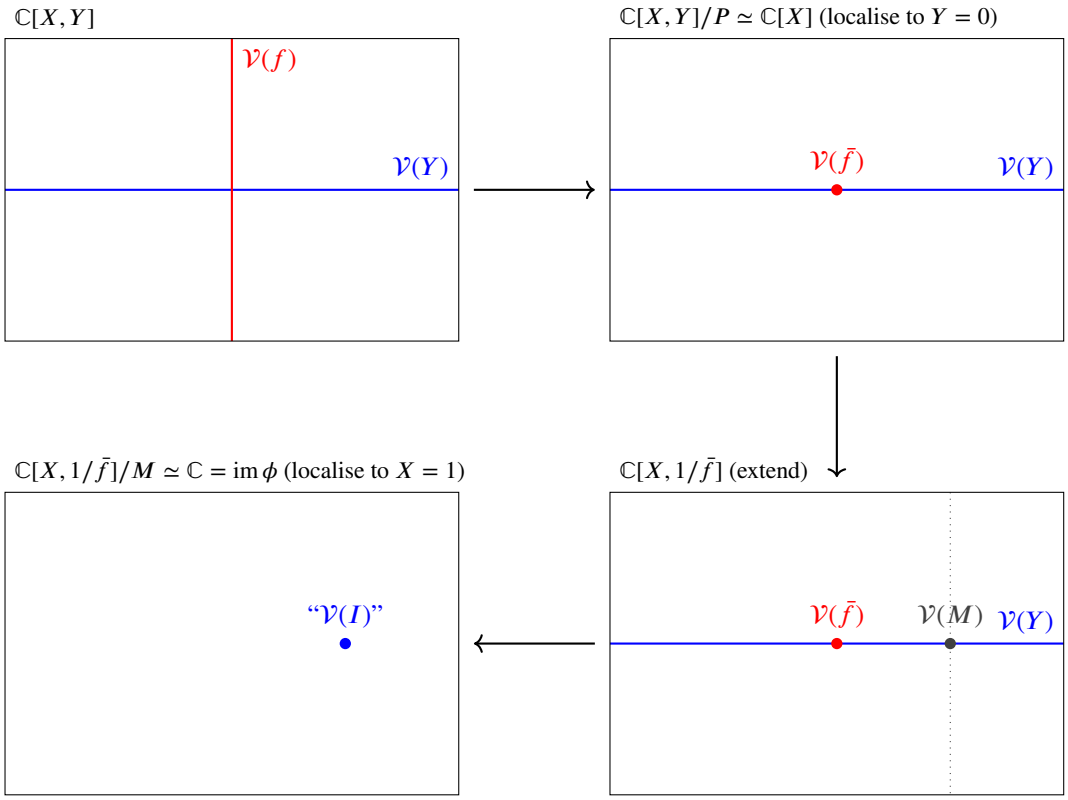


Figure 2.3: The sequence of killings made in example 2.11.6.

2.11.6 Example. Let $I = (Y) \subseteq \mathbb{C}[X,Y]$; so $\sqrt{I} = I$. Suppose $f \notin \sqrt{I}$; then, as always, we will attempt to construct a point in $\mathbb{C}^2$ which lies in $\mathcal{V}((Y))$ (i.e., lies on the y-axis) but at which f does not vanish.

There exists some prime ideal P containing I such that $f \notin P$; in this case, I itself is prime so we take $P = I$. In particular, $\mathbb{C}[X,Y]/P = \mathbb{C}[X,Y]/(Y) \simeq \mathbb{C}[X]$, and $\bar{f} = f$. Thus we pass to the extended quotient ring $\mathbb{C}[X,1/X]$. A maximal ideal of this ring is $(X-1)$; the final step is to quotient by this relation, forming $\mathbb{C}[X,1/X]/(X-1)$ (in effect, substituting $X=1$).

Note that we have already quotiented out by (Y), and so the point in $\mathbb{C}^2$ corresponding (via the isomorphism) to this final ideal M is actually the point $(0,1)$; and if $g \in I = (Y)$, then $g(0,1) = 0$. On the other hand, $f(0,1) = 1$.

The sequence of killings performed in this example is depicted in figure 2.3.

2.11.7 Exercise. Draw similar diagrams to example 2.11.6 for the case where $I = (X^2 + Y^2 - 1)$ and $f = X - Y$.

Show that $\mathbb{C}[X,Y]/(X^2 - Y) \simeq \mathbb{C}[X]$ and do the same thing for $I = (X^2 - Y)$ and $f = Y$.

Let us now study the case where K is *not* algebraically closed; as mentioned in the remarks above, the problem is that the point constructed must only lie in an algebraic extension of K, not necessarily in K itself.

2.11.8 Example ("quick maffs")**.** Let $K = \mathbb{R}$; we will work in $\mathbb{A}^2_{\mathbb{R}}$ for this example. Let $I = (X^2(X-1)^2 + Y^2)$; then I is prime (since $X^2(X-1)^2 + Y^2$ is irreducible, by example 2.7.13). By corollary 2.9.10,

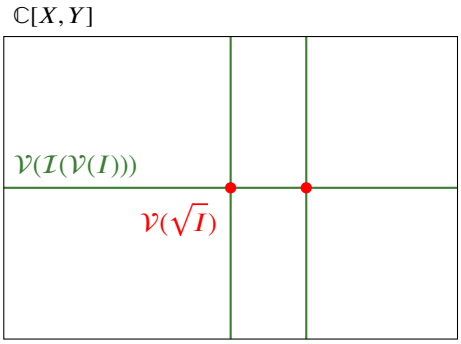


Figure 2.4: The relationship between $\sqrt{I}$ and $\mathcal{I}(\mathcal{V}(I))$ in example 2.11.8.

$I = \sqrt{I}$. On the other hand, let us compute $\mathcal{V}(I)$:

$$X^2(X-1)^2 + Y^2 = 0 \implies Y = 0 \text{ and } X(X-1) = 0 \tag{2.66}$$

so $\mathcal{V}(I) = \{(0,0),(1,0)\}$ and

$$\mathcal{I}(\mathcal{V}(I)) = \mathcal{I}(\{(0,0)\} \cup \{(1,0)\}) = (X,Y) \cap ((X-1),Y) = (X(X-1),Y) \tag{2.67}$$

where the second equality comes from theorem 2.5.18, and the third equality is an exercise. Note that I is clearly contained within $\mathcal{I}(\mathcal{V}(I))$ (this is guaranteed by proposition 2.4.4 anyway, but it is nice to see it concretely). The situation is pictured in figure 2.4.

We shall now see where the proof falls down, by picking some $f \notin \sqrt{I}$ such that $f \in \mathcal{I}(\mathcal{V}(I))$ and seeing what happens. For example, we may take $f = X$. We must pick a prime ideal $P \subseteq \mathbb{R}[X,Y]$ such that $P \not\ni f$ but $P \supseteq I$; since I is prime we may take $I = P$. Thus our localised[4] ring is $R = \mathbb{R}[X,Y]/(X^2(X-1)^2+Y^2)$. We need to pick a maximal ideal in $R[1/\bar{f}] = R[1/X]$; let $M = (X-1)^2 - Y^2$. It is not immediately clear that this is maximal, but consider

$$\frac{\frac{\mathbb{R}[X,Y]}{(X^2(X-1)^2+Y^2)}[1/X]}{(X-1)^2 - Y^2} \simeq \frac{\mathbb{R}[X,1/X,Y]}{(X^2Y^2+Y^2)} = \frac{\mathbb{R}[X,1/X,Y]}{(Y^2(X^2+1))} \simeq \frac{\mathbb{R}[X,1/X]}{(X^2+1)} \simeq \mathbb{C}. \tag{2.68}$$

Now let us see where (X,Y) is sent by our chain of morphisms. We have:

$$\begin{array}{c} X \\ \downarrow \\ X + (X^2(X-1)^2 + Y^2) \\ \downarrow \\ (X + (X^2(X-1)^2 + Y^2)) + ((X-1)^2 - Y^2) = X + (X^2+1) \\ \downarrow \\ i \end{array} \tag{2.69}$$

and so the point we have constructed in order to show that f does not go to zero at some point in $\mathcal{V}(I)$ actually ends up in $\mathbb{C}$, not in our original ring $\mathbb{R}$; in particular, it is here that the proof fails.

Note that the vanishing set of the maximal ideal $\mathcal{V}(M)$ actually does intersect the image of $\mathcal{V}(I)$ at a point of the image of $\mathbb{R}^2$ as well (see the third frame of figure 2.5), and so it is in some sense just bad luck that it actually corresponds to substitution of $X \mapsto i$. (Bad luck is perhaps the wrong word here, as it requires some thought to actually *pick* this bad maximal ideal!)

[4]The term 'local' here is used in a very imprecise sense and should not be confused with the technical manner in which it will be used in chapter 4.

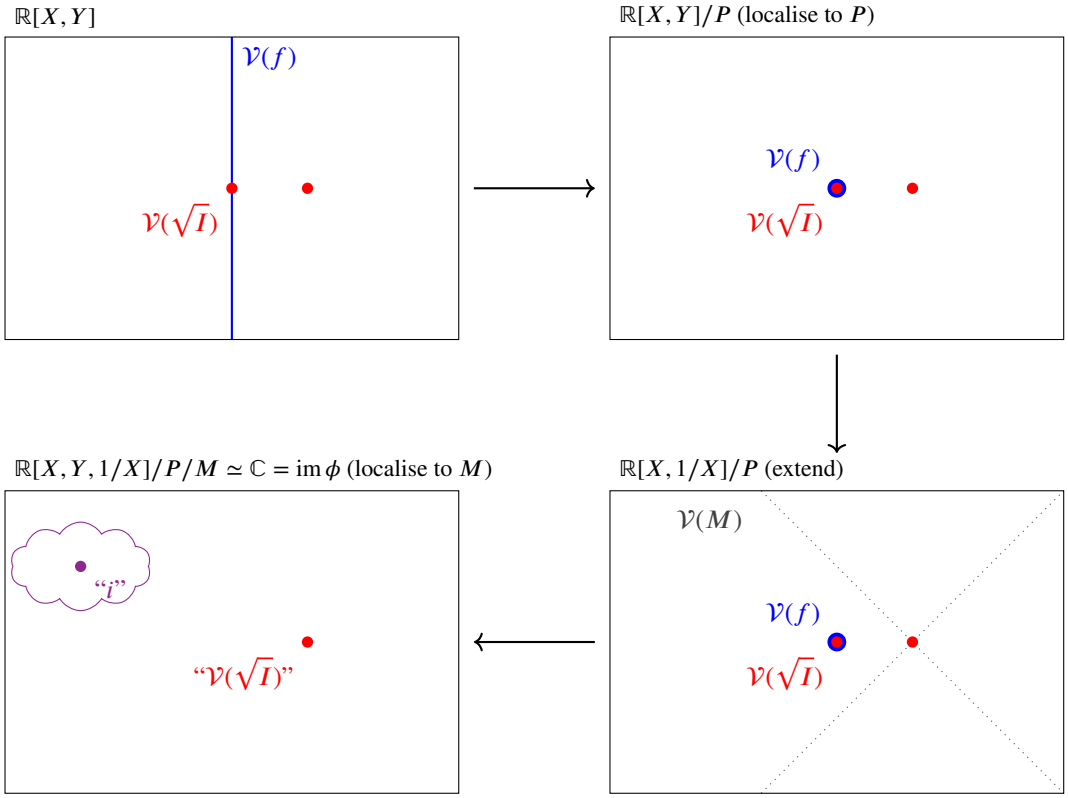


Figure 2.5: The sequence of killings made in example 2.11.8.

Corollaries to the Nullstellensatz

The Nullstellensatz is a very powerful theorem; it enables us to wrap up a lot of threads which we have left loose up to this point. For all the corollaries that follow, K is an algebraically closed field.

2.11.9 Corollary. *Every prime ideal in $K[X_1, ..., X_n]$ is the intersection of the ideals of the form $((X_1 - x_1), ..., (X_n - x_n))$ containing it.*

Proof. Suppose $I \subseteq K[X_1, ..., X_n]$ is prime. We have that $\mathcal{V}(I) = \cup_{P \in \mathcal{V}(I)}\{P\}$. Hence $\mathcal{I}(\mathcal{V}(I)) = \mathcal{I}(\cup_{P \in \mathcal{V}(I)}\{P\}) = \cap_{P \in \mathcal{V}(I)}\mathcal{I}(\{P\})$. By corollary 2.9.10, I is radical; so $I = \mathcal{I}(\mathcal{V}(I))$. In addition, by the characterisation of maximal ideals (corollary 2.11.4), the ideals $\mathcal{I}(\{P\}$ are of the desired form. ∎

2.11.10 Corollary. *A system of polynomial equations in $K[X_1, ..., X_n]$*

$$\begin{aligned} f_1(x_1, ..., x_n) &= 0 \\ &\vdots \\ f_m(x_1, ..., x_n) &= 0 \end{aligned} \tag{2.70}$$

has no solution in K^n iff there exist polynomials $p_1, ..., p_m \in K[X_1, ..., X_n]$ such that

$$1 = \sum_{i=1}^{m} p_i f_i. \tag{2.71}$$

Proof. If the system has no solutions, then $\mathcal{V}(f_1, ..., f_m) = \emptyset$ and so

$$\sqrt{(f_1, ..., f_m)} = \mathcal{I}(\mathcal{V}(f_1, ..., f_m)) = \mathcal{I}(\emptyset) = \mathcal{I}(\mathcal{V}(K[X_1, ..., X_n])) = K[X_1, ..., X_n] \tag{2.72}$$

(where the final equality comes from part 2 of the calculus). Hence $1 \in \sqrt{(f_1, ..., f_m)}$ and so $1 \in (f_1, ..., f_m)$.

Conversely, if $1 = \sum_{i=1}^{m} p_i f_i$ then there cannot be any point which is a zero of all the f_i: it would make the right hand side zero, violating the equality. ∎

2.11.11 Corollary. *Let $I \subseteq K[X_1, ..., X_n]$ be an irreducible ideal. Then $\mathcal{V}(I)$ is irreducible.*

Proof. Irreducible vanishing ideals are prime (proposition 2.6.11) and hence radical (corollary 2.9.10). Hence $I = \mathcal{I}(\mathcal{V}(I))$; suppose $\mathcal{V}(I) = \mathcal{V}(I_1) \cup \mathcal{V}(J)$. Then $I = \mathcal{I}(\mathcal{V}(I_1) \cup \mathcal{V}(J))) = \mathcal{I}(\mathcal{V}(I_1))\mathcal{I}(\mathcal{V}(I_2)) = \sqrt{I_1}\sqrt{I_2}$ by the calculus; by irreducibility, either $I = \sqrt{I_1}$ or $I = \sqrt{I_2}$. Without loss of generality, $I = \sqrt{I_1}$ and hence $\mathcal{V}(\sqrt{I_1}) = \mathcal{V}(I)$. Thus the decomposition of $\mathcal{V}(I)$ is trivial. ∎

2.11.12 Exercise. (a) Is the converse true? (b) Show that the corollary is false if K is not algebraically closed (i.e. exhibit an irreducible ideal with a reducible vanishing set over a field like $\mathbb{R}$).

(Hints: (a) the converse is false unless I is already a vanishing ideal. (b) Over $\mathbb{R}^n$ for some n, try to use proposition 2.6.3.)

2.11.13 Corollary. *If $f \in K[X_1, ..., X_n]$ has irreducible factorisation $f = f_1 \cdots f_r$ then*

$$\mathcal{V}(f) = \mathcal{V}(f_1) \cup \cdots \cup \mathcal{V}(f_r) \tag{2.73}$$

is a unique decomposition of $\mathcal{V}(f)$ into irreducibles (in the style of theorem 2.6.27).

Proof. The existence of the decomposition is clear from parts 5 and 3 of the calculus; each $\mathcal{V}(f_i)$ is irreducible by the previous corollary; and finally, the decomposition is unique by theorem 2.6.27. ∎

2.11.14 Corollary. *The bijective correspondence between radical ideals and algebraic sets may be further decomposed as follows:*

$$\begin{array}{ccc} \textit{Ideals in } K[X_1,...,X_n] & \leftrightsquigarrow & \textit{Subsets of } \mathbb{A}^n_K \\ \cup & & \cup \\ \textit{Radical ideals} & \leftrightarrow & \textit{Algebraic sets} \\ \cup & & \cup \\ \begin{array}{c}\textit{Irreducible ideals}\\ \textit{Prime ideals}\end{array} & \leftrightarrow & \textit{Irreducible sets} \\ \cup & & \cup \\ \textit{Maximal ideals} & \leftrightarrow & \textit{Points} \end{array}$$

∎

2.11.15 Exercise. Find counterexamples to all the above corollaries for a non algebraically-closed field K. Some counterexamples have already been given in these notes!

2.11.16 Exercise. Exercises 1.10, 1.11, and 1.12 of [Eis95] deal with geometric applications of the Nullstellensatz.

Chapter 3

A zoo of examples in affine space

In the previous chapter, we developed a lot of machinery to enable us to discuss algebraic sets in affine space. Before developing more machinery, this time of a geometric flavour, we will meet some basic examples of algebraic sets. We will continue to develop this zoo of examples as we develop our tools.

3.1 Plane conics

The conic sections are some of the most ancient mathematical objects.

3.1.1 Definition. A **plane conic** is a second degree polynomial over a field K in two variables.

We would like to categorise the conics over the complex plane $\mathbb{R}^2$. In classical geometry, we would do this by producing some group of natural transformations G acting on the space (in this case the affine transformations) and calling two curves α and β 'equivalent' in the classification if there was a transformation $g \in G$ such that $g(\alpha) = g(\beta)$.

In algebraic geometry, we are more interested in classifying 'everything at once', in some sense: rather than defining some equivalence relations and then trying to allocate the plane sets to the equivalence classes, we will try to develop tools to detect the subtle algebraic structures that are inherent in the space.

Our categorisation will be as follows for a plane conic C:

1. Reducible conics (traditionally called degenerate conics):-
   (a) A single point — e.g. $C = X^2 + Y^2$;
   (b) A double line — e.g. $C = X^2$;
   (c) A pair of intersecting lines — e.g. $C = X^2 - Y^2$;
   (d) A pair of parallel lines — e.g. $C = (Y - X - 1)(Y - X + 1)$;
2. Irreducible conics (nondegenerate conics):-
   (a) An ellipse — e.g. $C = X^2 + Y^2 - 1$;
   (b) A hyperbola — e.g. $C = X^2 - Y^2 - 1$;
   (c) A parabola — e.g. $C = X^2 - Y$.

In the language of algebraic geometry, these become (if V is the algebraic set related to the conic):

1. Reducible conics:-

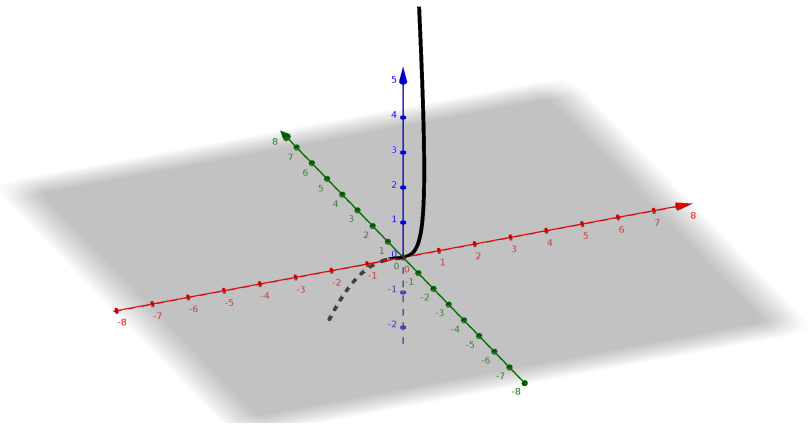


Figure 3.1: The twisted cubic in affine space.

(a) A single point — $\mathcal{I}(V) = I^2$ for a maximal ideal I..

(b) A double line — $\mathcal{I}(V) = I^2$ for some non-maximal ideal I.

(c) A pair of intersecting lines — $\mathcal{I}(V) = IJ$ for two distinct but non-coprime ideals I, J ;

(d) A pair of parallel lines — e.g. $\mathcal{I}(V) = IJ$ for two coprime ideals I, J;

2. Irreducible conics — $\mathcal{I}(V)$ is irreducible.

With the coarse algebraic tools we have built so far, we can distinguish between the various reducible conics. However, the irreducible conics are indistinguishable with respect to their vanishing ideals. The next set of tools we develop, a method of classifying algebraic sets using 'coordinate functions' out of them, will enable us to make sure that this indistinguishability is not just because our tools are too coarse.

We shall complete our initial discussion of conics in the plane with a couple of remarks.

Firstly, we consider intersections of conics. In our initial motivation (chapter 1), we studied the intersections of circles with lines; and we saw that when we moved to the complex plane, we ended up with the 'right number' of intersection points to do some interesting geometry. More generally, one might be interested in the intersections of conics with each other. It is possible to show by elementary methods ([Sal50, §18, §244]) that two conic sections will intersect in at most four points.

A dual problem to picking two conics and looking at their intersections is the problem of picking points and asking how many conics pass through these points. The general equation for a conic is

$$ax^2 + bx + cxy + dy + ey^2 + f = 0; \tag{3.1}$$

since at least one term has to be non-zero, there are five 'degrees of freedom' available and so a conic is uniquely determined by five points.

3.2 The twisted cubic

3.2.1 Definition. The image of the following map into $\mathbb{A}^3_{\mathbb{R}}$ is known as the **twisted cubic**:

$$\varphi : \mathbb{R} \ni t \mapsto (t, t^2, t^3) \in \mathbb{R}^3. \tag{3.2}$$

It is pictured in figure 3.1.

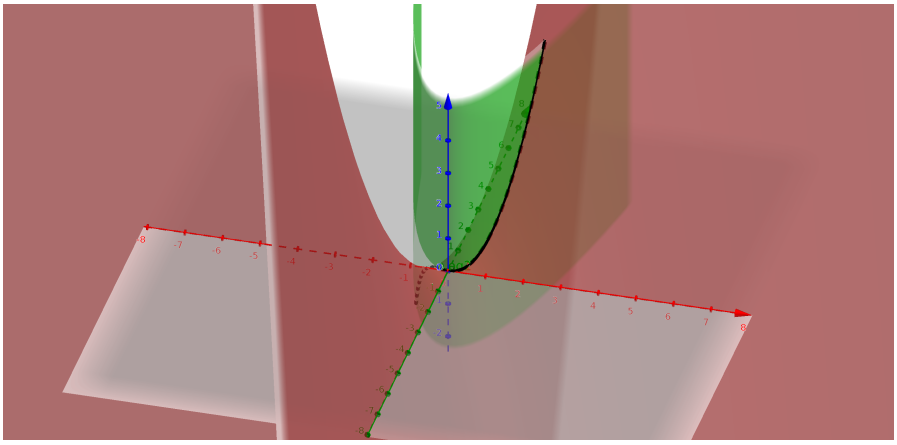


Figure 3.2: The twisted cubic as an intersection of quartic surfaces.

Suppose $P = (X, Y, Z) \in \operatorname{im}\varphi$. Then P satisfies the following equations in $\mathbb{R}[X, Y, Z]$:

$$\begin{aligned} F &:= XY - Z = 0, \\ G &:= X^2 - Y = 0. \end{aligned} \tag{3.3}$$

We have two equations, which seem to be independent, for a curve in three dimensions; it seems that the twisted cubic should be precisely the intersection of these two quadrics. If we graph these curves on one set of axes (figure 3.2), suspicion is aroused even further. And indeed,

3.2.2 Proposition. *The twisted cubic is an algebraic subset of* $\mathbb{A}^3_{\mathbb{R}}$ *given by* $V := \mathcal{V}(F) \cap \mathcal{V}(G)$.

Proof. If $(x, y, z) = (t, t^2, t^3)$ lies on the twisted cubic[1] then $F(P) = F(Q) = 0$ by direct evaluation.

Conversely, suppose $P = (x, y, z) \in \mathcal{V}(F) \cap \mathcal{V}(G)$. Then $y = x^2$ by evaluation of G at P, and $z = xy = x^3$ by substitution of the evaluation of G into the evaluation of F. Hence $(x, y, z) = (x, x^2, x^3)$. ∎

If we were doing analysis, we would note that the map φ is a homeomorphism from $\mathbb{R}$ to $\operatorname{im}\varphi$ and thus that the twisted cubic is connected. But we are not doing analysis, and so we need to work for our results.

3.2.3 Proposition. *The twisted cubic V is irreducible.*

Note that this proposition is not immediately obvious from the algebraic definition: the cubic is not given by the zeroes of a single irreducible polynomial!

Proof. The goal will be to show that $\mathcal{I}(V)$ is a prime ideal; then since $\mathbb{R}$ is an integral domain, $\mathcal{I}(V)$ is irreducible (by lemma 2.7.3) and so V is irreducible (by theorem 2.6.8).

As always, our primary tool for this kind of thing is the first homorphism theorem, corollary 2.1.8, together with knowledge of the evaluation map (and in particular theorem 2.5.18 and its following remarks). We shall consider the following map:

$$\mathbb{R}[X, Y, Z] \xrightarrow{\phi=\mathrm{eval}_{(T,T^2,T^3)}} \mathbb{R}[T]. \tag{3.4}$$

[1] Notationally, I am using lowercase letters to denote coordinates of points in $\mathbb{R}^3$, and uppercase letters for polynomial unknowns.

Consider the ideal (F, G). Clearly if $f \in (F, G)$ we have $\text{eval}_{(T,T^2,T^3)} f = 0$, so $(F, G) \subseteq \ker \phi$. Conversely, suppose $\text{eval}_{(T,T^2,T^3)} f = 0$ for some f. Then f vanishes on the twisted cubic, and so $f \in \mathcal{I}(V) = \sqrt{(F, G)}$.

Consider the following:

$$\begin{array}{c} \mathbb{R}[X, Y, Z] \\ \big\downarrow \phi_0 \\ \mathbb{R}[X, Y, Z, T] \\ \big\downarrow \phi_1 = \text{eval}_{(T,Y,Z,T)} \\ \mathbb{R}[X, Y, Z, T] \\ \big\downarrow \phi_2 = \text{eval}_{(X,T^2,Z,T)} \\ \mathbb{R}[X, Y, Z, T] \\ \big\downarrow \phi_3 = \text{eval}_{(X,Y,T^3,T)} \\ \mathbb{R}[X, Y, Z, T] \\ \big\downarrow \phi_4 = \text{eval}_{(T,T,T,T)} \\ \mathbb{R}[T] \end{array} \tag{3.5}$$

The composition $\phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1 \circ \phi_0$ is just ϕ. We now calculate the kernels of each morphism in the chain. Firstly, $\ker \phi_1 = (X - T, Y - Y, Z - Z, T - T) = (X - T) \subseteq \mathbb{R}[X, Y, Z, T]$; so $\ker(\phi_1 \circ \phi_0)$ is empty since nothing in the image of ϕ_0 is a multiple of $(X - T)$.

Secondly, $\ker \phi_2 = (X - X, Y - T^2, Z - Z, T - T) = (Y - T^2)$. An element g in the image of $\phi_1 \circ \phi_0$ is contained in $(Y - T^2)$ iff $g = \phi_1(g'(Y - X^2))$ for some $g' \in \mathbb{R}[X, Y, Z]$. Thus the kernel of $\phi_2 \circ \phi_1 \circ \phi_0$ is $(Y - X^2)$.

Now consider $\ker \phi_3 = (Z - T^3)$. An element may be mapped onto T^3 by ϕ_2 in two ways: $T^3 = \phi_2(T^3) = \phi_2(\phi_1(X^3))$, or $T^3 = \phi_2(YT) = \phi_2(\phi_1(XY))$ Hence $\ker \phi_3 \circ \phi_2 \circ \phi_1 \circ \phi_0 = ((XY - Z), (Z - X^3), (Y - X^2))$. Finally, ϕ_4 is the identity on the image of $\phi_3 \circ \phi_2 \circ \phi_1 \circ \phi_0$.

Thus $\ker \phi = ((XY - Z), (Z - X^3), (Y - X^2))$. But note that $Z - X^3 = X(Y - X^2) - (XY - Z)$; so $\ker \phi = ((XY - Z), (Y - X^2)) = (F, G)$.

In particular, ϕ is a surjective homomorphism $\mathbb{R}[X, Y, Z] \to \mathbb{R}[T]$; thus $\frac{\mathbb{R}[X,Y,Z]}{(F,G)} \simeq \mathbb{R}[T]$. Since $\mathbb{R}[T]$ is an integral domain (lemma 2.2.3), (F, G) is a prime ideal and we are done. ∎

3.2.4 Corollary. $\mathcal{I}(V) = (F, G)$. ∎

3.2.5 Exercise. The twisted cubic is not the vanishing set of a single polynomial.

3.3 Plane conchoids

Let $\ell \subseteq K^2$ be a line, let $P \in K^2$ be a point, and let $b \in \mathbb{R}$ be non-negative. Consider the pencil of lines at P which are not parallel to ℓ: call this family $\{m_\alpha\}$. For each α set Q_α to be the point $\ell \cap m_\alpha$. Define the set $CN(\ell, P, b)$ by the rule

$$CN(\ell, P, b) := \{R \in K^2 : \exists_\alpha(\rho(Q_\alpha, R) = b \text{ and } R \in m_\alpha)\} \tag{3.6}$$

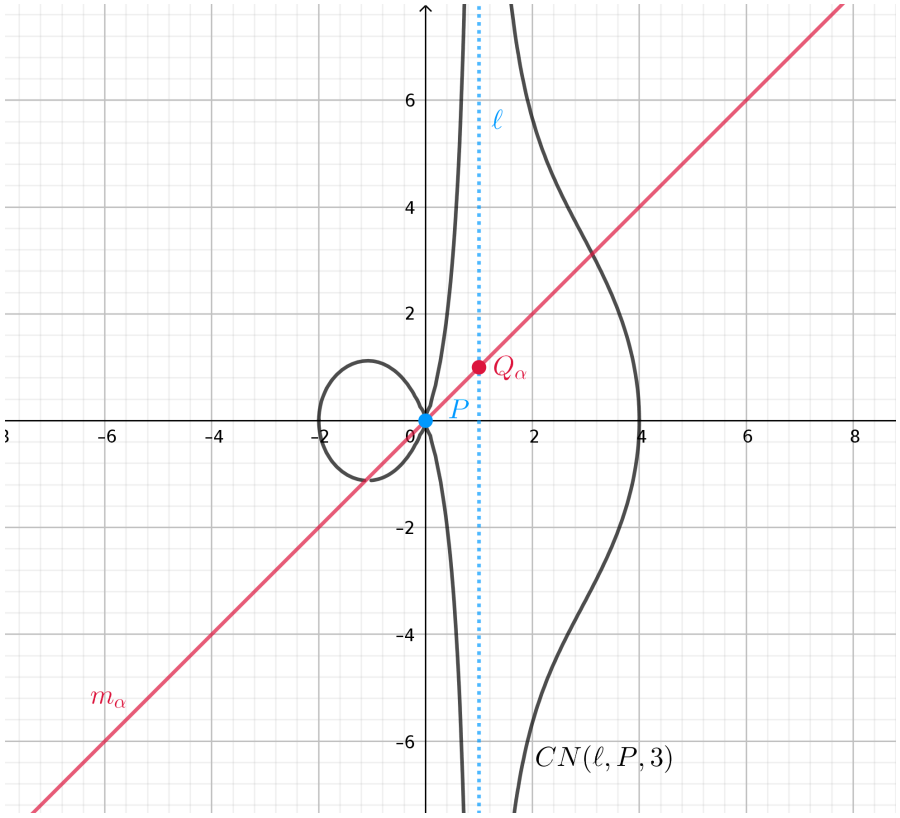


Figure 3.3: A conchoid of Nichomedes.

(where ρ denotes the Euclidean metric on K^2).

3.3.1 Definition. Each set $CN(\ell, P, b)$ is called a **conchoid of Nichomedes**.

3.3.2 Example. The set $CN(\mathcal{V}(X-1), (0,0), 3)$ is displayed as figure 3.3.

3.3.3 Proposition. *The set $CN(\ell, P, b)$ is an algebraic subset of $\mathbb{A}^2$.*

Proof. By example 4.2.15 below it will suffice to show this for a specific choice of ℓ and P (since changing this choice corresponds to a combination of translation, rotation, and magnification). For convenience we will choose $P = (0,0)$, and ℓ to be the line $\mathcal{V}(X-1)$. Thus the pencil of lines through P and skew to ℓ are the lines of the form $m_\alpha = \mathcal{V}(Y - \alpha X)$ for $\alpha \in K$. For each of these, $Q_\alpha = (1, \alpha)$, and the circle of radius b about this point is given by $\mathcal{V}((X-1)^2 + (Y-\alpha)^2 - b^2)$. Now consider:-

$$\begin{aligned} & (X-1)^2 + (Y-\alpha)^2 = b^2 \\ \Longrightarrow\ & X^2(X-1)^2 + X^2(Y-\alpha)^2 = X^2 b^2 \\ \Longleftrightarrow\ & X^2(X-1)^2 + (XY - \alpha X)^2 = X^2 b^2 \\ \Longleftrightarrow\ & X^2(X-1)^2 + (XY - Y)^2 = X^2 b^2 \\ \Longleftrightarrow\ & (X^2+Y^2)(X-1)^2 = X^2 b^2 \end{aligned} \tag{3.7}$$

and let $V = \mathcal{V}\left((X^2+Y^2)(X-1)^2 - X^2 b^2\right)$. By construction, if $R \in CN(\ell, P, b)$ then $R \in V$. Conversely, suppose $(x, y) \in V$. If $x = 0$ then the equation of V implies that $y = 0$; if $x \neq 0$ then we may set

$\alpha = x/y$ and run the chain of equations (3.7) backwards to see that (x, y) has distance b from Q_α and lies on the line m_α.

Hence $CN(\ell, P, b) = \mathcal{V}\left((X^2 + Y^2)(X-1)^2 - X^2b^2\right)$ and we are done. ∎

3.3.4 Proposition. *The algebraic set $CN(\ell, P, b)$ is irreducible if and only if $b \neq 0$.*

Proof. As above (now appealing to proposition 4.2.27 below as well), we need only show this for the curve given by $\ell = \mathcal{V}(X-1)$ and $P = (0,0)$. Thus $CN(\ell, P, b) = \mathcal{V}\left((X^2 + Y^2)(X-1)^2 - X^2b^2\right)$ and hence it suffices to show that $(X^2 + Y^2)(X-1)^2 - X^2b^2$ is irreducible iff $b \neq 0$.

Expanding as a polynomial in $K[X][Y]$, we obtain

$$p = (X^2 - 2X + 1)Y^2 + (X^4 - 2X^3 + (1 - b^2)X^2) \tag{3.8}$$

and so if p is reducible then it must factor as $p = fg$ where f and g are linear in Y; i.e. $f = \alpha Y + \beta$ and $g = \gamma Y + \delta$ for α and γ non-zero. Expanding fg we obtain $fg = \alpha\gamma Y^2 + (\alpha\delta + \beta\gamma)Y + \beta\delta$, and comparing coefficients we have the system of equations

$$\begin{aligned} \alpha\gamma &= (X^2 - 2X + 1) = (X-1)^2 \\ \alpha\delta + \beta\gamma &= 0 \\ \beta\delta &= (X^4 - 2X^3 + (1-b^2)X^2) = X^2(X^2 - 2X + (1-b^2)). \end{aligned} \tag{3.9}$$

Note that $X - 1$ is irreducible. Hence $(X-1) \mid \alpha$ or $(X-1) \mid \beta$ by the first of the equations 3.9. If $(X-1) \mid \beta$ then $(X-1) \mid \beta\delta$, so $X - 1 \mid X^2(X^2 - 2X + (1-b^2))$ by the third equation; if this is so then we must have $1(1 - 2 + (1 - b^2)) = 0$ and thus $b = 0$ On the other hand, if $(X-1) \mid \alpha$ then by the second equation we have $(X-1) \mid \beta\gamma$, and again we have $b = 0$ by the third equation. Thus if p is reducible then $b = 0$.

Conversely, if $b = 0$ then the polynomial $(X^2 + Y^2)(X-1)^2 - X^2b^2$ becomes $(X^2 + Y^2)(X-1)^2$; it is hard not to see that this polynomial is reducible. ∎

The line ℓ is, analytically speaking, an asymptote of the curve. We shall attempt later on to characterise this notion algebraically, without the use of any limiting process.

We can generalise the construction of the conchoid of Nichomedes to the case where our 'backbone' curve is not a line.

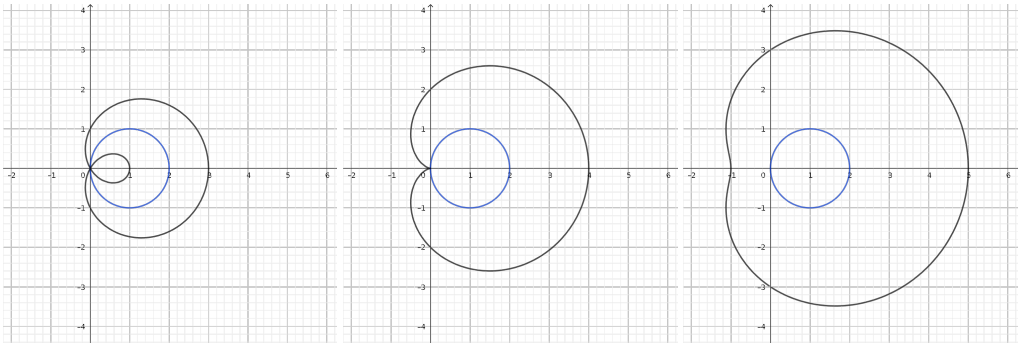
3.3.5 Definition. Let $V \subseteq K^2$ be a plane algebraic curve, let $P \in K^2$ be a point, and let $b \in K$ be non-negative. Let $\{m_\alpha\}$ be the pencil of lines at P; then the set $C(V, P, b)$ that is the locus of all points Q which lie on some m_α and which are a distance b from $m_\alpha \cap V$ is called a **conchoid**.

3.3.6 Example.

1. If V is a line, we clearly obtain the conchoid of Nichomedes.
2. If the curve V is a circle and $P \in V$ then the curve obtained is called a **limaçon of Pascal**. The case where b is twice the diameter of V is called a **cardioid**, for reasons that will become apparent when the reader glances at figure 3.4.

3.3.7 Exercise. What are the forms of the conchoids on other conics?

3.3.8 Exercise. Show that every conchoid is an algebraic curve and obtain a characterisation for the irreducibility of the conchoid based on V, P, and b in analogy with proposition 3.3.4.



(a) $b = 1$ (b) $b = 2$ (c) $b = 3$

Figure 3.4: The limaçon of Pascal (example 3.3.6) on the circle $X^2 + Y^2 = 1$.

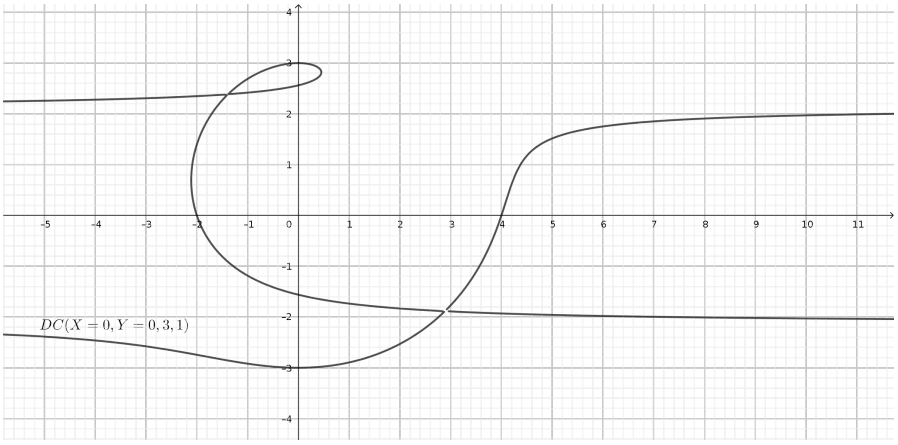


Figure 3.5: A conchoid of Dürer (exercise 3.3.9).

3.3.9 Exercise. There are curves which are called conchoids but are not strictly conchoids as defined above.

Suppose two perpendicular lines m_1 and m_2 are chosen in K^2, meeting at a point O. Fix $a, b \in R$. Let S be the set of pairs of points $(P_1, P_2) \in K^2 \times K^2$ such that $P_1 \in m_1$, $P_2 \in m_2$, and $\rho(P_1, O) + \rho(P_2, O) = b$. Then define

$$DC(m_1, m_2, a, b) := \{Q : \exists_{(P_1,P_2)\in S}(Q \in \overline{P_1P_2} \text{ and } d(P_1, Q) = a)\}. \tag{3.10}$$

Each set $DC(m_1, m_2, a, b)$ is called a **conchoid of Dürer** (see figure 3.5). Show that:-

1. Each conchoid of Dürer is an algebraic subset of K^2;
2. The curve is irreducible iff $b \neq 0$.

Further information on conchoids may be found in [Hol12, §3.7] and [Law72]; a detailed study of the conchoid of Dürer may also be found in [Fet83], and the original discussion by Dürer is reproduced as [BK86, pp. 48–51].

Chapter 4

Affine varieties

In this chapter we will enrich our understanding of vanishing sets and vanishing ideals in affine space by considering maps between them; we will use these maps (by analogy with more continuous subjects, like complex analysis or differential geometry) to study the interplay between the local and global properties of the objects themselves.

4.0.1 Assumption. From this point onwards, unless otherwise stated, K will be an algebraically closed field of characteristic zero (i.e. the unique homomorphism $\mathbb{Z} \to K$ is injective). One consequence of this is that K will be infinite (exercise 2.7.12). A number of the examples and theorems are actually valid over arbitrary fields (particularly the technical lemmata), but we leave the reader to look at a book on field theory like [Nag77] for the general statements.

4.0.2 Notation. From this point forward, we will use Fraktur letters ($\mathfrak{a}$, $\mathfrak{b}$, ...) to denote ideals.

4.1 Categories

> *"In other words, general set theory is pretty trivial stuff really, but, if you want to be a mathematician, you need some, and here it is; read it, absorb it, and forget it."* [Hal15b, p. vi]
> The same may be said about basic category theory.

For remainder of these notes, it will often be convenient to use terminology from category theory. We shall not use any deep results about categories: a good book for the beginner is [AHS90] and we will not use much beyond chapter 3 of that text. The series of online videos [CW] is also very useful.

4.1.1 Definition. A **category** C consists of the following data:

1. A collection Ob_{C} of 'objects';
2. For each pair A, B of objects, a collection $\mathrm{Arr}_{\mathsf{C}}(A, B)$ of 'arrows' or 'morphisms';
3. For every triplet A, B, C of objects, and every pair of arrows $f \in \mathrm{Arr}_{\mathsf{C}}(A, B)$ and $g \in \mathrm{Arr}_{\mathsf{C}}(B, C)$, an arrow $g \circ f \in \mathrm{Arr}\,\mathsf{C}(A, C)$, such that if h is a third arrow $h \in \mathrm{Arr}_{\mathsf{C}}(C, D)$ (where D is an arbitrary object) we have $(h \circ g) \circ f = h \circ (g \circ f)$;
4. For every object A an arrow $\iota_A \in \mathrm{Arr}\,\mathsf{C}(A, A)$ such that for all arrows $f \in \mathrm{Arr}_{\mathsf{C}}(A, B)$ and all arrows $g \in \mathrm{Arr}_{\mathsf{C}}(B, A)$ (for any other object B), $f \circ \iota_A = f$ and $\iota_A \circ g = g$. The arrow ι_A is called the **identity morphism** on A.

To avoid pathologies, we will additionally ensure that within a category C, all of the collections of arrows and objects are pairwise disjoint.

A **subcategory** of C is a category D whose collection of objects is a subcollection $\mathrm{Ob}_{\mathsf{D}} \subseteq \mathrm{Ob}_{\mathsf{C}}$, and such that if $A, B \in \mathrm{Ob}_{\mathsf{D}}$ then $\mathrm{Arr}_{\mathsf{D}}(A, B) = \mathrm{Arr}_{\mathsf{C}}(A, B)$.

4.1.2 Example. To fix notation, here are some familiar categories.

1. The category of sets, with set functions: Set.
2. The category of groups, with group homomorphisms: Grp.
3. The category of abelian groups, with group homomorphisms: Ab.
4. The category of rings, with ring homomorphisms: Ring.
5. The category of fields, with ring homomorphisms: Fld.
6. The category of topological spaces, with continuous functions: Top.
7. Given some fixed ring R, the category of R-algebras, with R-algebra homomorphisms: $\mathsf{Alg}(R)$.
8. Any poset $(A, \leq)$ is a category in a natural way: the objects are elements of a, and there is an arrow $a \to b$ iff $a \leq b$.

The majority of categories we will deal with here are **small**: that is, the collection of objects and all the collections of arrows are actually sets.

Note that we have some notion of 'inclusion' between categories (e.g. between Ab and Grp). Even stronger, we may define maps between categories.

4.1.3 Definition. If C and D are categories, a **functor** $\mathscr{F}$ from C to D is a collection of maps (all also denoted by $\mathscr{F}$) consisting of

1. a map $\mathscr{F} : \mathrm{Ob}_{\mathsf{C}} \to \mathrm{Ob}_{\mathsf{D}}$; and
2. for each pair A, B of objects in C, a map

$$\mathscr{F} : \mathrm{Arr}_{\mathsf{C}}(A, B) \to \mathrm{Arr}_{\mathsf{D}}(\mathscr{F}(A), \mathscr{F}(B)). \tag{4.1}$$

We require these maps to satisfy two additional properties:

F1. for all $A \in \mathrm{Ob}_{\mathsf{C}}$, $\mathscr{F}(\iota_A) = \iota_{\mathscr{F}(A)}$ (where ι_X denotes the identity on X); and

F2. if $f \in \mathrm{Arr}_{\mathsf{C}}(A, B)$ and $g \in \mathrm{Arr}_{\mathsf{C}}(B, C)$ then $\mathscr{F}(f \circ g) = \mathscr{F}(f) \circ \mathscr{F}(g)$.

4.1.4 Remark. The two properties F1 and F2 have two main consequences.

F1′. A functor preserves **isomorphism**: if A and B are objects of C such that there exists a morphism $f : A \to B$ and a morphism $g : B \to A$ such that $f \circ g = \iota_B$ and $g \circ f = \iota_A$, then the images of A, B, f, and g in D have the same property.

F2′. A functor preserves commutative diagrams: if a diagram in C commutes, then the diagram formed by the images of its component objects and morphisms in D commutes.

The reader should now forget definition 4.1.3 and view a functor as a map which has the properties F1′ and F2′.

Analogous to the concepts of 'injectivity' and 'surjectivity' for set-functions, we may classify functors as follows.

4.1.5 Definition. If C and D are categories, a functor $\mathscr{F} : \mathsf{C} \to \mathsf{D}$ is called

1. an **isomorphism** if it has an inverse (in the natural sense);
2. a **faithful** functor if every induced map
$$\mathscr{F} : \mathrm{Arr}_{\mathsf{C}}(A, B) \to \mathrm{Arr}_{\mathsf{D}}(\mathscr{F}(A), \mathscr{F}(b)) \tag{4.2}$$
is individually injective (i.e. it is injective on pairs of objects);
3. a **full** functor if every map (4.2) is surjective.

4.1.6 Example. The functor $\mathscr{U} : \mathsf{Ring} \to \mathsf{Set}$ obtained by 'throwing away the ring structure' (called a **forgetful** functor) is faithful and bijective on objects but not full.

4.1.7 Exercise. Most results in category theory are of the type that should not be proved in public; the proofs are not hard, but the notation quickly becomes messy.

1. A functor is an isomorphism if and only if it is (1) bijective on objects, (2) faithful, and (3) full.
2. Let $\mathscr{F} : \mathsf{A} \to \mathsf{B}$ and $\mathscr{G} : \mathsf{B} \to \mathsf{C}$ be functors.
   (a) If $\mathscr{F}$ and $\mathscr{G}$ are both (isomorphisms/faithful/full) then so is the composition $\mathscr{G} \circ \mathscr{F}$.
   (b) If $\mathscr{G} \circ \mathscr{F}$ is faithful, then so is $\mathscr{F}$.
   (c) If $\mathscr{G} \circ \mathscr{F}$ is full, then so is $\mathscr{G}$.
3. Recall that property F2′ is that functors preserve isomorphism (remark 4.1.4). Let $\mathscr{F} : \mathsf{A} \to \mathsf{B}$ be full and faithful.
   (a) For every arrow $f : \mathscr{F}(a) \to \mathscr{F}(b)$ in B, there is a unique arrow $g : a \to b$ in A such that $\mathscr{F}(g) = f$.
   (b) For the setup in (a), f is an isomorphism iff g is an isomorphism.

Unfortunately, the notion of isomorphism of categories is often too restrictive.

4.1.8 Example. Let MTop be the category of *metrisable* topological spaces with continuous maps as morphisms, and let Met be the category of metric spaces with continuous maps. Then we have a functor $\mathscr{F} : \mathsf{Met} \to \mathsf{Top}$ which assigns to each metric space its induced topological space has no well-defined inverse since every metrisable topological space is metrisable in more than one way (see [Mun00, theorem 20.1]); on the other hand, all metric spaces arising from the same metrisable topological space are homeomorphic (since homeomorphism is a topological property, not dependent on the particular metric).

Hence $\mathscr{F}$ is an 'isomorphism of categories, modulo some isomorphism of objects'.

4.1.9 Definition. If C and D are categories, a functor $\mathscr{F} : \mathsf{C} \to \mathsf{D}$ is called an **equivalence** (of categories) if it is full, faithful, and for every object $D \in \mathrm{Ob}_{\mathsf{D}}$ there exists some object $C \in \mathrm{Ob}_{\mathsf{C}}$ such that $\mathscr{F}(C)$ is isomorphic (in D) to D. We say that C and D are **equivalent**.

It turns out that this notion (slightly weaker than isomorphism) is the correct notion of 'sameness' for categories.

Another mathematical concept which occurs fairly often is a 'map which reverses relations'; for example, the $\mathcal{I}$-$\mathcal{V}$ correspondence, or the Galois correspondence, or the equivalence between finite dimensional vector spaces and their duals.

4.1.10 Definition. If C is a category, then C^{op} is the category whose objects are precisely the objects of C, and such that $f^{\mathrm{op}} : A \to B$ is a morphism in C^{op} if and only if $f : B \to A$ is a morphism in C. We refer to C^{op} as the **opposite category** of C.

A **contravariant functor** from C to D is a functor from $\mathsf{C}^{\mathrm{op}} \to \mathsf{D}$; to distinguish it, a 'usual' functor is sometimes called **covariant**.

Two categories C and D are **dually equivalent** if C^{op} is equivalent to D.

4.1.11 Example. Consider the (hopefully disjoint) union of the sets of humans and cats, and take this to be the set of objects for the category C. Define the morphisms as follows: if $a, b \in \mathrm{Ob}_{\mathsf{C}}$, then $a \to b \in \mathrm{Arr}_{\mathsf{C}}(a, b)$ iff a owns b. We shall make some assumptions, which may or may not reflect the real world: (1) every human owns themselves; (2) every cat owns themselves; (3) no human owns another human; (4) no cat owns another cat; (5) no cat owns a human.

Now consider a book on category theory written by a cat mathematician; they would naturally consider the opposite category C^{op} to be the more correct model of this situation.

4.2 Coordinate rings

In the proof of the Nullstellensatz, and even earlier in remark 1.2.1, we were interested in studying the restrictions of polynomials to particular algebraic sets; we saw that an algebraic way of doing this was to quotient out by the ideal corresponding to the algebraic set (since this is a generalisation of the idea of 'substitution' in one variable). In this section we shall formalise this idea, since in general knowing about the collection of maps out of an object over some ambient space usually gives us some idea of its global structure; and knowing about the collections of maps between objects tells us about their mutual relationship within that ambient space.

4.2.1 Definition. Let $V \subseteq \mathbb{A}^n_K$ be an algebraic set. A function $f : V \to K$ is called a **polynomial function** if there exists a polynomial $p \in K[X_1, ..., X_n]$ such that for all $P \in V$, $p(P) = f(P)$. The set of all polynomial functions on V is a ring, denoted by $K[V]$, called the **coordinate ring** of V.

We shall often write $[f]$ for the member of $K[V]$ corresponding to the polynomial f, although equally often we shall simply write f for both things. Since the two are for all intents and purposes equal 'as functions' this should lead to no confusion.

4.2.2 Example. Let $V \subseteq \mathbb{A}^n$; then for each i, $1 \leq i \leq n$, the restriction $\pi_i \restriction_V$ is a polynomial map. Indeed, $\pi_i : \mathbb{A}^n \to K$ is given for all P by $\pi(P) = \mathrm{eval}_P X_i$. We call these restrictions, which we will still denote by π_i, the **coordinate functions** of V.

4.2.3 Proposition. *If $V = \mathcal{V}(\mathfrak{a})$ for some ideal $\mathfrak{a}$, then $K[V] \simeq K[X_1, ..., X_n]/\mathfrak{a}$.*

Proof. Define $\phi : K[X_1, ..., X_n] \to K[V]$ to be the restriction map, $f \mapsto f \restriction_V$. Clearly ϕ is a ring homomorphism; and $f \in \ker \phi$ iff $\forall_{P \in V} f(P) = 0$ iff $f \in \mathcal{I}(V) = \mathcal{I}(\mathcal{V}(\mathfrak{a})) = \mathfrak{a}$. (Note we do require K to be algebraically closed!) Hence we are done by the first homomorphism theorem. ∎

4.2.4 Corollary. *By theorem 2.6.21, $K[V]$ is Noetherian.* ∎

4.2.5 Remark. This proposition explicitly shows that $K[V]$ is a finitely generated K-algebra over K, generated by the restriction maps π_i from the preceeding example. Indeed, the coordinate maps themselves generate $K[X_1, ..., X_n]$, and hence their residues modulo I generate the quotient ring. The only problem is that this generation might be redundant — we might not *need* all the generators we get (see parts 1 and 3 of example 4.2.6 below).

Another way of thinking about this is as follows: two polynomials $f, g \in K[X_1, ..., X_n]$ describe the same function on V if and only if $f - g = 0$ for all $P \in V$; i.e. if $f - g \in \mathfrak{a}$. Thus the equivalence relation $f \equiv g \iff f - g \in \mathfrak{a}$ is the relation we should quotient out by in order to make the polynomials become equal — and this is just quotienting out by $\mathfrak{a}$.

4.2.6 Example. Examples of coordinate rings include:

1. Let $P \in \mathbb{A}^n$ be a point; then $K[P] \simeq \frac{K[X_1,...,X_n]}{(P)}$. But (P) is maximal, and hence by the weak Nullstellensatz (lemma 2.11.3), $K[P] \simeq K$. (This is very natural, since a function out of a singleton is uniquely determined by its image. This is one reason we need to work over an algebraically closed field (assumption 4.0.1) — otherwise we lose this naturality.)

2. The coordinate ring $K[\mathbb{A}^n]$ is isomorphic to $K[X_1, ..., X_n]$; this is since (by part 1 of the calculus) $\mathcal{I}(\mathbb{A}^n) = (0)$.

3. Let $V = \mathcal{V}(X^2 - Y)$; then $\mathcal{I}(\mathcal{V}(Y - X^2)) = (Y - X^2)$ (since $Y - X^2$ is irreducible and hence radical, corollary 2.9.10) and so $K[V] \simeq K[X, Y]/(Y - X^2)$; note that $\text{eval}_{(X,X^2)} : K[X, Y] \to K[X]$ has kernel $(Y - X^2)$ (theorem 2.5.18) so $K[X, Y]/(Y - X^2) \simeq K[X]$; hence $K[V] \simeq K[X]$. In other words, every polynomial $f \in K[X]$ determines a polynomial function $f + Y - X^2$ on V. Note also that the function's values at a point on V are determined by the position of the projection onto the X-axis.

4. Let $W = \mathcal{V}(X^2 - Y^2)$. Then $\mathcal{I}(W) = (X^2 - Y^2)$ (check this is radical); so the coordinate ring is $K[W] = K[X, Y]/(X^2 - Y^2)$. But note that we have $[X + Y][X - Y] = [0]$ in this ring, so it is not an integral domain. (The problem is that the first function vanishes on one irreducible component, the second function on the other, and so their product identically vanishes.)

4.2.7 Exercise. Check the extreme cases: V is the empty set iff $K[V] = K[X_1, ..., X_n]$; $V = \mathbb{A}^n$ iff $K[V] = (0)$. (Are these true if K is not algebraically closed?)

4.2.8 Exercise. Show that $\mathcal{V}(XY - 1) \simeq K[X, X^{-1}]$.

Polynomial functions are fairly rigid: they can be reconstructed from their values in a dense open subset of the algebraic set.

4.2.9 Proposition. *Let V be an algebraic set and let $f, g \in K[V]$. If there exists a dense open subset $U \subseteq V$ such that $f(P) = g(P)$ for all $P \in U$, then $f(P) = g(P)$ for all $P \in V$.*

Proof. Let $S \subseteq V$ be the set of points where they disagree; then $V \setminus S = \mathcal{V}(f - g) \cap V$. So $V \setminus S$ is open in V, and thus either intersects U or is empty. Since the two polynomials functions agree on U, the set of points where they disagree must therefore be empty; hence $V \setminus S = \emptyset$, $S = V$, and the two polynomials functions agree everywhere on V. ∎

In $\mathbb{A}^n$, because every point is specified by coordinates, having functions in the coordinate ring is enough to define maps between algebraic sets in a natural way: we just send each point through n polynomial maps at once and then stick all of them together as the coordinates of a point in the image.

4.2.10 Definition. Let $V \subseteq \mathbb{A}^m$ and $W \subseteq \mathbb{A}^n$ be algebraic sets. A function $f : V \mapsto W$ is a **polynomial map** is the restriction of a function $\tilde{f} : V \mapsto K[X_1, ..., X_n]$ such that

1. For every projection map $\pi_i : K^n \to K$, $\pi_i \circ \tilde{f}$[1] is a polynomial function on V, and

[1] Formally, since polynomials are not functions, this is shorthand for $\text{eval}_{\text{eval}_{(X_1,...,X_n)}\,\tilde{f}}\,\pi_i : V \to K$. See remark 4.2.17 below.

2. The image of $\tilde{f}$ lies in W.

4.2.11 Remark. A function $\tilde{f}$ determines a polynomial map $f : V \to W$ in the above sense iff both the following hold:

1. Condition 1 in the definition holds, and
2. For every $g \in \mathcal{I}(W)$ we have $g \circ f = 0$ (as functions).

4.2.12 Example.

1. The identity function $\iota_V : V \to V$ is a polynomial map as it is the restriction of $\tilde{\iota}_V : V \to K[X_1, ..., X_n]$ given by $\tilde{f} = (X_1, ..., X_n)$. Alternatively, we may view the identity as the map $(\pi_1, ..., \pi_n)$ where each π_i is the standard coordinate function.
2. Every polynomial function $f : V \to K$ is a polynomial map.

4.2.13 Example. Recall that the twisted cubic (definition 3.2.1) is the image V of the map

$$\varphi : K \ni t \mapsto (t, t^2, t^3) \in K^3. \tag{4.3}$$

The map φ is a polynomial map $\mathbb{A}^1 \to \mathbb{A}^3$. Further, the polynomial map has an inverse: $\psi : V \to K$ given by the restriction $\pi_1 \restriction_V$.

Motivated by this, we may make the following definition:

4.2.14 Definition. A polynomial map $f : V \to W$ is an **isomorphism** if there exists a polynomial map $g : W \to V$ such that $(f \circ g)(P) = P$ for all $P \in W$ and $(g \circ f)(Q) = Q$ for all $Q \in V$. We say that V and W are **isomorphic** if there is an isomorphism between them, and we write $V \simeq W$.

4.2.15 Example.

1. By the above example, the twisted cubic is isomorphic to the affine line, $\mathbb{A}^1$.
2. An **affine transformation** is a function $f : K^n \to K^m$ defined by

$$f(P) = AP + b \tag{4.4}$$

for some $A \in \mathrm{GL}_n(K)$ and some $b \in K^n$. Every affine transformation, restricted to an algebraic set $V \subseteq \mathbb{A}^n$, is an isomorphism $V \simeq f(V)$.

In particular, if $V \subseteq \mathbb{A}^n$ is an algebraic subset then so is $f(V)$.

4.2.16 Exercise. Show that every ellipse in $\mathbb{R}^2$ (i.e. curve of the form $\mathcal{V}(aX^2 + bX + cXY + dY + eY^2)$ for $ae > 0$) is isomorphic to the circle $\mathcal{V}(X^2 + Y^2 - 1)$. Perform a similar act for the hyperbolae (i.e. curves of the form $\mathcal{V}(aX^2 + bX + cXY + dY + eY^2)$ for $ae < 0$) are isomorphic; and the parabolae (curves of the form $\mathcal{V}(aX^2 + bX + cXY + dY + eY^2)$ for $ae = 0$ but one of a or e nonzero).

The natural question to ask is, therefore, the following:-

If V and W are isomorphic, what is the relationship between $K[V]$ and $K[W]$? (4.5)

4.2.17 Remark. Based on the above remarks, for each field K we have a category $\mathsf{AffAlgSet}(K)$ such that

1. $\mathrm{Ob}_{\mathsf{AffAlgSet}(K)}$ is the set of the affine algebraic sets over K;
2. $\mathrm{Arr}_{\mathsf{AffAlgSet}(K)}(V, W)$ is the set of polynomial maps from V to W;

3. Composition $f \circ g$ (for $g \in \mathrm{Arr}(V_1, V_2)$, $f \in \mathrm{Arr}(V_2, V_3)$) is defined by composition of evaluation maps between the rings: $f \circ g := \mathrm{eval}_{\mathrm{eval}_{(X_1,\ldots,X_n)}\, g}\, f \in \mathrm{Arr}(V_1, V_3)$;

4. Every object V has an identity morphism, $\iota : V \to V$.

Thus a slightly more precise way to phrase question 4.5 is:

What is the relationship between the category $\mathsf{AffAlgSet}(K)$ *and the category of* K*-algebras?* (4.6)

The answer to this formulation, contained within the following theorem and theorem 4.2.21 below, is (roughly speaking) that 'the relationship is as nice as could be expected'.

4.2.18 Theorem. *There is a subcategory* C *of* $\mathsf{Alg}(K)$ *which is dually equivalent to* $\mathsf{AffAlgSet}(K)$*, by some contravariant functor which we will denote* $\cdot^*$*; the inverse of* $\cdot^*$ *will be denoted by* $\cdot_*$*.*
More precisely:-

1. *For every algebraic set* V*, there is a* K*-algebra* V^* *such* $V \simeq W$ *as algebraic sets if and only if* $V^* \simeq W^*$ *as* K*-algebras;*

2. *For every pair of algebraic sets* V, W *there is a bijection between the set of polynomial maps from* V *to* W *and the set of* K*-algebra homomorphisms from* W^* *to* V^**. If* $f : V \to W$ *then the corresponding* K*-algebra homomorphism is denoted by* $f^* : W^* \to V^*$*; and if* $g : W^* \to V^*$ *then the corresponding polynomial map will be denoted by* $g_* : V \to W$*.*

3. *The bijection of part 2 reverses compositions:* $f^* \circ g^* = (g \circ f)^*$*, and* $f_* \circ g_* = (g \circ f)_*$*.*

4.2.19 Remark. Clearly V^* will be the coordinate ring $K[V]$, so the only thing here that really needs to be 'intuitively' justified is why this correspondence 'should be' contravariant; the answer is that if $V \subseteq W$ there is already a natural map $f : V \mapsto W$ (namely the inclusion map); by the $\mathcal{I}$-$\mathcal{V}$ correspondence we have an inclusion $K[W] \subseteq K[V]$, with another natural map $g : K[W] \to K[V]$; and for the whole thing to be 'nice' we should have that $g = f^*$.

Proof of theorem 4.2.18. The proof is lengthy, but does not really require any creative thinking.

As remarked, for every algebraic set V define V^* to be the coordinate ring $K[V]$. We need to show that this mapping is injective up to isomorphism; in order to do this, we shall use the results from parts 2 and 3 of the theorem. Thus we will prove those first.

Part 2: The functor on arrows. In order to obtain a contravariant functor we need to additionally define a map on morphisms that reverses the arrows. We will first construct this map in both directions; then we shall show it has an inverse, which will complete the proof of part 2.

Part 2a: Definition of $\cdot^*$ on morphisms. Now let $f : V \to W$ be a polynomial map; we wish to associate to it a homomorphism $f^* : K[W] \to K[V]$. that will map polynomial functions $g : W \to K$ to functions $f^*g : V \to K$. We have the following diagrams:

$$V \xrightarrow{\ f\ } W$$

$$\begin{array}{ccc} V & & W \\ & {\scriptstyle f^*g}\searrow & \downarrow{\scriptstyle g} \\ & & K \end{array} \qquad (4.7)$$

The obvious thing to do is to glue these diagrams together and define f^*g such that the triangle formed commutes; in other words, we define $f^*g := g \circ f$. The reader should check that f^*, defined in this way, is indeed a K-algebra homomorphism.

Part 2b: Definition of $\cdot_*$ on morphisms. Now, suppose $\varphi : K[W] \to K[V]$ is a K-algebra homomorphism; we wish to associate to it a polynomial map $\varphi_* : V \to W$. Suppose we have embedded $V \subseteq \mathbb{A}^m$ and $W \subseteq \mathbb{A}^n$.

Since each point $P \in W$ may be written as $(\pi_1, ..., \pi_n)(P)$, it is natural to consider the polynomial map $\varphi_* : V \to K^n$ given coordinate-wise by

$$\varphi_* = (\varphi(\pi_1), ..., \varphi(\pi_n)). \tag{4.8}$$

Note that each coordinate is a polynomial function from V into K, since each $\varphi(\pi_i)$ is a function in $K[V]$. To show that $\operatorname{im} \varphi_* \subseteq W$, it suffices to show that for all $g \in \mathcal{I}(W)$, $g \circ \varphi_* = 0$. Indeed, $[g] = [0] \in K[W]$; hence $\varphi([g]) = [0] \in K[V]$ (here we used that φ is a homomorphism), so if $P \in V$, $(g \circ \varphi_*)(P) = (\varphi(g))(P) = 0$, where the first equality comes from the definition of the π_i as restrictions of the X_i and the fact that φ is a K-algebra (so it fixes K):

$$g(\varphi_*(P)) = g(\varphi([X_1]), ..., \varphi([X_n]))(P) = \varphi(g(X_1, ..., X_n))(P) = \varphi(g)(P). \tag{4.9}$$

Thus φ_* is a polynomial map from V to W, as desired.

Part 2c: The maps $\cdot^*$ and $\cdot_*$ are inverses on morphisms. Now to complete the proof of part 2, we need to show that for all $f : V \to W$, $(f^*)_* = f$; and for all homomorphisms $g : K[W] \to K[V]$, $(g_*)^* = g$.

Fix $P \in V$; then $(f^*)(\pi_i) = \pi_i \circ f$ for all i, so $(f^*)_* = (\pi_1 \circ f, ..., \pi_n \circ f)$ and hence $((f^*)_*)(P) = (\pi_1 \circ f, ..., \pi_n \circ f)(P) = f(P)$.

On the other hand, fix $r \in k[W]$. Then $((g_*)^*)(r) = (g(\pi_1), ..., g(\pi_n))^*(r) = r \circ (g(\pi_1), ..., g(\pi_n)) = g(r)$ (where the final equality comes from a similar argument to equation 4.9 above, since g is a K-algebra homomorphism).

Part 3: The functor is contravariant. Let V_1, V_2, and V_3 be algebraic sets; let $g : V_1 \to V_2$ and $f : V_2 \to V_3$ be polynomial maps. We need to show that $(f \circ g)^* = g^* \circ f^*$. Indeed,

$$(f \circ g)^*(r) = r \circ (f \circ g) = f^*(r) \circ g = g^*(f^*(r)) = (g^* \circ f^*)(r) \tag{4.10}$$

for all $r \in K[V_3]$.

Now let $g : K[V_1] \to K[V_2]$ and $f : K[V_2] \to K[V_3]$; we need to show that $(f \circ g)_* = g_* \circ f_*$. Indeed,

$$(f \circ g)_* = ((f \circ g)(\pi_1), ..., (f \circ g)(\pi_n)) = g_*(f(\pi_1), ..., f(\pi_n)) = (g_* \circ f_*). \tag{4.11}$$

Part 1: The functor on objects. We finally need to show that $V^* \simeq W^*$ if and only if $V \simeq W$. Suppose that $V \simeq W$; then there is a bijective polynomial map $f : V \to W$. In particular, by parts 2 and 3 there exist homomorphisms $f^* : K[W] \to K[V]$ and $(f^{-1})^* : K[V] \to K[W]$ such that $f^* \circ (f^{-1})^* = (f^{-1} \circ f)^* = (\iota_V)^*$; the reader is invited to verify that this is the identity homomorphism on $K[V]$, and that the opposite composition $(f^{-1})^* \circ f^*$ is the identity on $K[W]$. Hence f^* is an isomorphism between $K[W]$ and $K[V]$.

Conversely, if $K[V] \simeq K[W]$ then there is a K-algebra isomorphism $f : K[V] \to K[W]$. By parts 2 and 3 there exist polynomial maps $f_* : W \to V$ and $f_*^{-1} : V \to V$ such that $f_* \circ f_*^{-1} = (f^{-1} \circ f)_* = (\iota_{K[V]})_*$; the reader is invited to check that this is the identity polynomial map on V; and that the opposite composition is the identity on W. Hence f_* is an isomorphism between W and V.

This concludes the proof of the theorem. ∎

In order to complete our description of this correspondence, we will describe the subcategory of algebras which the category of algebraic sets is dually equivalent to.

4.2.20 Definition. An algebra is called **reduced** if it has no non-zero nilpotents (definition 2.9.6). An **affine K-algebra** is a reduced finitely generated K-algebra.

4.2.21 Theorem. *The image of the functor $\cdot^*$ (that is, the subcategory* C *from the statement of theorem 4.2.18) is the category* $\mathtt{AffAlg}(K)$ *of affine K-algebras.*

Proof. Suppose V is an algebraic set. Then $K[V] \simeq K[X_1, ..., X_n]/\mathcal{I}(V)$ is finitely generated since the canonical morphism $K[X_1, ..., X_n] \to K[V]$ is surjective. Suppose $[f] \in K[X_1, ..., X_n]/\mathcal{I}(V)$ is such that $[f]^k = [0]$; then $f^k \in \mathcal{I}(V)$, so $f \in \mathcal{I}(V)$ since $\mathcal{I}(V)$ is radical; hence $[f] = [0]$, and all nilpotents are zero in $K[V]$.

Conversely, if A is an affine K-algebra then $A \simeq K[X_1, ..., X_n]/\mathfrak{a}$ for some ideal $\mathfrak{a}$. The ideal $\mathfrak{a}$ is finitely generated since $K[X_1, ..., X_n]$ is Noetherian; suppose $\mathfrak{a} = (f_1, ..., f_m)$. Consider $\mathcal{V}(\mathfrak{a})$; we shall show that $\mathcal{I}(\mathcal{V}(\mathfrak{a})) = \mathfrak{a}$ and so $A \simeq K[X_1, ..., X_n]/\mathcal{I}(\mathcal{V}(\mathfrak{a})) \simeq K[\mathcal{V}(\mathfrak{a})]$. Since $\mathcal{I}(\mathcal{V}(\mathfrak{a})) = \sqrt{\mathfrak{a}}$, if $f \in \mathcal{I}(\mathcal{V}(\mathfrak{a}))$ then $f^k \in \mathfrak{a}$ for some k. Hence $[f]^k = 0 \in A$; since A has no non-zero nilpotents, we must have $[f] = 0 \in A$, hence $f \in \mathfrak{a}$. Hence $\mathfrak{a} \supseteq \mathcal{I}(\mathcal{V}(\mathfrak{a}))$; and $\mathfrak{a} \subseteq \sqrt{\mathfrak{a}} = \mathcal{I}(\mathcal{V}(\mathfrak{a}))$; so we are done. ∎

The correspondence we have set up makes proving isomorphisms a purely algebraic matter.

4.2.22 Example. The parabola $\mathcal{P} = \mathcal{V}(X^2 - Y)$ is isomorphic to the x-axis, $\mathcal{X} = \mathcal{V}(Y)$. Indeed, $K[\mathcal{P}] \simeq K[X]$ (example 4.2.6); and $K[\mathcal{X}] = K[X, Y]/(Y) \simeq K[X]$; thus $K[\mathcal{P}] \simeq K[\mathcal{X}]$ and so by theorem 4.2.18 we are done.

4.2.23 Exercise.

1. Show that there is no isomorphism between $K[X]$ and $K[X, X^{-1}]$. Use example 4.2.6 and exercise 4.2.8 above to conclude that the parabola $\mathcal{V}(X^2 - Y)$ is not isomorphic to the hyperbola $\mathcal{V}(XY - 1)$.
2. On the other hand, show that the hyperbola $\mathcal{V}(X^2 - Y^2 - 1)$ *is* isomorphic to the circle $\mathcal{V}(X^2 + Y^2 - 1)$.
3. Finally, show that the two hyperbolae $\mathcal{V}(XY - 1)$ and $\mathcal{V}(X^2 - Y^2 - 1)$ are isomorphic.

Compare this with the classification of plane conics in section 3.1 above. In particular, use this exercise and the ideas of 4.2.16 above to show that in $\mathbb{R}^2$, the regular conics form three isomorphism classes (ellipses, hyperbolae, and parabolae); but over $\mathbb{C}^2$, there are only two isomorphism classes.

Note that even though coordinate rings will not have nilpotents, they still might have zero divisors. Example 4.2.6.4 above furnishes us with an example. More generally, if V is an algebraic set which is not irreducible then V will have zero divisors since $\mathcal{I}(V)$ will not be prime. This situation is not ideal (if you pardon the pun). Thus we make the fundamental definition of (classical) algebraic geometry:

4.2.24 Definition. An **affine variety** is an irreducible algebraic set.

We may also consider subsets of varieties:

4.2.25 Definition. An **affine subvariety** is a closed subset of an affine variety. A **quasi-affine variety** is an open subset of an affine variety.

4.2.26 Remark. We can extend the notion of isomorphism to arbitrary subvarieties and quasi-affine varieties as follows: if $X \subseteq V, Y \subseteq W$ are subsets of affine varieties V and W then a morphism $f : X \to Y$ is a polynomial map $f : V \to W$ such that $\operatorname{im} f\restriction_V \subseteq W$; the notion of an isomorphism is defined in the obvious way.

Being a variety is preserved (as one would hope) by isomorphism.

4.2.27 Proposition. *Let V and W be isomorphic algebraic sets. Then V is irreducible iff W is irreducible.*

Proof. V is irreducible iff $K[V]$ is an integral domain, and being an integral domain is preserved by isomorphism. ∎

4.2.28 Example. By example 4.2.13, the twisted cubic is isomorphic to the affine line; by example 2.8.3, if K is infinite then the affine line is irreducible; hence the twisted cubic is irreducible. Compare this with the original proof that the twisted cubic is irreducible (proposition 3.2.3).

Note that we may define a variety (or more generally, an algebraic set) without reference to any embedding in affine space; in particular, the properties of a variety are independent of the specific equations we use to define it. The definition would be as follows:

4.2.24′ Alternative definition. An **affine variety** is a set V together with an affine K-algebra $K[V]$ of functions $V \to K$ without zero divisors; an **embedding** of V in $\mathbb{A}^n$ is a choice of generators $\pi_1, ..., \pi_n$ of $K[V]$, and the zero-zet of V under the embedding is the set

$$\mathcal{V}(V) = \{(x_1, ..., x_n) \in \mathbb{A}^n : \exists_{P \in V} \forall_i x_i = \pi_i(P)\}. \tag{4.12}$$

The rigorous development and generalisation of this point of view is due in part to Grothendeick and Serre[2] and is known as the theory of **schemes**; we shall glimpse some small part of this very modern theory in the final chapter of these notes.

4.3 Rational maps

We have defined what it means for two algebraic sets to be related via some map; however, this notion is slightly too rigid for many examples. For example, in theorem 1.2.2 we saw that the circle (minus a point) is given as the image of the map

$$\lambda \mapsto \left(\frac{2r}{\lambda^2 + 1}, \frac{2\lambda r}{\lambda^2 + 1} \right). \tag{4.13}$$

which is not a polynomial map over $\mathbb{C}$ since $\lambda^2 + 1 = 0$ for some λ.

We need to allow functions that may have singularities. In order to do this, we want to follow the example of the complex analysts with their magical 'meromorphic functions' and look at functions of the form f/g for polynomials f and g; the natural construction for this is to form the field of fractions (proposition 2.1.18) of the coefficient ring. However, recall (remark 2.1.20) that in order to form this field we need the coefficient ring to be an integral domain. Hence we shall restrict ourselves to affine varieties.

4.3.1 Definition. Let V be an affine variety; then the **function field** of V, denoted by $K(V)$, is the field

$$K(V) := \operatorname{Frac} K[V]. \tag{4.14}$$

Members of $K(V)$ are of the form $[f]/[g]$ for $[f], [g] \in K[V]$ such that $[g] \neq [0]$ (i.e. $g \notin \mathcal{I}(V)$), modulo the equivalence relation $[f]/[g] = [f']/[g'] \iff [f][g'] - [f'][g] = [0]$ (i.e. $fg' - f'g \in \mathcal{I}(V)$). The members of $K(V)$ are called **rational functions** on V.

The **domain** of $\rho \in K(V)$ is the set

$$\operatorname{dom} \rho = \{P \in V : \exists_{[f],[g] \in K[V]} \rho = [f]/[g] \text{ and } g(P) \neq 0\}; \tag{4.15}$$

if $P \in \operatorname{dom} \rho$ then ρ is said to be **regular** at ρ.

[2] As in many parts of mathematics attributing a subject to a small group of people is often misleading; see, for example, [Die85] for a more balanced viewpoint.

We would like to treat the elements of $K(V)$ as functions which we can evaluate, defining $([f]/[g])(P)$ to be the value of $f(P)/g(P)$. However, there are some problems: firstly, this function will not be defined everywhere — only at points where $g(P) \neq 0$. We also need to check that if $[f]/[g] = [f']/[g']$ then $f(P)/g(P) = f'(P)/g'(P)$ wherever both sides of this latter equality are defined; this is not as trivial as it seems:

4.3.2 Example. Consider the cone $V = \mathcal{V}(XY - Z^2)$; this is irreducible, but still not particularly nice as $K[V] = K[X, Y, Y]/(XY - Z^2)$ is not a UFD and so the equality of rational functions $X/Z = Z/Y$ holds despite the functions on the right looking very different.

4.3.3 Proposition. *Let V be an affine variety and let $\rho \in K(V)$. The **ideal of denominators** of ρ,*

$$\mathfrak{d}_\rho := \{[g] \in K[V] : [g]\rho \in K[V]\}, \tag{4.16}$$

is an ideal of $K[V]$.

(Recall that we may embed $K[V]$ inside $K(V)$ in a natural way, so this definition does indeed make sense.)

Proof. Suppose $[g], [g'] \in \mathfrak{d}_\rho$ and $[h] \in K[V]$. If $[g]\rho \in K[V]$ and $[g']\rho \in K[V]$ then $([g] + [g'])\rho \in K[V]$; in addition, $[h]([g']\rho) \in K[V]$. ∎

4.3.4 Proposition. *Let V be an affine variety and let $\rho \in K(V)$. Then* $\operatorname{dom}\rho$ *is a dense open subset of V.*

Proof. Since $K[V]$ is Noetherian (corollary 4.2.4), $\mathfrak{d}_\rho = ([g_1], ..., [g_m])$ for $[g_1], ..., [g_m] \in K[V]$. Note that $\operatorname{dom}\rho = V \setminus \mathcal{V}(\mathfrak{d}_\rho)$, and $\mathcal{V}(\mathfrak{d}_\rho)$ is a closed subset of $\mathbb{A}^n$ which is not the full space since $\mathfrak{d}_\rho \neq (0)$ (indeed, $\rho = [f]/[g]$ for some nonzero $[g] \in K[V]$, so at the very least $[g] \in \mathfrak{d}_\rho$). Hence $\operatorname{dom}\rho$ is a non-empty open subset of V, and by proposition 2.6.6 this set is dense. ∎

The following theorem completes our discussion by telling us that two elements of $K(V)$ are equal in that ring iff their evaluations agree everywhere that they are both defined.

4.3.5 Theorem. *Let V be an affine variety, and let $[f]/[g]$ and $[f']/[g']$ be two elements of $K(V)$. If there exists some dense open subset $U \subseteq V$ such that $U \subseteq \operatorname{dom}[f]/[g]$, $U \subseteq \operatorname{dom}[f']/[g']$, and $f(P)/g(P) = f'(P)/g'(P)$ for all $P \in U$, then $[f]/[g] = [f']/[g']$ as rational functions.*

Proof. If $f(P)/g(P) = f'(P)/g'(P)$ for all $P \in U$, then the two polynomial functions fg' and $f'g$ agree on U. Thus by proposition 4.2.9, fg' and $f'g$ agree on V; hence $fg' - f'g = 0$ on the whole of V, and so the two rational functions agree on the entire intersection of their domains of definition. ∎

In particular, we may make the following definition without worry.

4.3.6 Definition. Let V be an affine variety; then for every point $P \in V$, define the homomorphism $\operatorname{eval}_P : K(V) \to K$ by $\operatorname{eval}_P([f]/[g]) := f(P)/g(P)$. We also define $\mathcal{V}(S)$ for $S \subseteq K(V)$ in the obvious fashion.

Finally, we shall check that the set of rational functions which is defined on the whole of a variety is just the coordinate ring. Since every member of the coordinate ring is a rational function regular everywhere, we only need to check the converse.

4.3.7 Proposition. *A rational function $\rho \in K(V)$ such that* $\operatorname{dom}\rho = V$ *is a polynomial function.*

Proof. Suppose $\rho \in K(V)$. The ideal of denominators $\mathfrak{d}_\rho$ (proposition 4.3.3) is finitely generated, say by $g_1, ..., g_n \in K[V]$. In particular, since these functions cannot have a common zero on V, by corollary 2.11.10 to the Nullstellensatz we can find $p_1, ..., p_n \in K[V]$ such that $\sum_{i=1}^n p_i g_i = 1$. Thus $f = \sum_{i=1}^n f p_i g_i$; and $g_i f \in K[V]$ for every i by definition of $\mathfrak{d}_\rho$, so we are done. ∎

In analogy with the polynomial case, we shall now define rational maps between varieties.

4.3.8 Definition. Let $V \subseteq \mathbb{A}^m$ and $W \subseteq \mathbb{A}^n$ be affine varieties. A **rational map** from V to W is a function φ from a dense open set $U \subseteq V$ to $\mathbb{A}^n$ such that

1. For each i, $\pi_i \circ \varphi : U \to K$ is a rational function on U; and
2. $\operatorname{im} \varphi \subseteq W$.

We will write $\varphi : V \dashrightarrow W$ to notate that φ is defined on a dense open subset of V, and we will write $\operatorname{dom} \varphi$ for the **domain of definition** (the set U); the image of φ is the set $\operatorname{im} \varphi = \varphi(\operatorname{dom} \varphi)$. We say that φ is **regular** at $P \in V$ if $P \in \operatorname{dom} \varphi$.

There is one small problem: if $\varphi : V_1 \dashrightarrow V_2$ and $\psi : V_2 \dashrightarrow V_3$ are rational maps, the domain of the composition $\psi \circ \varphi$ is only defined on the set $\varphi^{-1}(\operatorname{dom} \psi)$; and this set might be empty. In order to fix this problem, we will usually only consider rational maps whose images are dense open subsets of their codomain.

4.3.9 Definition. If $\varphi : V \to W$ is a rational map such that $\operatorname{im} \varphi$ is dense in W, then φ is called **dominant**.

4.3.10 Proposition. *If $\varphi : V_1 \dashrightarrow V_2$ and $\psi : V_2 \dashrightarrow V_3$ are rational maps and φ is dominant, then $\operatorname{dom}(\psi \circ \varphi) = \varphi^{-1}(\operatorname{dom} \psi)$ is non-empty.*

Proof. By density of $\operatorname{im} \varphi$, $\operatorname{dom} \psi \cap \operatorname{im} \varphi$ is non-empty. Hence $\varphi^{-1}(\operatorname{dom} \psi)$ is non-empty. ∎

4.3.11 Example. Let $V = \mathcal{V}\left((X-r)^2 + Y^2 - r^2\right)$; in the proof of theorem 1.2.2, we saw that the map

$$\varphi : (x, y) \mapsto \frac{y}{x} \tag{4.17}$$

is a map from the dense open subset $V \setminus (0,0)$ to the affine line $\mathbb{A}^1$, and so φ is a regular map $V \dashrightarrow \mathbb{A}^1$.

This map has another special property: as we showed in the referenced theorem, this map $U \to \mathbb{A}^1$ is bijective and its inverse (reproduced in equation 4.13) is also a rational map $\mathbb{A}^1 \to V$. By analogy with complex analysis we shall view such maps as 'almost isomorphisms'.

4.3.12 Definition. If V and W are affine varieties, a dominant rational map $\varphi : V \dashrightarrow W$ is called **birational** if there exists a dominant rational map $\psi : W \dashrightarrow V$ such that $\varphi \circ \psi = \iota_W$ and $\psi \circ \varphi = \iota_V$. The two varieties are called **birationally equivalent** if there is a birational map between them, and we will write $V \asymp W$ in this case.

4.3.13 Exercise. Adapt the proof of the circle parameterisation to show that every conic is birationally equivalent to $\mathbb{A}^1$.

4.3.14 Exercise. Decompose $\mathcal{V}(Y^2 - XZ, Z^2 - Y^3)$ into irreducible components and show that each component is birationally equivalent to $\mathbb{A}^1$.

4.3.15 Exercise. Show that there is a birational map φ from the conchoid of Dürer (exercise 3.3.9) onto $\mathbb{A}^1$ which is regular at all but two points, and that every point of $\mathbb{A}^1$ has precisely two preimages in the conchoid [Fet83]. Hints (illustrated by figure 4.1):

1. It suffices to prove the statement for $DC(X = 0, Y = 0, a, b)$.
2. Consider the set S which is used in the definition of the Dürer conchoid. For every pair $(P_1, P_2) \in S$, we have $P_1 = (0, b - t)$ and $P_2 = (t, 0)$ for some $t \in K$.

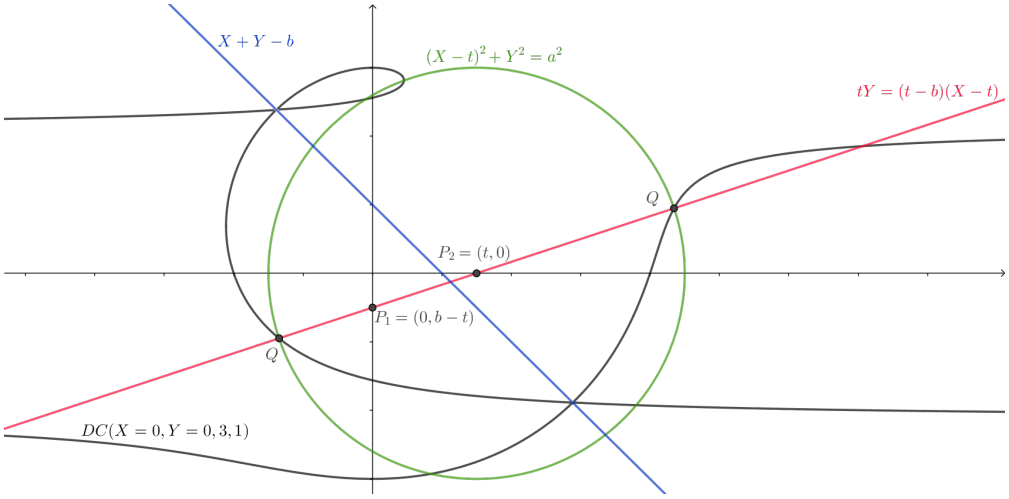


Figure 4.1: A conchoid of Dürer (exercise 4.3.15).

3. The points on the conchoid on the line $\overline{P_1 P_2}$ are precisely the endpoints of the segment with midpoint $(t, 0)$ and length $2a$. Hence the conchoid is the locus of the intersection points between the lines of the form $Yt = (t - b)(X - t)$ and circles of the form $(X - t)^2 + Y^2 = a^2$.

4. Use part 3 to show that for (x, y) on the conchoid then

$$t = \frac{x^2 + y^2 - bx - a^2}{x + y - b} \tag{4.18}$$

wherever the latter is defined.

5. Show that there are precisely two points where equation 4.18 is undefined.

Philosophically speaking, since the Dürer conchoid is irreducible, we have an example of a curve with two branches such that each branch is almost birationally equivalent to the affine line — the problem is that the two preimages of the line are algebraically indistinguishable so we cannot separate the function into two.

Question: does there exist a birational equivalence between the conchoid and $\mathbb{A}^1$?

4.3.16 Exercise. Is $\mathcal{V}(X^2 - Y^3)$ isomorphic to $\mathbb{A}^1$? Is it birationally equivalent to $\mathbb{A}^1$?

The previous few exercises are fairly important as they are indicative of a major theme in this subject: the idea of parameterisation. In particular, we have parameterised various curves using rational maps out of an affine line.

4.3.17 Definition. A variety V is called **rational** if $V \asymp \mathbb{A}^m$ for some m.

We shall prove later in this chapter (theorem 4.5.12) that every variety is at least birationally equivalent to a hypersurface of $\mathbb{A}^m$ for some m.

We will now check that rational maps are continuous. The bulk of the work comes from showing that rational *functions* are continuous, which is proved as the following lemma.

4.3.18 Lemma. *Let $\varphi : V \to K$ be a rational function. If $U \subseteq K$ is closed, then $\varphi^{-1}(U)$ is closed in* dom φ.

Proof. Since $K[X]$ is Noetherian, we may assume that $\mathcal{I}(U)$ is generated by finitely many polynomials; suppose $\mathcal{I}(U) = (u_1, ..., u_k)$. Let $P \in \operatorname{dom}\varphi$, and find $f, g \in K[X]$ such that $P \in \operatorname{dom} f/g$ and $f/g = \varphi$. Then $U_P := \operatorname{dom}(f/g) \cap \operatorname{dom}\varphi$ is an open neighbourhood of P on which the rational function f/g is defined and agrees with φ; by theorem 4.3.5, f/g agrees with φ everywhere that both are defined. Now,

$$Q \in (f/g)^{-1}(U) \iff (f/g)(Q) \in U = \mathcal{V}(u_1, ..., u_k) \iff \forall_{1\leq i\leq k}(u_i \circ (f/g))(Q) = 0. \tag{4.19}$$

In particular, $(f/g)^{-1}(U) \cap U_P = U_P \cap \mathcal{V}(\{u_i \circ f/g\}_{i=1}^k)$. But each $u_i \circ f/g$ is a rational function with a domain including the set U_P; say $u_i \circ f/g = f_i'/g_i'$ on this set. Since the zeros of f_i'/g_i' in its domain are precisely the zeros of $f_i' \in K[X]$ restricted to that domain, we have that $\mathcal{V}(\{u_i \circ f/g\}_{i=1}^k) = \mathcal{V}(\{f_i'\}_{i=1}^k)$; and so $(f/g)^{-1}(U) \cap U_P = U_P \cap \mathcal{V}(\{f_i'\}_{i=1}^k)$.

Thus $\varphi^{-1}(U)$ is closed in each U_P; more explicitly, for each $P \in \operatorname{dom}\varphi$ there exists some closed set V_P of $\operatorname{dom}\varphi$ such that $\varphi^{-1}(U) \cap U_P = U_P \cap V_P$. The open sets U_P cover $\operatorname{dom}\varphi$, and so (by exercise 2.6.28) we can choose finitely many points of $\operatorname{dom}\varphi$, say $P_1, ..., P_m$, such that $\{U_{P_i}\}_{i=1}^m$ is an open cover of $\operatorname{dom} f$. We can therefore write

$$\varphi^{-1}(U) = \cup_{i=1}^m(\varphi^{-1}(U) \cap U_{P_i}) = \cup_{i=1}^m(U_{P_i} \cap V_{P_i}) = \operatorname{dom} f \cap \left(\cup_{i=1}^m V_{P_i}\right); \tag{4.20}$$

and a finite union of closed sets is closed, so $\varphi^{-1}(U)$ is closed in $\operatorname{dom} f$. ∎

4.3.19 Theorem. *Let $\varphi : V \dashrightarrow W$ be a rational map. If $U \subseteq W$ is closed, then $\varphi^{-1}(U)$ is closed in* $\operatorname{dom}\varphi$.

Proof. Suppose $W \subseteq \mathbb{A}^n$, $U \subseteq W$, and $U = W \cap \mathcal{V}(u_1, ..., u_k)$ for polynomials $u_i \in K[X_1, ..., X_n]$. Each $u_i \circ \varphi$ is a rational map $V \dashrightarrow K$, and so is continuous by the previous lemma. In particular,

$$\varphi^{-1}(U) = \varphi^{-1}(\mathcal{V}(u_1, ..., u_k)) = \cap_{i=1}^k \mathcal{V}((\alpha_i \circ f)^{-1}) \tag{4.21}$$

is an intersection of closed sets and is therefore closed. ∎

4.3.20 Corollary. *Birational maps are homeomorphisms.* ∎

4.3.21 Corollary. *Polynomial maps are continuous.* ∎

The reader is now invited to prove an analogue of theorems 4.2.18 and 4.2.21 for dominant rational maps:

4.3.22 Theorem. *The category of affine varieties over K with dominant rational maps as morphisms is dually equivalent to the category of function fields over K.*

More explicitly,

1. *For every affine variety V there is a unique function field $K(V)$ such that $W \asymp V$ iff $K(W) \simeq K(V)$.*
2. *For every dominant rational map $\varphi : V \dashrightarrow W$ there is a unique K-homomorphism $\varphi^* : K(W) \to K(V)$;*
3. *For every K-homomorphism $\varphi : K(W) \to K(V)$ there is a unique dominant rational map $\varphi_* : V \dashrightarrow W$;*
4. *$(\varphi\circ\psi)^* = \psi^*\circ\varphi^*$ and $(\varphi\circ\psi)_* = \psi_*\circ\varphi_*$.* ∎

4.3.23 Corollary. *Two varieties W and V are birationally equivalent iff $K(W) \simeq K(V)$.* ∎

We will prove the intuitive result that two varieties are birationally equivalent iff they are isomorphic 'almost everywhere' — think about the circle minus a point and the affine line, for example: we wrap the affine line around, and there is only a single point missing.

4.3.24 Proposition. *Let V and V' be varieties; then $V \asymp V'$ iff there exist dense open subsets $U \subseteq V$ and $U' \subseteq V'$ such that $U \simeq U'$ (recall that a definition of morphisms of subsets of varieties was given in remark 4.2.26).*

Proof. Suppose $\varphi : V \to V'$ is birational. Then $\operatorname{im}\varphi \subset V'$ and $\operatorname{im}\varphi^{-1} \subset V$ are open since φ is a homeomorphism. Since φ is regular on its image, it is a polynomial map on its image; the same is true for φ^{-1}. Thus φ is an isomorphism between $\operatorname{im}\varphi^{-1}$ and $\operatorname{im}\varphi$.

Conversely, if $U \simeq U'$ for dense open subsets $U \subset V$ and $U' \subset V'$, then there is an isomorphism $f : U \to U'$; for every $(x_1, ..., x_n) \in V \setminus U$ let $f_P = f\frac{(X_1-x_1)\cdots(X_n-x_n)}{(X_1-x_1)\cdots(X_n-x_n)}$; then $f_P \in K(V)$, and all the f_P agree on a dense open subset (namely U) so all agree everywhere they are defined. In particular we have constructed a dominant rational map $V \dashrightarrow V'$. We may extend f^{-1} to a dominant rational map on V' in the same way; and the rational maps thus constructed are inverses on U and U'. ∎

4.4 Introduction to dimension theory

In linear algebra, the 'dimension' of a space is a very rigid invariant that is preserved or transformed in very predictable ways by homomorphisms (linear transformations): indeed, every finite dimensional vector space over some field is isomorphic to all the others with the same dimension.

In this subject, our objects are not necessarily linear; but they are still defined by polynomials over some field. We would therefore like to describe something that feels a bit like dimension. We begin with two special cases. Firstly, subsets of $\mathbb{A}^1$:

4.4.1 Example. Exercise 2.5.8 claims that the closed sets of $\mathbb{A}^1$ are precisely $\mathbb{A}^1$ itself and finite sets of points. The first of these, the entire affine line, feels like an algebraic set of dimension 1; the second of these, finite sets of points, feel like sets of dimension 0.

We now consider subsets of $\mathbb{A}^2$:

4.4.2 Proposition. *Let $X \subseteq \mathbb{A}^2$ be an algebraic set. Then precisely one of the following holds:*

1. *X is a finite set of points.*
2. *X is the union of a plane algebraic curve and a finite set of points.*
3. *X is the whole plane $\mathbb{A}^2$.*

These correspond intuitively to sets which should have respective dimensions of 0, 1, and 2.

Proof. Suppose $X = \mathcal{V}(f_1, ..., f_n)$. If all the f_i are identically zero then $X = \mathbb{A}^2$ and so we are in case 3; so suppose this is not the case.

Since $K[X, Y]$ is a UFD, there exists a polynomial g of maximal degree such that $g \mid f_i$ for all i. Suppose g is a constant, and suppose there are infinitely many points $P_\alpha \in \mathbb{A}^2$ such that $P_\alpha \in X$. Then by remark 2.8.2, since $f_i(P_\alpha) = 0$ for all i and all P_α, each f_i is identically zero. Thus X is a finite set of points, and we are in case 1.

Finally, if g is non-constant then we may write $X = \mathcal{V}(g) \cup \mathcal{V}(f_1/g, ..., f_n/g)$; by the argument in the previous paragraph $\mathcal{V}(f_1/g, ..., f_n/g)$ is a finite set of points, and $\mathcal{V}(g)$ is an plane algebraic curve; so we are in case 2, and have exhaused all possibilities. ∎

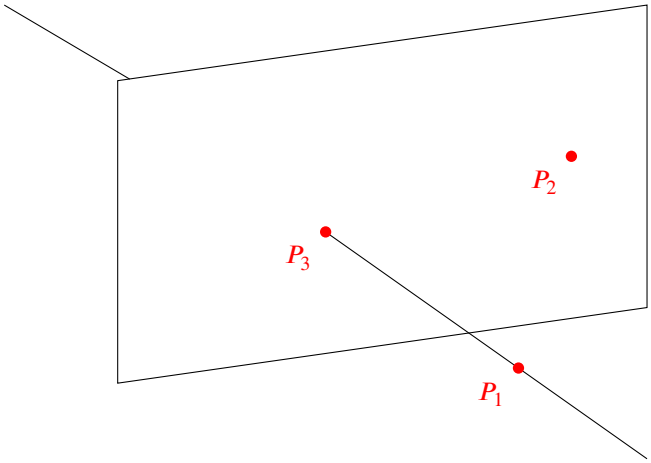


Figure 4.2: A depiction of dimension at a point.

Note that in each case we may write a chain of proper inclusions of irreducible subsets: in the case of $\mathbb{A}^1$, we may write $\mathbb{A}^1 \supset \{0\}$; and in the case of $\mathbb{A}^2$, we may write $\mathbb{A}^2 \supset \mathcal{V}(Y) \supset \{(0,0)\}$. This motivates the following set of definitions:-

4.4.3 Definition. Let $X_0 \supset X_1 \supset \cdots X_n$ be a proper chain of inclusion. The **chain length** is defined to be n (the number of inclusions).

4.4.4 Definition. If V is an affine variety, then the **dimension** of V, denoted $\dim V$, is the supremum

$$\dim V := \sup\{n \in \mathbb{N}_0 : n \text{ is the length of a chain of irreducible subsets } V \supset V_1 \supset \cdots \supset V_n\}. \quad (4.22)$$

If X is an arbitrary affine algebraic set, and $P \in X$, then the **dimension of X at P**, denoted $\dim_P X$, is the number

$$\dim_P X := \sup\{\dim V : V \text{ an irreducible subset of } X \text{ containing } P\}. \quad (4.23)$$

The dimension of X itself, $\dim X$, is simply $\sup\{\dim_P X : P \in X\}$.

4.4.5 Example. Here is the standard intuitive example. Consider the algebraic set X depicted in figure 4.2; then we have $\dim_{P_1} X = 1$, $\dim_{P_2} X = 2$, and $\dim_{P_3} X = 2$.

Clearly by the decomposition theorem, theorem 2.6.27, the supremums in the definitions of $\dim X$ and $\dim_P X$ are actually maximums and so it suffices to show that $\dim V$ is finite for affine varieties V. The proof of this will require us to understand some more of the theory of field extensions.

4.4.6 Remark. The theory of dimension is a major subfield of commutative algebra. A very good introductory reference is [Eis95, part II]; we shall primarily follow the proofs in chapter 13 of that book, though our treatment of transcendence bases will follow [Lan71, chapter X].

First, we shall transform our geometric definition into the language of algebra.

4.4.7 Definition. Let R be an integral domain. The **Krull dimension** of R, denoted $\dim R$, is the supremum of the lengths of ascending chains of distinct prime ideals from the zero ideal in R; i.e. the supremum of the lengths of prime ideal chains of the form $(0) \subset P_1 \subset \cdots$. The Krull dimension is also called the **height** of R.

4.4.8 Proposition. *If V is an affine variety, then* $\dim V = \dim K[V]$.

Proof. We will show that there is a length-preserving bijection between descending chains of irreducibles in V and ascending chains of prime ideals in $K[V]$.

Let $V \supset V_1 \supset \cdots$ be a descending chain of irreducibles in V. Then we can construct, by the $\mathcal{I}$-$\mathcal{V}$ correspondence, an ascending chain of prime ideals $\mathcal{I}(V) \subset \mathcal{I}(V_1) \subset \cdots$ in $K[X_1, ..., X_n]$[3] and this clearly has the same length. Then by the prime correspondence theorem (proposition 2.1.16) we have a chain of prime ideals $(0) \subset \mathcal{I}(V_1)/\mathcal{I}(V) \subset \cdots$ in $K[V]$.

Conversely, every chain of prime ideals $(0) \subset \mathfrak{p}_1 \subset \cdots$ corresponds bijectively to a chain of prime ideals in $K[X_1, ..., X_1]$ which corresponds bijectively to a descending chain of irreducibles from V. ∎

It will turn out that computing the dimension of a ring is difficult in general. To show that it is at least finite in this case will require a good deal of machinery. The setup for our main theorem is as follows.

4.4.9 Definition. Let $L : K$ be a field extension, and let $\alpha_1, ..., \alpha_n \in L$ be elements in the extension. We say that the elements are **algebraically independent** over K if there is no nonzero polynomial $f \in K[X_1, ..., X_n]$ such that $f(\alpha_1, ..., \alpha_n) = 0$.

The set $S = \{\alpha_1, ..., \alpha_n\}$ is called a **transcendence basis** for L over K if S is a maximal (under inclusion) algebraically independent set and L is algebraic over $K(\alpha_1, ..., \alpha_n)$.

Our main example is the following.

4.4.10 Example. Let $K : L$ be a finitely generated field extension (i.e. $K = L(\alpha_1, ..., \alpha_n)$). Then some subset of $\{\alpha_1, ..., \alpha_n\}$ is a transcendence basis for L over K. Indeed, if some α_i is algebraically dependent on the others then we can throw it away; iterating this process will produce a transcendence basis.

4.4.11 Proposition. *If $\{\alpha_i\}_{i=1}^m$ and $\{\beta_j\}_{j=1}^n$ are transcendence bases for L over K, then $m = n$. This number is called the* ***transcendence degree*** *of L over K, and is denoted by* $\operatorname{trdeg}_{L:K}$.

Proof. Let $\{\alpha_i\}_{i=1}^m$ be a transcendence basis; suffices to show that if $\{\beta_j\}_{j=1}^n$ is any set of algebraically independent elements then $n \leq m$. Since $\{\alpha_i\}$ is maximal, there exists a polynomial $p_1 \in K[X_0, ..., X_m]$ such that $p_1(\beta_1, \alpha_1, ..., \alpha_m) = 0$. The indeterminate X_0 must appear in p_1, since otherwise p_1 would be a polynomial in m variables over K such that $(\alpha_1, ..., \alpha_m)$ is a root in L. Similarly, at least one of the indeterminates $X_1, ..., X_m$ must appear. By reordering, we may assume that X_1 appears. Hence α_1 (and hence L) is algebraic over $K(\beta_1, \alpha_1, ..., \alpha_m)$.

We now proceed inductively. Suppose that after a suitable reordering of $X_2, ..., X_m$ and $\alpha_2, ..., \alpha_m$ we have found $\beta_1, ..., \beta_r$ $(r < m)$ such that L is algebraic over $K(\beta_1, ..., \beta_r, \alpha_{r+1}, ..., \alpha_m)$. Then there is a nonzero polynomial $p_{r+1} \in K[X_0, ..., X_m]$ such that $p_{r+1}(\beta_{r+1}, \beta_1, ..., \beta_r, \alpha_{r+1}, ..., \alpha_m) = 0$; in addition, X_0 must appear in p_{r+1}, as must (after a suitable reordering) X_{r+1}; hence α_{r+1} is algebraic over $K(\beta_1, ..., \beta_{r+1}, \alpha_{r+2}, ..., \alpha_m)$; and so L is algebraic over this set.

Repeating this procedure, if $n \geq m$ we may replace each α_i with some β_j to see that L is algebraic over $K(\beta_1, ..., \beta_m)$; in particular, $\beta_{m+1}, ..., \beta_n$ cannot be algebraically independent of $\beta_1, ..., \beta_m$; to avoid a contradiction we must have $m = n$. Thus it is impossible to have $n > m$ and we are done. ∎

4.4.12 Definition. If A is a reduced K-algebra, then $\operatorname{trdeg}_{A:K}$ is defined to be $\operatorname{trdeg}_{(\operatorname{Frac} A):K}$.

4.4.13 Example. Let A be an affine reduced K-algebra; suppose A is generated by $\pi_1, ..., \pi_n$. Then $\operatorname{trdeg}_{A:K} \leq n$, by example 4.4.10 above.

4.4.14 Remark. The theory of transcendence bases may be generalised to infinite bases; then the above proposition still holds, in the sense that all transcendence bases have the same *cardinality*. The relevant results may be found in [Nag77, chapter 4]; in our setting, all the bases will be finite and so luckily we need not dirty ourselves unnecessarily in the realm of set theory.

[3] Recall that we have a standing assumption (assumption 4.0.1) that the base field is algebraically closed.

Note the clear analogy with the theory of bases of vector spaces — compare the proof of the previous proposition with that of theorem 8.1 of [Hal15a], for example. In the linear case, the size of a basis was what we used to define dimension; in analogy with that case, we would expect the transcendence basis to give us something 'like' dimension. And indeed, our main theorem is as follows:

4.4.15 Theorem (Main dimension theorem)**.** *If A is an affine reduced K-algebra, then* $\dim A = \mathrm{trdeg}_{A:K}$.

References to proof. We omit the proof, as it requires commutative-algebraic technicalities that would detract from the main theme of the theory. It may be found in [Eis95, pp. 281–291] (it is 'theorem A', stated on page 221). An alternative presentation of the same proof may be found in [Nag77, theorem 4.9.4]; this latter treatment has the advantage of including the majority of the field-theoretic prerequisites. The standard reference [AM69] also includes a proof (theorem 11.25) but the crucial step depends on an exercise (exercise 5.16) which is worked out in detail in the other texts.

The reader who wishes to read from Eisenbud (recommended) will need to read the following results in order: lemma 13.2(c) (note that a 'finitely generated module' is what we called a 'linearly generated algebra'); theorem 13.3; proposition 13.10; (which comes within the proof of) theorem 13.9; finally the *proof of theorem A* given on page 289 (the statement is on page 286).

Proposition 13.10 uses the language of Galois theory; the main definition required is as follows.

Definition. Let $L : K$ be a field extension. The **Galois group** of the extension, denoted $\Gamma(L : K)$, is the group of automorphisms of L (that is, isomorphisms $L \to L$) such that $\phi \in \Gamma(L : K)$ iff $\forall_{l \in L} \phi(l) = l$.

The reader should also be familiar with Zorn's lemma; see, for example, [Mun00, §9] or [Hal15b, §16]. ∎

4.4.16 Corollary. *If V is an affine variety, then* $\dim V$ *is a finite integer.*

Proof. Indeed, $\mathrm{trdeg}_{K[V]:K}$ is finite by example 4.4.13 above, since $K[V]$ is finitely generated and reduced over K; then $\dim K[V] = \mathrm{trdeg}\, K[V] : K$ by theorem 4.4.15, so $\dim K[V]$ is finite; and $\dim V = \dim K[V]$ by proposition 4.4.8 above. ∎

Let us check that various things do indeed work, by working out some easy examples.

4.4.17 Example.

1. $\dim \mathbb{A}^n = n$, since $\mathrm{trdeg}_{K(\mathbb{A}^n):K} = n$. (Indeed, $K(\mathbb{A}^n) = K([X_1], ..., [X_n])$, and the X_i are independent over K.)

2. An irreducible plane curve has dimension 1; indeed, if $V = \mathcal{V}(f)$ for some irreducible $f \in K[X, Y]$ then $K[V] \simeq K[X, Y]/(f)$. Since $f(X, Y) = 0$, $X + (f)$ and $Y + (f)$ are not algebraically independent in $K(V)$; hence $\mathrm{trdeg}_{K[V]:K} \leq 1$. On the other hand, $X + (f)$ is transcendental over K, since if there existed $g \in K[Z]$ such that $g(X + (f)) = 0$ then $g(X) + (f) = 0 \in K[V]$ and so $g(X) \mid f$; so either g is constant (impossible, since it has a root), or f is reducible (impossible by assumption). Thus $\mathrm{trdeg}_{K[V]:K} \geq 1$ and so $\dim V = \mathrm{trdeg}_{K[V]:K} = 1$.

3. If $P \in \mathbb{A}^n$ then $\dim P = 0$ since $\mathcal{I}(P)$ is maximal.

4.4.18 Exercise.

1. Show that $X \subseteq \mathbb{A}^n$ is finite if and only if $\dim X = 0$.

2. If C is a regular conic over K^2, show that $\dim C = 1$.

3. What is the dimension of the twisted cubic? (Proposition 4.4.19 below makes this trivial; try not to use it.)

The dimension is a birational invariant.

4.4.19 Proposition. *If $V \asymp W$ then* $\dim V = \dim W$.

Proof. If $V \asymp W$ then $K(V) \simeq K(W)$ (corollary 4.3.23); hence $\mathrm{trdeg}_{K(V):K} = \mathrm{trdeg}_{K(W):K}$, and we are done. ∎

4.4.20 Corollary. *$\mathbb{A}^n \asymp \mathbb{A}^m$ iff $n = m$.*

Proof. Clearly if $n = m$ then $\mathbb{A}^n \asymp \mathbb{A}^m$. Conversely, if $n \neq m$ then $\dim \mathbb{A}^m = m \neq n = \dim \mathbb{A}^n$ (example 4.4.17) and so $V \not\asymp W$ by the proposition. ∎

We now describe how dimension acts with respect to subvarieties.

4.4.21 Theorem. *Let V be a variety, and let $W \subseteq V$ be a subvariety. Then* $\dim W \leq \dim V$. *If* $\dim W = \dim V$, *then $W = V$.*

Proof. Suppose $W \subseteq V \subseteq \mathbb{A}^m$, and set $n = \dim V$. Then any $n+1$ elements of $K(V)$ are algebraically dependent over K; by the inclusion-reversing nature of the function field correspondence (theorem 4.3.22), these $n+1$ elements lie in $K(W)$ too, so $\dim W = \mathrm{trdeg}_{K(W):\mathbb{A}^n} < n+1$; i.e. $\dim W \leq n = \dim V$.

Now suppose $\dim V = \dim W$; call this dimension n. There exists a transcendence basis for $K(V)$ over K consisting of generators of V, say $\pi_1, ..., \pi_n$. These elements all lie in $K(W)$, and they must be algebraically independent in $K(W)$ as well. Hence $\pi_1, ..., \pi_n$ also forms a transcendence basis of $K(V)$.

We will now show that $\mathcal{I}(W) \subseteq \mathcal{I}(V)$; this will complete the proof of the equality. Suppose $f \notin \mathcal{I}(V)$. Since $f \in K(V)$ is algebraically dependent on $\pi_1, ..., \pi_n$, there is a polynomial $p \in K(\pi_1, ..., \pi_n)$ such that $p(f) = 0$. We may therefore write

$$0 = p_0(\pi_1, ..., \pi_n) + \cdots + p_k(\pi_1, ..., \pi_n) f^k \tag{4.24}$$

for polynomials $p_i \in K[X_1, ..., X_n]$ (here we are using lemma 2.10.14). Since $K[X_1, ..., X_n]$ is a UFD, we may arrange for p to be irreducible (in particular, $p_0(\pi_1, ..., \pi_n)$ is non-zero in $K(V)$); and since $\pi_1, ..., \pi_n$ are a transcendence basis for $K(W)$, the relation 4.4 holds in $K(W)$ as well. Suppose $f \in \mathcal{I}(W)$; then $f(w) = 0$ for all $w \in W$, and so by relation we must have that $[p_0] = [0]$ in $K(W)$. Since the π_i are algebraically independent in $K(W)$, this implies that p_0 must be the zero polynomial over K; i.e. p_0 is zero on the whole of $\mathbb{A}^n$, contradicting that $p_0 \notin \mathcal{I}(V)$. Hence if $f \notin \mathcal{I}(V)$ then $f \notin \mathcal{I}(W)$, so $\mathcal{I}(W) \subseteq \mathcal{I}(V)$ and we are done. ∎

We can use dimension theory to formalise the idea that a hypersurface is a 'maximal subvariety' of $\mathbb{A}^n$.

4.4.22 Definition. Let V be an algebraic set and let $W \subseteq V$ be closed. Then the **codimension** of W in V is the integer $\mathrm{codim}_V W = \dim V - \dim W$.

4.4.23 Theorem. *If V is a hypersurface in $\mathbb{A}^n$, then every irreducible component of V has* $\mathrm{codim}_{\mathbb{A}^n} V = n - 1$.

Proof. Suppose $V = \mathcal{V}(f)$ for $f \in K[X_1, ..., X_n]$, and f is factorised into irreducibles as $f = f_1 \cdots f_k$. By corollary 2.11.13 we may decompose V as a union of irreducible components given by $\mathcal{V}(f) = \mathcal{V}(f_1) \cup \cdots \cup \mathcal{V}(f_k)$. For each i, f_i is nonzero and so there is some indeterminate X_j which actually appears in the polynomial expression of f. By reordering if necessary, suppose $j = n$. We claim that $\{\pi_1, ..., \pi_{n-1}\}$ is algebraically independent in $K(\mathcal{V}(f_i))$. Indeed, if $g \in K[X_1, ..., X_{n-1}]$ satisfies $g(\pi_1, ..., \pi_{n-1}) = 0$ then $g \in \mathcal{I}(\mathcal{V}(f_i)) = (f_i)$; i.e. $f_i \mid g$. But this is impossible since g does not involve the indeterminate X_n and f_i does; it follows that $\dim V \geq n - 1$ and so since $V \neq \mathbb{A}^n$ we may apply theorem 4.4.21 above to conclude that $\dim V = n - 1$. ∎

4.5 Classification of hypersurfaces

We will prove the classification theorem for varieties which we have been promising. We will need some (more!) technical results about field extensions.

4.5.1 Definition. Let $L : K$ be an algebraic field extension. For every $l \in L$, a **minimal polynomial** for l over K is a polynomial $p \in K[X]$ of minimal degree such that $p(l) = 0$ in L. If $\mu_l p = 1$ then we say that l is **separable** over K (the reader should quickly verify that this definition makes sense: $\mu_l p = \mu_l q$ for all minimal polynomials $p, q \in K[X]$ for l over K). The extension $L : K$ as a whole is called **separable** if every element of L is separable over K.

4.5.2 Exercise. Let $L : K$ be an algebraic field extension, and let $l \in L$ be arbitrary.

1. The element l has a minimal polynomial p over K.
2. The polynomial p is irreducible over K; hence $(p) = \sqrt{(p)}$ as ideals in $K[X]$.
3. If $f \in K[X]$ is such that $f(l) = 0$, we have $p \mid f$.

4.5.3 Exercise. Check that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is separable over $\mathbb{Q}$.

4.5.4 Theorem (Primitive element theorem, classical version)**.** *Let $L = K(\alpha_1, ..., \alpha_n)$ be a separable field extension of K. Then $L = K(\theta)$ for some $\theta \in L$. (Such a θ is called a **primitive element** for L over K.)*

We shall follow the proof given by van der Waerden in [Wae49, §40] (but with some simplification as K is algebraically closed).

Proof. We shall prove this by induction on n. If $n = 1$ then there is nothing to prove. The inductive step is the following claim:

$$\textit{If } L = K(x, y) \textit{ and } y \textit{ is separable over } K \textit{ then there exists } \theta \in L \textit{ such that } L = K(\theta). \tag{4.25}$$

Indeed, once we prove this then the induction will work as follows: if $L = K(\alpha_1, ..., \alpha_n)$ then $L = K(\alpha_1, .., \alpha_{n-1})(\alpha_n)$; by induction there exists some $\theta' \in K(\alpha_1, .., \alpha_{n-1})$ such that $L = K(\theta')(\alpha_n)$; and we can apply claim (4.25) to conclude the existence of a primitive element for L over K.

Let f and g be minimal polynomials for x and y respectively over K factorise (up to a unit) $f = (X - x_1) \cdots (X - x_r)$ and $g = (X - y_1) \cdots (X - y_s)$. By reordering, we may assume that $x = x_1$ and $y = y_1$. Since y is separable over K, $y_i = y$ iff $i = 1$. Thus for all $i \neq 1$, and all j $(1 \leq j \leq r)$, there exists at most one $z_{i,j} \in K$ such that

$$x_i + z_{i,j} y_j = x + z_{i,j} y. \tag{4.26}$$

Since K is infinite, there exists some $z \in K$ such that $z \neq z_{i,j}$ for all $z_{i,j}$. Let $\theta = x + zy \in L$.

We will be done if we show that $K(x, y) = K(\theta)$. By definition of θ we have

$$g(y) = 0 \text{ and } f(\theta - zy) = f(x) = 0; \tag{4.27}$$

these are polynomial relations with coefficients in $K(\theta)$ (since $f, g \in K[X]$). We claim that y is the *only* common root of the polynomials $g, f(\theta - zY) \in K(\theta)[Y]$ in L; indeed, the roots of g are at most $y_1, ..., y_n$, and y_i for $i \neq 1$ is not a root of the second equation since $f(\theta - zy_i) = 0$ iff $\theta - zy_i = x_j$ for some j; i.e. $x_j + zy_i = \theta = x + zy$; but $z_{i,j}$ is the only value of Z for which $x_j + Zy_i = x + Zy$, and $z \neq z_{i,j}$.

By separability, y has multiplicity 1 as a root of g. Hence $(Y - y)$ is a common factor of highest degree of g and $f(\theta - zY)$. The coefficients of $(Y - y)$ thus lie in $K(y)$ and in particular $y \in K(\theta)$. Since $x = \theta - zy$, $x \in K(\theta)$; so $K(x, y) \subseteq K(\theta)$ which is sufficient to complete the proof. ∎

4.5.5 Remark. The modern version of the primitive element theorem, formulated by Emil Artin, is [Lan71, theorem VII.14].

4.5.6 Exercise. Check that the above proof only requires that $K(\alpha_1, ..., \alpha_n)$ be separable over $K(\alpha_1)$ (i.e. we proved something a little stronger than the statement claims).

4.5.7 Exercise. Find $\theta \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$ such that $\mathbb{Q}(\theta) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

We will now define an operation which looks like the analytic concept of differentiation; the main purpose of it at this stage is to allow us to look at polynomial multiplicity in a somewhat elegant fashion.

4.5.8 Definition. Suppose R is an integral domain, and let $f \in R[X]$ be given by $\sum_{i=0}^{n} f_i X^n$. Then the **algebraic derivative** of f is the polynomial $D_X f := \sum_{i=0}^{n-1} n f_i X^{n-1}$.

4.5.9 Exercise. Show using formal algebraic calculations in $R[X]$ that:

1. if $f, g \in R[X]$ then $D_X(fg) = (D_X f)g + f(D_X g)$;
2. if $r \in R$ and $n \in \mathbb{N}$ then $D_X(X - r)^n = n(X - r)^{n-1}$;
3. if $f \in R[X]$, then $\partial(D_X f) < \partial f$. Is it necessarily true that $\partial(D_X f) = \partial f - 1$?

4.5.10 Lemma. *Suppose $f \in R[X]$ for an integral domain R. If $r \in R$ and $\mu_r f > 1$ then $(D_X f)(r) = 0$.*

Proof. If $\mu_r f > 1$ then we can write $f = (X - r)^n g$ for some $n > 1$ and some $g \in R[X]$ such that $g(r) \neq 0$. Then $D_X f = n(X - r)^{n-1} g + (X - r)^n (D_X g) = (X - r)^{n-1}(n + (X - r))(D_X g)$, so $(D_X f)(r) = 0$ (since $n - 1 \geq 1$). ∎

4.5.11 Lemma. *Let $L : K$ be an algebraic field extension. Then L is separable over K.*

Proof. Suppose $l \in L$; let $p \in K[X]$ such that $\mu_l p > 1$. Then $(D_X p)(l) = 0$ by the previous lemma; and by exercise 4.5.9, $\partial(D_X p) < \partial p$. Thus p cannot be a minimal polynomial for l. Let $f \in K[X]$ be the minimal polynomial for l over K; we must have $\mu_l f \geq 1$, and so $\mu_l f = 1$. Hence l is separable over K). ∎

4.5.12 Theorem. *If $V \subseteq \mathbb{A}^n$ is a variety of dimension d, then V is birationally equivalent to a hypersurface in $\mathbb{A}^{d+1}$.*

Proof. By the discussion of dimension we may write $K(V) \simeq K(\pi_1, ..., \pi_d)(\pi_{d+1}, ..., \pi_n)$ after some reordering of the coordinate functions, where $\pi_1, ..., \pi_d$ are transcendental over K and $\pi_{d+1}, ..., \pi_n$ are algebraic. By lemma 4.5.11, $K(\pi_1, ..., \pi_d)(\pi_{d+1}, ..., \pi_n)$ is separable over $K(\pi_1, ..., \pi_d)$. By the primitive element theorem, theorem 4.5.4, there exists some $\theta \in K(\pi_1, ..., \pi_d)(\pi_{d+1}, ..., \pi_n)$ such that $K(\pi_1, ..., \pi_d)(\pi_{d+1}, ..., \pi_n) = K(\pi_1, ..., \pi_d, \theta)$.

Let the minimal polynomial for θ over $K(\pi_1, ..., \pi_d)$ be $f \in K(\pi_1, ..., \pi_d)[X_{d+1}]$. Since $K(\pi_1, ..., \pi_d) \simeq K(X_1, ..., X_d)$ we may (after clearing denominators) view f as a polynomial in $K[X_1, ..., X_{d+1}]$. Define $W \subseteq \mathbb{A}^{d+1}$ by $W = \mathcal{V}(f)$.

Now we claim that $\frac{K[X_1,...,X_{d+1}]}{(f)} \simeq K[\pi_1, ..., \pi_d, \theta]$. Indeed, consider the evaluation homomorphism $\varphi := \mathrm{eval}_{(\pi_1,...,\pi_d,\theta)} : K[X_1, ..., X_{d+1}] \to K[\pi_1, ..., \pi_d, \theta]$. We may decompose this homomorphism as follows:

$$\begin{array}{ccccc} K[X_1, ..., X_{d+1}] & & & & \\ & \searrow^{\mathrm{eval}_{(\pi_1,...,\pi_d,X_{d+1})}} & & & \\ & & K[\pi_1, ..., \pi_d][X_{d+1}] & & \\ & & & \searrow^{\mathrm{eval}_\theta} & \\ & & & & K[\pi_1, ..., \pi_d, \theta] \end{array} \tag{4.28}$$

The kernel of the second morphism is the set of all polynomials over $K[\pi_1, ..., \pi_d]$ which map to zero after substitution by θ; this set is (f) by exercise 4.5.2. Since $\pi_1, ..., \pi_d$ are transcendental over K, the first map is actually an isomorphism (lemma 2.10.14); hence the kernel of the whole map φ is just (f), and by the first homomorphism theorem we have proved the claim.

Since $(f) = \sqrt{(f)} = \mathcal{I}(\mathcal{V}(f))$, we have therefore shown that $K[W] \simeq K[\pi_1, ..., \pi_d, \theta]$. Taking the fraction field, we have $K(W) \simeq K(\pi_1, ..., \pi_d, \theta)$; and by construction, $K(V) \simeq K(\pi_1, ..., \pi_d, \theta)$; thus $K(W) \simeq K(V)$ and so by corollary 4.3.23 we have $W \asymp V$. ∎

4.6 Global and local properties

In the previous chapters and sections, we primarily studied the 'global' features of algebraic sets: irreducibility, isomorphism, birational equivalence, dimension of a surface. We can vaguely characterise a global property as being a function from the set of all algebraic sets to some characterising set. Irreducibility is a function from the set $\mathrm{Ob}_{\mathsf{AffAlgSet}}$ to the set $\{\mathrm{True}, \mathrm{False}\}$; dimension is a function from the set of affine varieties to $\mathbb{N}_0$; and so forth.

In other fields of mathematics, there is significant interplay between global properties like these and 'local' properties: families of functions (one function for each object) from *points* on the objects of interest to some shared characterising set. For example, in elementary mathematics a local property of the set $D_{[0,1]}$ of differentiable functions $[0, 1] \to \mathbb{R}$ is the derivative (the family of all derivatives is a set of functions indexed by $D_{[0,1]}$ such that each point on a differentiable function is assigned a number, namely the 'slope'). Further, there is a global property on the same set, the function

$$\mathcal{J} : D_{[0,1]} \ni f \mapsto \int_0^1 f \in \mathbb{R}, \tag{4.29}$$

such that the local and global properties are linked by the fundamental theorem of calculus; namely, if $f \in D_{[0,1]}$ then the local property $D_f : x \mapsto f'(x)$ satisfies the relation $\mathcal{J}(D_f) = f(1) - f(0)$.

The question we wish to answer, then, is twofold: firstly, what are the interesting local properties of algebraic sets; and secondly, how do they interact with the global properties we have already studied?

By theorem 4.2.18, knowledge of the isomorphism classes of coordinate rings is equivalent to knowledge of the isomorphism classes of algebraic sets: in other words, the coordinate ring of an algebraic set X (equivalently, the set of all rational functions regular everywhere on X) is the primary algebraic object encoding the global properties of X. If $P \in X$, then the equivalent object for local properties at P is the set of all rational functions regular at the point P. Such functions should be of the form f/g for $f, g \in K[X]$ and $g(P) \neq 0$.

In order to actually construct a ring of these functions, we cannot simply take the subring of the function field because the function field is only defined for varieties (since for arbitrary affine varieties, the coordinate ring is not an integral domain). We cannot just restrict ourselves to varieties either, since we will often be interested in 'degenerate' points like $(0, 0) \in \mathcal{V}(X^2 - Y^2)$ (figure 2.2). Thus we need to generalise the construction of fraction fields to arbitrary rings by hand.

4.6.1 Definition. If R is an arbitrary ring, and $S \subseteq R$, then the **multiplicative closure** of S is the set

$$S_\times := \{ab : a, b \in S\} \cup \{1\}. \tag{4.30}$$

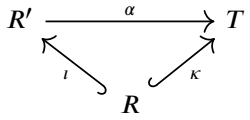
(The subset S has an associative binary operation with identity; it is called a **submonoid** of R.)

If $S = S_\times$, then S is called **multiplicatively closed**.

For example, the set of units of a ring R is a multiplicatively closed subset of R.

4.6.2 Proposition. *Let R be a ring, and let S be a multiplicatively closed subset of R. Then there exists a unique (up to isomorphism) ring R' and a unique injective homomorphism $\iota : R \hookrightarrow R'$ such that*

1. *every element of $\iota(S)$ is a unit in R', and*
2. *for every ring T and every injective homomorphism $\kappa : R \hookrightarrow T$ such that every element of $\kappa(S)$ is a unit in T, there exists a unique homomorphism of rings $\alpha : R' \to T$ such that the following diagram commutes:*



(4.31)

*The field R' is called the **localisation** of R at S, and is often denoted by $R[S^{-1}]$, or $\mathrm{Frac}_S R$. We shall use the latter notation in these notes.*

Proof. The proof is similar to the proof of proposition 2.1.18, and so we only sketch it. We will define R' to be the set consisting of ordered pairs $(r, s) \in R \times S$, modulo the equivalence relation $(r, s) \sim (r', s')$ iff there exists $t \in S$ such that $t(rs' - r's) = 0$; then the operations are given by $(r, s) \cdot (r', s') = (rr', ss')$ and $(r, s) + (r', s') = (rs' + r's, ss')$. As always, we will abuse notation and write r/s for (r, s). ∎

4.6.3 Example.

1. If R is an integral domain then $\mathrm{Frac}_{R^*} R \simeq \mathrm{Frac}\, R$.
2. If $0 \in S$, then $R' \simeq 0$.

Our main example of interest the following.

4.6.4 Proposition. *Suppose $\mathfrak{p} \subseteq R$ is a prime ideal. Then:*

1. *$R \setminus \mathfrak{p}$ is an algebraically closed subset of R, and hence $\mathrm{Frac}_{R\setminus\mathfrak{p}} R$ is well-defined;*
2. *$\mathrm{Frac}_{R\setminus\mathfrak{p}} R$ has only one maximal ideal $\mathfrak{m}$, which consists of all elements of the form f/g for $f \in \mathfrak{p}$ and $g \in R \setminus \mathfrak{p}$.*

*We denote $\mathrm{Frac}_{R\setminus\mathfrak{p}} R$ by $R_\mathfrak{p}$, and call it the **local ring** of R at $\mathfrak{p}$. (More generally, a local ring is a ring with a single maximal ideal.)*

Proof.

1. Since $\mathfrak{p}$ is prime, $\mathfrak{p} \neq R$ and so $1 \in R \setminus \mathfrak{p}$. It remains to check closure under multiplication; if $ab \in \mathfrak{p}$ then $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$, thus if neither a nor b lies in $\mathfrak{p}$ we must have that $ab \notin \mathfrak{p}$.
2. It is clear that $\mathfrak{m}$ is indeed an ideal of $R_\mathfrak{p}$. Further, suppose $\mathfrak{a}$ is an ideal of $R_\mathfrak{p}$ which is not contained in $R/\mathfrak{p}$, and pick $f/g \in \mathfrak{a} \setminus \mathfrak{m}$; then $f \notin \mathfrak{p}$, so $g/f \in R/\mathfrak{p}$ and f/g is a unit. In particular, $\mathfrak{a} = R$. This shows that $\mathfrak{m}$ is maximal and is the only maximal ideal. ∎

The geometric discussion above motivates the particular local ring we are interested in. There is a bijective correspondence between points of X and maximal ideals of $K[X]$ (this follows by the Nullstellensatz and the prime correspondence theorem, proposition 2.1.16); if $P \in X$ we shall write $\mathfrak{m}_{P,X}$ for this maximal ideal.

4.6.5 Definition. Let X be an algebraic set, and let $P \in X$. The **local ring** of X at P is the ring $\mathcal{O}_{P,X}$ defined by

$$\mathcal{O}_{P,X} := \{f/g : K[X]_{\mathfrak{m}_{P,X})}\}. \tag{4.32}$$

The local ring $\mathcal{O}_{P,X}$ inherits many of the algebraic properties of the coordinate ring: we can view it as being the ring which has same structure as $K[X]$ but which throws away all the information pertaining to the components of X that do not pass through P. Indeed, the irreducible components of X passing through P correspond to the prime ideals of $K[X]$ (by corollary 2.11.14 and the prime correspondence theorem, proposition 2.1.16); and so the following proposition makes this notion of localising precise.

4.6.6 Proposition. *There is a bijective correspondence between the prime ideals of $\mathcal{O}_{P,X}$ and the prime ideals of $K[X]$ contained in $\mathfrak{m}_{P,X}$.*

Proof. The correspondence will be set up as follows: we set up maps

$$\phi : \mathcal{O}_{P,X} \supseteq \mathfrak{p} \mapsto \{f \in K[X] : f/1 \in \mathfrak{p}\} \subseteq K[X] \tag{4.33}$$

$$\psi : K[X] \supseteq \mathfrak{p} \mapsto \{f/g \in \mathcal{O}_{P,X} : f \in \mathfrak{p}\} \subseteq \mathcal{O}_{P,X}. \tag{4.34}$$

First, it is clear by inspection that if $\mathfrak{p} \subseteq \mathcal{O}_{P,X}$ is an ideal then $\phi(\mathfrak{p})$ is an ideal of $K[X]$ and if $\mathfrak{p} \subseteq K[X]$ is an ideal then $\psi(\mathfrak{p})$ is an ideal of $\mathcal{O}_{P,X}$.

We next prove that ϕ and ψ are inverses. Suppose $\mathfrak{p} \subseteq \mathcal{O}_{P,X}$ is an ideal. Then if $f/g \in \mathfrak{p}$ we have $f/1 = g(f/g) \in \mathfrak{p}$ and thus $f \in \phi(\mathfrak{p})$, so $f/g \in \psi(\phi(\mathfrak{p}))$. Conversely, if $f/g \in \psi(\phi(\mathfrak{p}))$ then $f \in \mathfrak{p}$ and so $f/g = (1/g)f \in \mathfrak{p}$. This shows that $\psi(\phi(\mathfrak{p})) = \mathfrak{p}$. The reader should check that $\phi(\psi(\mathfrak{p})) = \mathfrak{p}$ for an ideal $\mathfrak{p} \subseteq K[X]$.

We next prove that ϕ and ψ preserve primality when we restrict to the sets of prime ideals in the statement of the proposition. Suppose $\mathfrak{p} \subseteq K[X]$ is a prime ideal contained in $\mathfrak{m}_{P,X}$. Then if $(f_1/g_1)(f_2/g_2) \in \psi(\mathfrak{p})$ we have that $f_1 f_2 \in \mathfrak{p}$ and thus either $f_1 \in \mathfrak{p}$ or $f_2 \in \mathfrak{p}$; by definition of ψ, in the first case we have $f_1/g_1 \in \psi(\mathfrak{p})$ and in the second case we have $f_2/g_2 \in \psi(\mathfrak{p})$. We need only check that $\psi(\mathfrak{p})$ is not the whole local ring. Suppose for the sake of contradiction that $u/v \in \psi(\mathfrak{p})$ is a unit. Then $v/u \in \psi(\mathfrak{p})$, so $v/1 = (u/1)(v/u) \in \psi(\mathfrak{p})$ and $v \in \mathfrak{p}$; but if $u/v \in \mathcal{O}_{P,X}$ then $v \notin \mathfrak{m}_{P,X} \supset \mathfrak{p}$, contradiction.

Conversely, suppose $\mathfrak{p} \subset \mathcal{O}_{P,X}$ is a prime ideal. If $fg \in \phi(\mathfrak{p})$ then $(fg)/1 = (f/1)(g/1) \in \mathfrak{p}$, so either $f/1$ or $g/1$ lies in $\mathfrak{p}$ and so either f or g lies in $\phi(\mathfrak{p})$. It remains to show that $\phi(\mathfrak{p}) \subseteq \mathfrak{m}_{P,X}$. Indeed, suppose $f \in K[X] \setminus \mathfrak{m}_{P,X}$; then $1/f \in \mathcal{O}_{P,X}$ and so $f/1$ is a unit in $\mathcal{O}_{P,X}$. Since $\mathfrak{p}$ is prime, $f/1 \notin \mathfrak{p}$ and hence $f \notin \phi(\mathfrak{p})$. ∎

4.6.7 Corollary. *The local ring $\mathcal{O}_{P,X}$ has finitely many prime ideals.* ∎

4.6.8 Proposition. *The local ring $\mathcal{O}_{P,X}$ is Noetherian.*

Proof. The map $\iota : K[X] \to \mathcal{O}_{P,X}$ given by $\iota(f) = f/1$ is the natural inclusion; hence if $\mathfrak{a}$ is any ideal of the local ring, $\iota^{-1}(\mathfrak{a})$ is an ideal of $K[X]$ and is finitely generated by corollary 4.2.4. Suppose $\iota^{-1}(\mathfrak{a}) = (f_1, ..., f_n)$. If $f \in \mathfrak{a}$, then $f/1 = \iota(g)/\iota(h)$ for some $g, h \in K[X]$ such that $h(P) \neq 0$. Thus there exists $t \in K[X] \setminus$
$mfa(P)$ such that $t(fh - g) = 0$ (this is by definition of equality in the local ring) and therefore $tg \in \iota^{-1}(\mathfrak{a})$; we may therefore write $tg = \sum r_i f_i$ for some $r_i \in K[X]$. Since neither t nor h vanish at P, $1/\iota(th) \in \mathcal{O}_{P,X}$; thus

$$f = \frac{\iota(g)}{\iota(h)} = \frac{\iota(tg)}{\iota(th)} = \frac{\iota(\sum r_i f_i)}{\iota(th)} = \sum \frac{\iota(r_i)}{\iota(th)} \iota(f_i) \tag{4.35}$$

and so f is generated by the $\iota(f_i)$; since f was arbitary, $\mathfrak{a} = (\iota(f_1), ..., \iota(f_n))$. ∎

As a first example of how local properties manifest themselves on this ring, the reader may wish to check that an earlier local definition may be rewritten to use this concept.

4.6.9 Exercise. Check that $\dim_P X = \dim \mathcal{O}_{P,X}$ (where the left hand side is given by definition 4.4.4, and the right hand side is the Krull dimension 4.4.7. By proposition 4.6.6, the Krull dimension of $\mathcal{O}_{P,X}$ is equal to the supremum of the lengths of chains of prime ideals downward from $\mathfrak{m}_{P,X}$ in X.

The maximal descending chain length is important enough that we distinguish it with a special name.

4.6.10 Definition. If $\mathfrak{p}$ is any prime ideal, the **height** of $\mathfrak{a}$ is the supremum of the lengths of the descending chains of prime ideals from $\mathfrak{p}$.

In particular, the above exercise shows that the dimension of X at a point P is the height of $\mathfrak{m}_{P,X}$. The following fact will be useful.

4.6.11 Exercise. In a ring, let $\mathfrak{p}$ be a prime ideal and let $\mathfrak{m}$ be maximal. If the height of $\mathfrak{p}$ is at least the height of $\mathfrak{m}$, we must have $\mathfrak{p} = \mathfrak{m}$.

4.7 Tangent spaces

In analogy with differential geometry, we want to attach to every point P on an algebraic set X a vector space $\Theta_{P,X}$. We shall first do this for the case where our set is embedded in $\mathbb{A}^n$, and then we will look for a definition that is independent of the ambient space (after all, a circle should have the same tangent space regardless of whether it is seen in $\mathbb{A}^2$ or $\mathbb{A}^3$).

We will define $\Theta_{P,X}$ to be the set of points on the lines 'tangent' to X at P. We shall now proceed by analogy with remark 1.2.1, and view L_α as being tangent to X if it 'intersects multiply' there. More precisely, we make the following definitions:

4.7.1 Definition. Let $X = \mathcal{V}(f_1, ..., f_k)$ be an algebraic set, P be a point of X, and L be a line through P of the form $L = \{t\alpha + P : t \in K\}$ for some $\alpha \in K^n$. Then the zero set of the polynomials

$$f_1(T\alpha + P), ..., f_k(T\alpha + P) \in K[T] \tag{4.36}$$

is precisely the set $L \cap X$. Since $K[T]$ is a UFD, there exists some $f \in K[T]$ of maximal degree such that $f \mid f_i(T\alpha + P)$ for all i. The **multiplicity** of L and X at P is the largest $n \in \mathbb{N}$ such that $T^n \mid f$; we denote it by $\mu_P(X, L)$.

4.7.2 Exercise. Check that the definition of $\mu_P(X, L)$ does not depend on the choice of generators $f_1, ..., f_k$.

4.7.3 Definition. A line L is **tangent** to X at P if $\mu_P(X, L) > 1$. The **tangent space** to X at P is the set

$$\Theta_{P,X} = \bigcup \{L \subseteq \mathbb{A}^n : L \text{ is a tangent line to } X \text{ at } P\}. \tag{4.37}$$

Recall that an **affine subspace** of $\mathbb{A}^n$ (also known as a **hyperplane** if the subspace has codimension 1) is an additive coset of a linear subspace of K^n. For the reader less confident in classical geometry, an excellent reference for the theory of affine spaces (a simple generalisation of linear algebra) is [Ber09, chapters 2 and 9].

4.7.4 Exercise. If $X = \mathcal{V}(f_1, ..., f_k)$, and if $P = (x_1, ..., x_n) \in \mathbb{A}^n$, then define $X - P$ to be the set $\{Q - P : Q \in X\}$. Show that:-

1. $X - P = \mathcal{V}(f_1(X_1 + x_1, ..., X_n + x_n), ..., f_k(X_1 + x_1, ..., X_n + x_n))$.
2. $\{t\alpha + Q : t \in K, Q \in X\}$ is tangent to X at Q iff $\{t\alpha + Q - P : t \in K, Q \in X\}$ is tangent to $X - P$ at $Q - P$.
3. $\Theta_{P,X} = \Theta_{Q-P,X-P} + P$.

4.7.5 Proposition. *$\Theta_{P,X}$ is an affine subspace of $\mathbb{A}^n$. (See figure 4.3.)*

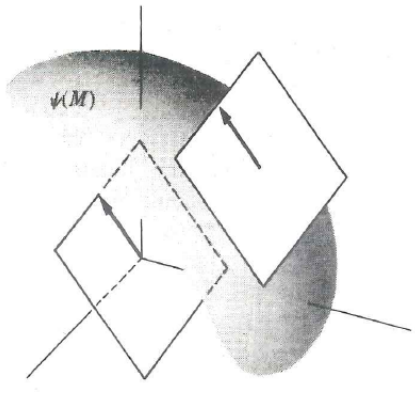


Figure 4.3: The tangent space of an algebraic set is an affine subspace. Figure from [LS14, p. 375].

Proof. First, suppose $0 \in X$ and $P = 0$. Write $X = \mathcal{V}(f_1, ..., f_k)$, let $Q \in X$, and consider the line $L := \{tQ : t \in K\}$ joining 0 and Q. Write $f_i = c_i + l_i + g_i$ for $\partial c_i < 1$, $\partial l_i = 1$, and $\partial g_i > 1$. For all i, since $f_i(0) = 0$, we must have $c_i = 0$. Now, $Q \in \Theta_{P,X}$ iff $\forall_i, T^2 \mid f_i(TQ)$ iff $\forall_i, T^2 \mid l_i(TQ) + g_i(TQ) = Tl_i(Q) + g_i(TQ)$ iff $\forall_i, l_i(Q) = 0$ iff $Q \in \mathcal{V}(l_1, ..., l_k)$; hence $\Theta_{P,X} = \mathcal{V}(l_1, ..., l_k)$ which is a linear subspace of $\mathbb{A}^n$ (it is the nullspace of a linear transformation).

Now clearly if X is an arbitrary algebraic set containing some point P, then we may simply perform an affine translation to reduce the situation to the case just proved. Indeed, by part 3 of exercise 4.7.4, we have that $\Theta_{P,X} = \Theta_{0,X-P} + P$; by part 1, $X - P$ is indeed an algebraic set; thus by the case we proved above, $\Theta_{0,X-P}$ is a linear subspace of $\mathbb{A}^n$; and so $\Theta_{P,X}$ is an affine subspace. ∎

4.7.6 Example.

1. Let $P \in \mathbb{A}^n$; then $\Theta_{P,\mathbb{A}^n} = \mathbb{A}^n$. Indeed, $\mathbb{A}^n$ is the vanishing set of (0); hence $Q \in \Theta_{P,\mathbb{A}^n} \iff Q \in \cap_{1 \le i \le k} \operatorname{null} 0 + P \iff Q \in \mathbb{A}^n + P = \mathbb{A}^n$.

2. Consider the cusped cubic in $\mathbb{A}^2$, $V = \mathcal{V}(X^2 - Y^3)$. At $(0,0)$, $\Theta_{P,V} = \mathbb{A}^2$ since the linear part of $X^2 - Y^3$ is zero. On the other hand, at $(1,1)$, we have

$$V - (1,1) = \mathcal{V}((X+1)^2 - (Y+1)^3) = \mathcal{V}(X^2 + 2X - Y^3 + 3Y^2 - 3Y); \tag{4.38}$$

   the linear terms are $l = 2X - 3Y$, and so the tangent space $\Theta_{0,V-(1,1)}$ is $\mathcal{V}(l)$; finally, $\Theta_{(1,1),V} = \Theta_{0,V-(1,1)} + (1,1) = \mathcal{V}(2(X-1) - 3(Y-1))$ (figure 4.4).

We will now recall a couple of basic notions from linear algebra.

4.7.7 Definition. Let V be a vector space over a field K. Then a linear map $\lambda : V \to K$ is called a **linear functional**, and the set of all such linear maps forms a vector space over K called the **dual space**, denoted by V^*.

If $S \subset \mathbb{A}^n$ is the affine subspace $V + P$ for a linear subspace V and point P, a linear functional on S is a function $\lambda : S \to K$ such that for all $Q \in S$, $\lambda(Q) = \tilde{\lambda}(Q - P)$ for some $\tilde{\lambda} \in V^*$. The set of all such maps forms a vector space over K which we call the dual space of S, and which we denote by S^*.

The following proposition is well-known.

4.7.8 Proposition. *Let V be a finite-dimensional vector space over K.*

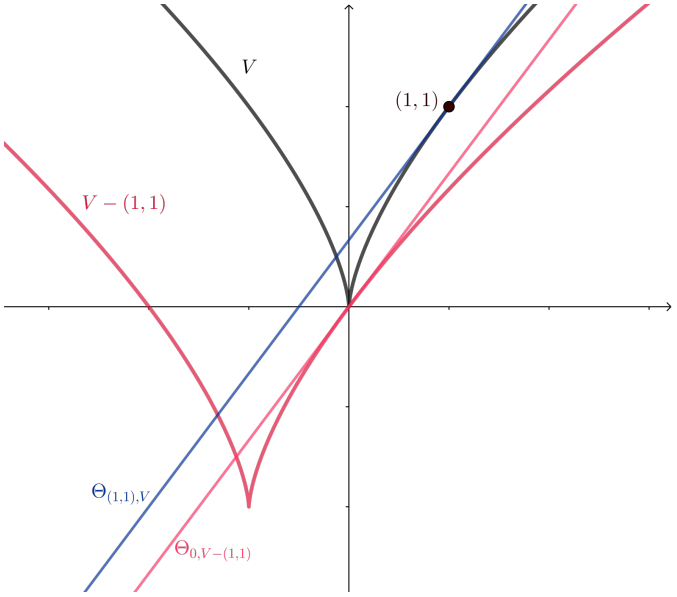


Figure 4.4: The tangent space of the cusped cubic at (1, 1).

1. *If* $(e_1, ..., e_n)$ *is a basis for* V*, then the tuple* $(f_1, ..., f_n)$ *where each* f_i *is defined by* $f_i(e_j) = \delta_{i,j}$ *is a basis for* V^**. In particular,* $\dim V = \dim V^*$ *if* V *is finite dimensional.*

2. *There is a natural isomorphism between* V *and* V^{**} *given on the basis by* $e_i \mapsto (f \mapsto f(e_i))$ *for all* i*.*

4.7.9 Exercise. Prove proposition 4.7.8. The confused reader should see [Hal15a, §§13–15].

4.7.10 Exercise. Show that the category $\mathsf{Vec}(K)$ of finite-dimensional vector spaces over K is dually equivalent to itself.[4]

This linearisation l_i of f_i at a point P in the proof of proposition 4.7.5 is analogous to the differential $\mathrm{d}\, f$ of a continuous function f in differential geometry. It will be useful to define it for arbitrary polynomials.

4.7.11 Definition. Let $f \in K[X_1, ..., X_n]$, and let $P = (x_1, ..., x_n) \in K^n$. Then the **differential** $\mathrm{d}_P f$ is the polynomial consisting of the linear terms of $f(X_1 - x_1, ..., X_n - x_n)$.

https://mathoverflow.net/q/345136

4.7.12 Proposition. *Let* $X = \mathcal{V}(f_1, ..., f_k)$*; then* $Q \in \Theta_{P,X} \iff \forall_i \mathrm{d}_P f_i(Q) = 0$*.*

Proof. This follows directly from the proof of proposition 4.7.5. ∎

Recall from definition 4.5.8 that we have defined algebraic derivatives for polynomials $f \in R[X]$. We have a simple equality, familiar from calculus, relating these to the differential.

4.7.13 Proposition. $\mathrm{d}_P f = \sum_{i=0}^{n} D_{X_i} f(P)(X_i - x_i)$.

Proof. This may be shown by a lengthy but elementary calculation involving the binomial theorem. ∎

4.7.14 Corollary. *If* $\lambda \in K$ *and* $f, g \in k[X_1, ..., X_n]$*, then* $\mathrm{d}_P(\lambda) = \lambda\, \mathrm{d}_P f$*,* $\mathrm{d}_P(f + g) = \mathrm{d}_P f + \mathrm{d}_P g$*, and* $\mathrm{d}_P(fg) = (\mathrm{d}_P f)g + f(\mathrm{d}_P g)$*.* ∎

[4] See also `https://mathoverflow.net/q/345136`.

Now suppose X is an algebraic set, and let $f \in K[X_1, ..., X_n]$. We cannot simply define $\mathrm{d}_P[f] = \mathrm{d}_P f$ for $[f] \in K[X]$, since it is not necessarily true that if $[f] = [g]$ then $[\mathrm{d}_P f] = [\mathrm{d}_P g]$; more precisely, $f - g \in \mathcal{I}(X)$ implies only that $\mathrm{d}_P[f] - \mathrm{d}_P[g] = \mathrm{d}_P[h]$ for some $h \in \mathcal{I}(X)$, and derivatives may be non-zero at points where the primitives are zero so $\mathrm{d}_P[h]$ is not necessarily in $\mathcal{I}(X)$. Luckily, if we restrict ourselves to local behaviour, everything works.

4.7.15 Proposition. *Let X be an algebraic set, and let $P \in X$. For $f \in K[X_1, ..., X_n]$, define*

$$\mathrm{d}_P[f] = \mathrm{d}_P f \restriction_{\Theta_{P,X}} \in \Theta^*_{P,X}. \tag{4.39}$$

If $f, g \in K[X_1, ..., X_n]$ satisfy $[f] = [g]$, then $\mathrm{d}_P[f](Q) = \mathrm{d}_P[g](Q)$ for all $Q \in \Theta_{P,X}$.

Proof. If $[f] = [g]$ then $f - g \in \mathcal{I}(X)$. Let $Q \in \Theta_{P,X}$. Then $\mathrm{d}_P f(Q) - \mathrm{d}_P g(Q) = \mathrm{d}_P(f - g)(Q)$; By proposition 4.7.12, this last quantity is zero; and since $Q \in \Theta_{P,X}$, $\mathrm{d}_P f(Q) - \mathrm{d}_P g(Q) = \mathrm{d}_P f \restriction_{\Theta_{P,X}}(Q) - \mathrm{d}_P g \restriction_{\Theta_{P,X}}(Q) = \mathrm{d}_P[f](Q) - \mathrm{d}_P[f](Q)$. Hence $\mathrm{d}_P[f](Q) - \mathrm{d}_P[f](Q) = 0$. ∎

Note that every linear functional $\lambda : \Theta_{P,X} \to K$ is a linear map which sends P to zero. If we denote the set of polynomial functions in $K[X]$ which vanish at P by $\mathfrak{m}_{P,X}$, then $\lambda \in \mathfrak{m}_{P,X}$; hence $\Theta^*_{P,X}$ is a subset of $\mathfrak{m}_{P,X}$. Further, for every polynomial in $\mathfrak{m}_{P,X}$ we can obtain a linear functional by 'throwing away the non-linear terms at P' similarly to how we did so in the proof of proposition 4.7.5. When we do this, we obtain an equivalence relation: $f \sim g$ if f equals g modulo the non-linear terms. This motivates the following theorem.

4.7.16 Theorem. *Let X be an algebraic set, and let $P = (x_1, ..., x_n) \in X$. There is an isomorphism between the vector spaces $\mathfrak{m}_{P,X}/\mathfrak{m}^2_{P,X}$ and $\Theta^*_{P,X}$.*

Proof. As the preceeding discussion suggests, we will construct this isomorphism by way of linearisation. Note that d_P is a linear map from $K[X] \to \Theta^*_{P,X}$ (by corollary 4.7.14). We will use the first homomorphism theorem for linear spaces ([Hal15a, §22]); first, we compute $\mathrm{d}_P(\mathfrak{m}_{P,X})$. By definition, this set will be a subset of $\Theta^*_{P,X}$; and if $\lambda \in \Theta^*_{P,X}$ then λ is a polynomial in $K[X_1, ..., X_n]$ which vanishes at P, hence lies in $\mathfrak{m}_{P,X}$; and $\mathrm{d}_P \lambda = \lambda$ by definition. Hence $\mathrm{d}_P(\mathfrak{m}_{P,X}) = \Theta^*_{P,X}$.

Now suppose $f \in \ker \mathrm{d}_P$. Then the linear terms of f at P are zero; in particular, if $(X_i - x_i) \mid f$ then $(X_i - x_i)^2 \mid f$ for all i and so $f \in \mathfrak{m}^2_{P,X}$ by corollary 2.11.4. Conversely, if $f \in \mathfrak{m}^2_{P,X}$ then $\mathrm{d}_P f = 0$; thus $\ker \mathrm{d}_P = \mathfrak{m}^2_{P,X}$.

Thus $\mathrm{d}_P f$ is an onto homomorphism with $\ker \mathrm{d}_P = \mathfrak{m}^2_{P,X}$, $\operatorname{im} \mathrm{d}_P = \Theta^*_{P,X}$; hence $\mathfrak{m}_{P,X}/\mathfrak{m}^2_{P,X} \simeq \Theta^*_{P,X}$. ∎

4.7.17 Proposition. *Let $f : X \to Y$ be an isomorphism of algebraic sets. If $P \in X$ then $\Theta_{P,X} \simeq \Theta_{f(P),Y}$.*

Proof. The isomorphism f induces an isomorphism $f^* : K[Y] \to K[X]$. Since $f^*(\mathfrak{m}_{f(P)}) \subset \mathfrak{m}_{P,X}$ and $f^*(\mathfrak{m}^2_{f(P)}) \subset \mathfrak{m}^2_P$, we obtain a restriction $f^* \restriction_{\mathfrak{m}^2_{f(P)}/\mathfrak{m}_{f(P)}}$ into $\mathfrak{m}^2_P/\mathfrak{m}_{P,X}$; by bijectivity of f, this restriction is bijective; hence f^* is an isomorphism between $\Theta^*_{f(P),Y} \simeq \mathfrak{m}^2_{f(P)}/\mathfrak{m}_{f(P)}$ and $\Theta^*_{P,X} \simeq \mathfrak{m}^2_P/\mathfrak{m}_{P,X}$. Taking duals, we obtain the desired isomorphism of tangent spaces. ∎

The tangent space is not only a property of the point, it is a property of the local ring. Because the local ring is a set of rational functions, we need to extend our notion of differential to this case.

4.7.18 Definition. Let $f/g \in \mathcal{O}_{P,X}$. Then the **differential** of ρ, $\mathrm{d}_P \rho$, is defined to be the linear map[5]

$$\mathrm{d}_P \rho := \frac{g(P)\,\mathrm{d}_P f - f(P)\,\mathrm{d}_P g}{[g(P)]^2}. \tag{4.40}$$

[5] If $P = (x_1, ..., x_n)$, then we could in fact define it to be the linear part of $f(X_1 - x_1, ..., X_n - x_n)/g(X_1 - x_1, ..., X_n - x_n)$ in analogy with the earlier definition 4.7.11. An excellent exposition of the properties of differentials, which is developed in a way compatible with the current situation, may be found in [LS14, §3.6].

4.7.19 Exercise. Check that the differential of $\rho \in K(X)$ at $P \in X$ is independent of the representative of ρ as a quotient.

4.7.20 Proposition. *Let X be an algebraic set, and let $P = (x_1, ..., x_n) \in X$. There is an isomorphism between the vector spaces $\mathfrak{m}/\mathfrak{m}^2$ and $\Theta^*_{P,X}$, where $\mathfrak{m}$ is the unique maximal ideal of $\mathcal{O}_{P,X}$ (proposition 4.6.4).*

Note that $\mathfrak{m}/\mathfrak{m}^2$ is a vector space over the field $\mathcal{O}_{P,X}/\mathfrak{m}$, not over the field K. Thus this proposition also shows that the dimensions of $\Theta^*_{P,X}$ over the fields $\mathcal{O}_{P,X}/\mathfrak{m}$ and K are equal.

Proof. The proof is essentially similar to the proof of theorem 4.7.16 above: one must show first that $\mathrm{d}_P \rho : \mathfrak{m} \to \Theta^*_{P,X}$ is linear; then one shows that its kernel is $\mathfrak{m}^2$; and then one concludes using the first homomorphism theorem for linear maps. ∎

The vector space $\Theta^*_{P,X}$ is called the **cotangent space** to X at P; since by part 2 of proposition 4.7.8 we have $\Theta^{**}_{P,X} \simeq \Theta_{P,X}$, we have a purely algebraic characterisation of the tangent space: namely:

4.7.21 Corollary. $\Theta_{P,X} \simeq (\mathfrak{m}/\mathfrak{m}^2)^*$ *(where $\mathfrak{m}$ is the maximal ideal of $\mathcal{O}_{P,X}$).* ∎

We would now like to study the dimension of the tangent space; it would be nice if $\dim \Theta_{P,X} = \dim_P X$. However, we do have issues at 'bad' points like $(0,0)$ on the cusped cubic (example 4.7.6) where the dimension is too large.

4.7.22 Definition. Let X be an algebraic set. Then $P \in X$ is called **nonsingular** if $\dim \Theta_{P,X} = \min_{Q \in X} \dim \Theta_{Q,X}$; the point P is called **singular** if $\dim \Theta_{P,X} > \min_{Q \in X} \dim \Theta_{Q,X}$. The set of singular points of X will be denoted by $\operatorname{Sing} X$.

The entire set X is called **smooth** if every $P \in X$ is nonsingular.

We want to formalise and then prove the following two statements:

1. Almost every point in an algebraic set X is non-singular.
2. At every non-singular point P, the tangent space has equal dimension to the dimension of X at P.

We will achieve this in theorem 4.7.26 below.

For the next two lemmata, fix an algebraic set X and a point $P \in X$. Let $\mathfrak{m}$ be the maximal ideal of $\mathcal{O}_{P,X}$.

We shall first prove a commutative algebra result; the idea is that if X has dimension d at P then the point P itself is a vanishing point of X from 'at most d directions'.

4.7.23 Lemma (Annoying lemma)**.** *If $\dim \mathcal{O}_{P,X} = d$, then there exists an ideal $\mathfrak{a} \subseteq \mathcal{O}_{P,X}$ satisfying the following properties:*

1. *$\sqrt{\mathfrak{a}} = \mathfrak{m}$; and*
2. *$\mathfrak{a}$ has a generating set consisting of at most d elements of $\mathcal{O}_{P,X}$.*

The proof of this lemma is annoying, and I am not entirely sure why as it is also fairly straightforward.

Proof. We will construct $\mathfrak{a}$ by finding d elements $f_1, ..., f_d \in \mathcal{O}_{P,X}$ such that for each i, the height of every prime ideal containing $(f_1, ..., f_i)$ is at least i. Set $\mathfrak{a} = (f_1, ..., f_d)$; by proposition 2.9.9, to verify part 1 of the lemma we need only check that $\mathfrak{m}$ is the intersection of the prime ideals containing $\mathfrak{a}$. Indeed, if $\mathfrak{p}$ is a prime ideal containing $\mathfrak{a}$ then the height of $\mathfrak{p}$ is at least d; since the height of $\mathfrak{m}$ is d and $\mathfrak{p} \subseteq \mathfrak{m}$ we must have $\mathfrak{m} = \mathfrak{p}$ (exercise 4.6.11) and so the only prime ideal containing $\mathfrak{a}$ is $\mathfrak{m}$ itself. We now proceed to construct $\mathfrak{a}$.

Suppose $i \in \mathbb{N}$ is such that f_j has been constructed for $j < i$ (in particular, for the base case i could be 1). Let S be the set of prime ideals of $\mathcal{O}_{P,X}$ containing $(f_1, ..., f_{i-1})$ with height exactly $i-1$; by corollary 4.6.7, S is finite. Since $\dim \mathcal{O}_{P,X} = d$ is the height of $\mathfrak{m}$ and $i - 1 < d$, $\mathfrak{m} \notin S$ and so by proposition 2.1.17, $\mathfrak{m} \not\subseteq \cup S$.

We may therefore pick some $f_i \in \cup S \setminus \mathfrak{m}$; we will check that every prime ideal $\mathfrak{p} \supseteq (f_1, ..., f_i)$ has height at least i. If $\mathfrak{p}$ contains some ideal $\mathfrak{s} \in S$ by inclusion then $\mathfrak{s} \subset \mathfrak{p}$ is a strict chain of prime ideals and so the height of $\mathfrak{s}$ ($i-1$ by definition) is strictly less than the height of $\mathfrak{p}$: hence the height of $\mathfrak{p}$ is at least i. On the other hand, if $\mathfrak{p}$ contains no such ideal then the height of $\mathfrak{p}$ cannot be precisely $i-1$ (otherwise $\mathfrak{p} \in S$) and hence the height is at least i. ∎

We are actually only interested in the following awkward corollary.

4.7.24 Corollary. *The number δ defined by*

$$\delta := \min\{n \in \mathbb{N} : \exists_{\mathfrak{a} \subseteq \mathcal{O}_{P,X}} \sqrt{\mathfrak{a}} = \mathfrak{m} \text{ and } \mathfrak{a} \text{ has a generating set of at most } n \text{ elements}\} \tag{4.41}$$

satisfies the inequality $\delta \leq \dim \mathcal{O}_{P,X}$.

We can now check that the dimension of the tangent space at P is bounded below by the dimension of X at P.

4.7.25 Lemma. *The inequality $\dim \mathfrak{m}/\mathfrak{m}^2 \geq \dim_P X$ holds.*

Proof. By exercise 4.6.9 it suffices to show that $\dim \mathfrak{m}/\mathfrak{m}^2 \geq \dim \mathcal{O}_{P,X}$. Suppose $(v_i)_{i=1}^d$ is a tuple in $\mathfrak{m}$ such that $(v_i + \mathfrak{m}^2)_{i=d}^d$ is a basis for $\mathfrak{m}/\mathfrak{m}^2$.

▶ **Claim A.** *The set $\{v_i\}$ generates $\mathfrak{m}$ as an algebra over $\mathcal{O}_{P,X}$.* ◀

By corollary 4.7.23 we have that $d \leq \delta \leq \dim \mathcal{O}_{P,X}$ ∎

We can now prove our two main theorems. The first one gives us an equivalent characterisation of singularity based on the differentials of the generating polynomials of our algebraic set.

4.7.26 Theorem. *Let $X \subseteq \mathbb{A}^n$ be an algebraic set such that $\mathcal{I}(X) = (f_1, ..., f_k)$. Then X is nonsingular at P if and only if*

$$\operatorname{rank} \begin{bmatrix} D_{X_1} f_1 & \cdots & D_{X_n} f_1 \\ \vdots & \ddots & \vdots \\ D_{X_1} f_k & \cdots & D_{X_n} f_k \end{bmatrix} = n - \dim X. \tag{4.42}$$

Our second main theorem will show the two goals we outlined above.

4.7.27 Theorem. *Let X be an algebraic set. Then:*

1. *$\operatorname{Sing} X$ is a proper closed subset of X.*
2. *If $P \in X \setminus \operatorname{Sing} X$ then $\dim_P X = \dim \Theta_{P,X}$.*

4.8 Intersection numbers

Chapter 5

Projective varieties

Chapter 6

Schemes

Bibliography and further reading

[AHS90] Jiří Adámek, Horst Herrlich, and George E. Strecker. *Abstract and concrete categories*. John Wiley & Sons, 1990.

[Alu09] Paolo Aluffi. *Algebra: Chapter 0*. Graduate studies in mathematics 104. American Mathematical Society, 2009.

[AM69] M. F. Atiyah and I. G. Macdonald. *Introduction to commutative algebra*. Addison-Wesley Publishing Company, 1969.

[Art91] Michael Artin. *Algebra*. Prentice-Hall Ltd, 1991.

[Ber09] Marcel Berger. *Geometry I*. Trans. by M. Cole and S. Levy. Fourth revised printing. Universitext. Springer, 2009.

[BK86] Egbert Brieskorn and Horst Knörrer. *Plane algebraic curves*. Trans. by John Stillwell. Birkhäuser, 1986.

[CLO04] David A. Cox, John Little, and Donal O'Shea. *Using Algebraic Geometry*. 2nd ed. Graduate texts in mathematics 185. Springer, 2004.

[CW] Eugenia Cheng and Simon Willerton. *The Catsters*. URL: `http://www.youtube.com/thecatsters`.

[Die85] Jean Dieudonné. *History of algebraic geometry. An outline of the history and development of algebraic geometry*. Trans. by Judith D. Sally. Wadsworth Advanced Books & Software, 1985.

[DO07] Erik D. Demaine and Joseph O'Rourke. *Geometric folding algorithms*. Cambridge University Press, 2007.

[Edw77] Harold M. Edwards. *Fermat's last theorem. A genetic introduction to algebraic number theory*. Graduate texts in mathematics 50. Springer-Verlag, 1977.

[Eis95] David Eisenbud. *Commutative algebra with a view toward algebraic geometry*. Graduate texts in mathematics 150. Springer-Verlag, 1995.

[Fet83] Henry E. Fettis. 'The geometry of Dürer's conchoid'. In: *Crux Mathematicorum* 9.2 (Feb. 1983), pp. 32–37. URL: `https://cms.math.ca/crux/backfile/Crux_v9n02_Feb.pdf`.

[Hal15a] Paul R. Halmos. *Finite-dimensional vector spaces*. Oxford: Benediction Classics, 2015.

[Hal15b] Paul R. Halmos. *Naïve set theory*. Oxford: Benediction Classics, 2015.

[Har77] Robin Hartshorne. *Algebraic Geometry*. Graduate texts in mathematics 52. Springer-Verlag, 1977.

[Hol12] Audun Holme. *A royal road to algebraic geometry*. Springer, 2012.

[IR82] Kenneth Ireland and Michael Rosen. *A classical introduction to modern number theory*. Graduate texts in mathematics 84. Springer-Verlag, 1982.

[Lan71] Serge Lang. *Algebra*. Addison Wesley, 1971.

[Law72] J. Dennis Lawrence. *A catalog of special plane curves*. Dover Publications, 1972.

[Lin71] Carl E Linderholm. *Mathematics made difficult*. Butler & Tanner Ltd, 1971.

[LS14] Lynn Loomis and Shlomo Sternberg. *Advanced calculus*. Revised edition. World Scientific, 2014.

[Mun00] James Munkres. *Topology*. 2nd ed. Pearson, 2000.

[Nag77] Masayoshi Nagata. *Field theory*. Pure and applied mathematics 40. Marcel Dekker, 1977.

[Rud66] Walter Rudin. *Real and complex analysis*. 3rd ed. McGraw Hill, 1966.

[Sal50] George Salmon. *A treatise on conic sections. Containing an account of some of the most important modern algebraic and geometric methods*. 2nd ed. Facsimile Publisher, 1850.

[Sha94a] Igor R. Shafarevich. *Basic algebraic geometry*. 2nd ed., revised and expanded. Vol. 1. Springer-Verlag, 1994.

[Sha94b] Igor R. Shafarevich. *Basic algebraic geometry*. 2nd ed., revised and expanded. Vol. 2. Springer-Verlag, 1994.

[Sin97] Simon Singh. *Fermat's last theorem*. Fourth Estate, 1997.

[SS95] Lynn Arthur Steen and J. Arthur Seebach Jr. *Counterexamples in topology*. Dover Publications, 1995.

[Ste15] Ian Stewart. *Galois theory*. 4th ed. CRC Press, 2015.

[Wae49] B. L. van der Waerden. *Modern algebra*. Trans. by Fred Blum. Springer, 1949.

Index