Commutative Algebra

Andreas Gathmann

Class Notes TU Kaiserslautern 2013/14

Contents

0. Introduction

Commutative algebra is the study of commutative rings. In this class we will assume the basics of ring theory that you already know from earlier courses (e. g. ideals, quotient rings, the homomorphism theorem, and unique prime factorization in principal ideal domains such as the integers or polynomial rings in one variable over a field), and move on to more advanced topics, some of which will be sketched in Remark 0.14 below. For references to earlier results I will usually use my German notes for the "Algebraic Structures" and occasionally the "Foundations of Mathematics" and "Introduction to Algebra" classes [G1, G2, G3], but if you prefer English references you will certainly have no problems to find them in almost any textbook on abstract algebra.

You will probably wonder why the single algebraic structure of commutative rings deserves a full one-semester course for its study. The main motivation for this is its many applications in both *algebraic geometry* and *(algebraic) number theory*. Especially the connection between commutative algebra and algebraic geometry is very deep — in fact, to a certain extent one can say that these two fields of mathematics are essentially the same thing, just expressed in different languages. Although some algebraic constructions and results in this class may seem a bit abstract, most of them have an easy (and sometimes surprising) translation in terms of geometry, and knowing about this often helps to understand and remember what is going on. For example, we will see that the Chinese Remainder Theorem that you already know [G1, Proposition 11.22] (and that we will extend to more general rings than the integers in Proposition 1.14) can be translated into the seemingly obvious geometric statement that "giving a function on a disconnected space is the same as giving a function on each of its connected components" (see Example 1.15 (b)).

However, as this is not a geometry class, we will often only sketch the correspondence between algebra and geometry, and we will never actually use algebraic geometry to prove anything. Although our "Commutative Algebra" and "Algebraic Geometry" classes are deeply linked, they are deliberately designed so that none of them needs the other as a prerequisite. But I will always try to give you enough examples and background to understand the geometric meaning of what we do, in case you have not attended the "Algebraic Geometry" class yet.

So let us explain in this introductory chapter how algebra enters the field of geometry. For this we have to introduce the main objects of study in algebraic geometry: solution sets of polynomial equations over some field, the so-called *varieties*.

Convention 0.1 (Rings and fields)**.** In our whole course, a *ring* R is always meant to be a commutative ring with 1 [G1, Definition 7.1]. We do not require that this multiplicative unit 1 is distinct from the additive neutral element 0, but if $1 = 0$ then R must be the zero ring [G1, Lemma 7.5 (c)]. Subrings must have the same unit as the ambient ring, and ring homomorphisms are always required to map 1 to 1. Of course, a ring $R \neq \{0\}$ is a *field* if and only if every non-zero element has a multiplicative inverse.

Definition 0.2 (Polynomial rings)**.** Let R be a ring, and let $n \in \mathbb{N}_{>0}$. A **polynomial** over R in n variables is a formal expression of the form

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

with coefficients $a_{i_1,\dots,i_n} \in R$ and formal variables $x = (x_1, \dots, x_n)$, such that only finitely many of the coefficients are non-zero (see [G1, Chapter 9] how this concept of "formal variables" can be defined in a mathematically rigorous way).

Polynomials can be added and multiplied in the obvious way, and form a ring with these operations. We call it the **polynomial ring** over R in n variables and denote it by $R[x_1, \dots, x_n]$.

Definition 0.3 (Varieties)**.** Let K be a field, and let $n \in \mathbb{N}$.

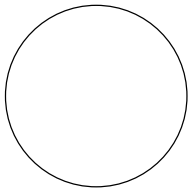
(a) We call
$$\mathbb{A}^n_K := \{(c_1,\dots,c_n) : c_i \in K \text{ for } i = 1,\dots,n\}$$
the **affine n-space** over K. If the field K is clear from the context, we will write $\mathbb{A}^n_K$ also as $\mathbb{A}^n$.

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

(b) For a polynomial $f \in K[x_1,\dots,x_n]$ as above and a point $c = (c_1,\dots,c_n) \in \mathbb{A}^n_K$ we define the **value** of f at c to be
$$f(c) = \sum_{i_1,\dots,i_n \in \mathbb{N}} a_{i_1,\dots,i_n}\, c_1^{i_1} \cdot \dots \cdot c_n^{i_n} \quad \in K.$$
If there is no risk of confusion we will sometimes denote a point in $\mathbb{A}^n_K$ by the same letter x as we used for the formal variables, writing $f \in K[x_1,\dots,x_n]$ for the polynomial and $f(x)$ for its value at a point $x \in \mathbb{A}^n_K$.

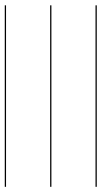
(c) Let $S \subset K[x_1,\dots,x_n]$ be a set of polynomials. Then
$$V(S) := \{x \in \mathbb{A}^n_K : f(x) = 0 \text{ for all } f \in S\} \quad \subset \mathbb{A}^n_K$$
is called the **zero locus** of S. Subsets of $\mathbb{A}^n_K$ of this form are called **(affine) varieties**. If $S = (f_1,\dots,f_k)$ is a finite set, we will write $V(S) = V(\{f_1,\dots,f_k\})$ also as $V(f_1,\dots,f_k)$.

Example 0.4. Varieties, say over the field $\mathbb{R}$ of real numbers, can have many different "shapes". The following picture shows a few examples in $\mathbb{A}^2_{\mathbb{R}}$ and $\mathbb{A}^3_{\mathbb{R}}$.



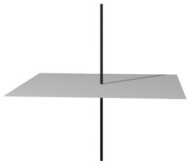
(a) $V(x_1^2+x_2^2-1) \subset \mathbb{A}^2$

(b) $V(x_2^2-x_1^3) \subset \mathbb{A}^2$

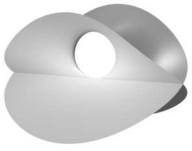


(c) $V(x_1^3-x_1) \subset \mathbb{A}^2$

(d) $V(x_1^6+x_2^6+x_3^6-1) \subset \mathbb{A}^3$



(e) $V(x_1x_3, x_2x_3) \subset \mathbb{A}^3$



(f) $V(x_2^2+x_3^3-x_3^4-x_1^2x_3^2) \subset \mathbb{A}^3$

Of course, the empty set $\emptyset$ and all of $\mathbb{A}^n$ are also varieties in $\mathbb{A}^n$, since $\emptyset = V(1)$ and $\mathbb{A}^n = V(0)$.

It is the goal of algebraic geometry to find out the geometric properties of varieties by looking at the corresponding polynomials from an algebraic point of view (as opposed to an analytical or numerical approach). However, it turns out that it is not a very good idea to just look at the defining polynomials given initially — simply because they are not unique. For example, the variety (a) above was given as the zero locus of the polynomial $x_1^2+x_2^2-1$, but it is equally well the zero locus of $(x_1^2+x_2^2-1)^2$, or of the two polynomials $(x_1-1)(x_1^2+x_2^2-1)$ and $x_2(x_1^2+x_2^2-1)$. In order to

remove this ambiguity, it is therefore useful to consider *all* polynomials vanishing on X at once. Let us introduce this concept now.

Construction 0.5 (Rings and ideals associated to varieties)**.** For a variety $X \subset \mathbb{A}^n_K$ (and in fact also for any subset X of $\mathbb{A}^n_K$) we consider the set

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

of all polynomials vanishing on X. Note that this is an ideal of $K[x_1,\dots,x_n]$ (which we write as $I(X) \trianglelefteq K[x_1,\dots,x_n]$): it is clear that $0 \in I(X)$, and if two polynomials f and g vanish on X, then so do $f+g$ and $f \cdot h$ for any polynomial h. We call $I(X)$ the **ideal** of X.

With this ideal we can construct the quotient ring

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

in which we identify two polynomials $f,g \in K[x_1,\dots,x_n]$ if and only if $f-g$ is the zero function on X, i. e. if f and g have the same value at every point $x \in X$. So one may think of an element $\overline{f} \in A(X)$ as being the same as a function

$$X \to K,\ \ x \mapsto f(x)$$

that can be given by a polynomial. We therefore call $A(X)$ the **ring of polynomial functions** or **coordinate ring** of X. Often we will simply say that $A(X)$ is the *ring of functions* on X since functions in algebra are always given by polynomials. Moreover, the class of a polynomial $f \in K[x_1,\dots,x_n]$ in such a ring will usually also be written as $f \in A(X)$, dropping the explicit notation for equivalence classes if it is clear from the context that we are talking about elements in the quotient ring.

Remark 0.6 (Polynomials and polynomial functions)**.** You probably know that over some fields there is a subtle difference between *polynomials* and *polynomial functions*: e. g. over the field $K = \mathbb{Z}_2$ the polynomial $f = x^2 + x \in K[x]$ is certainly non-zero, but it defines the zero function on $\mathbb{A}^1_K$ [G1, Remark 9.16 (b)]. In our current notation this means that the ideal $I(\mathbb{A}^1_K)$ of functions vanishing at every point of $\mathbb{A}^1_K$ is non-trivial, in fact that $I(\mathbb{A}^1_K) = (x^2+x)$, and that consequently the ring $A(\mathbb{A}^1_K) = K[x]/(x^2+x)$ of polynomial functions on $\mathbb{A}^1_K$ is not the same as the polynomial ring $K[x]$.

In this class we will skip over this problem entirely, since our main geometric intuition comes from the fields of real or complex numbers where there is no difference between polynomials and polynomial functions. We will therefore usually assume silently that there is no polynomial $f \in K[x_1,\dots,x_n]$ vanishing on all of $\mathbb{A}^n_K$, i. e. that $I(\mathbb{A}^n_K) = (0)$ and thus $A(\mathbb{A}^n_K) = K[x_1,\dots,x_n]$.

Example 0.7 (Ideal of a point)**.** Let $a = (a_1,\dots,a_n) \in \mathbb{A}^n_K$ be a point. We claim that its ideal $I(a) := I(\{a\}) \trianglelefteq K[x_1,\dots,x_n]$ is

$$I(a) = (x_1 - a_1, \dots, x_n - a_n).$$

In fact, this is easy to see:

- "$\subset$" If $f \in I(a)$ then $f(a) = 0$. This means that replacing each x_i by a_i in f gives zero, i. e. that f is zero modulo $(x_1 - a_1,\dots,x_n - a_n)$. Hence $f \in (x_1 - a_1,\dots,x_n - a_n)$.
- "$\supset$" If $f \in (x_1 - a_1,\dots,x_n - a_n)$ then $f = \sum_{i=1}^n (x_i - a_i) f_i$ for some $f_1,\dots,f_n \in K[x_1,\dots,x_n]$, and so certainly $f(a) = 0$, i. e. $f \in I(a)$.

Construction 0.8 (Subvarieties)**.** The ideals of varieties defined in Construction 0.5 all lie in the polynomial ring $K[x_1,\dots,x_n]$. In order to get a geometric interpretation of ideals in more general rings it is useful to consider a relative situation: let $X \subset \mathbb{A}^n_K$ be a fixed variety. Then for any subset $S \subset A(X)$ of polynomial functions on X we can consider its zero locus

$$V_X(S) = \{x \in X : f(x) = 0 \text{ for all } f \in S\} \quad \subset X$$

just as in Definition 0.3 (c), and for any subset $Y \subset X$ as in Construction 0.5 the ideal

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

of all functions on X that vanish on Y. It is clear that the sets of the form $V_X(S)$ are exactly the varieties in $\mathbb{A}^n_K$ contained in X, the so-called **subvarieties** of X.

If there is no risk of confusion we will simply write $V(S)$ and $I(Y)$ again instead of $V_X(S)$ and $I_X(Y)$. So in this notation we have now assigned to every variety X a ring $A(X)$ of polynomial functions on X, and to every subvariety $Y \subset X$ an ideal $I(Y) \trianglelefteq A(X)$ of the functions that vanish on Y. This assignment of an ideal to a subvariety has some nice features:

Lemma 0.9. *Let X be a variety with coordinate ring $A(X)$. Moreover, let Y and Y' be subsets of X, and let S and S' be subsets of $A(X)$.*

(a) *If $Y \subset Y'$ then $I(Y') \subset I(Y)$ in $A(X)$; if $S \subset S'$ then $V(S') \subset V(S)$ in X.*

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

(c) *If Y is a subvariety of X then $Y = V(I(Y))$.*

(d) *If Y is a subvariety of X then $A(X)/I(Y) \cong A(Y)$.*

Proof.

(a) Assume that $Y \subset Y'$. If $f \in I(Y')$ then f vanishes on Y', hence also on Y, which means that $f \in I(Y)$. The second statement follows in a similar way.

(b) Let $x \in Y$. Then $f(x) = 0$ for every $f \in I(Y)$ by definition of $I(Y)$. But this implies that $x \in V(I(Y))$. Again, the second statement follows analogously.

(c) By (b) it suffices to prove "$\supset$". As Y is a subvariety of X we can write $Y = V(S)$ for some $S \subset A(X)$. Then $S \subset I(V(S))$ by (b), and thus $V(S) \supset V(I(V(S)))$ by (a). Replacing $V(S)$ by Y now gives the required inclusion.

(d) The ring homomorphism $A(X) \to A(Y)$ that restricts a polynomial function on X to a function Y is surjective and has kernel $I(Y)$ by definition. So the result follows from the homomorphism theorem [G1, Proposition 8.12]. □

Remark 0.10 (Reconstruction of geometry from algebra)**.** Let Y be a subvariety of X. Then Lemma 0.9 (c) says that $I(Y)$ determines Y uniquely. Similarly, knowing the rings $A(X)$ and $A(Y)$, together with the ring homomorphism $A(X) \to A(Y)$ that describes the restriction of functions on X to functions on Y, is enough to recover $I(Y)$ as the kernel of this map, and thus Y as a subvariety of X by the above. In other words, we do not lose any information if we pass from geometry to algebra and describe varieties and their subvarieties by their coordinate rings and ideals.

This map $A(X) \to A(Y)$ corresponding to the restriction of functions to a subvariety is already a first special case of a ring homomorphism associated to a "morphism of varieties". Let us now introduce this notion.

Construction 0.11 (Morphisms of varieties)**.** Let $X \subset \mathbb{A}^n_K$ and $Y \subset \mathbb{A}^m_K$ be two varieties over the same ground field. Then a **morphism** from X to Y is just a set-theoretic map $f : X \to Y$ that can be given by polynomials, i. e. such that there are polynomials $f_1, \dots, f_m \in K[x_1, \dots, x_n]$ with $f(x) = (f_1(x), \dots, f_m(x)) \in Y$ for all $x \in X$. To such a morphism we can assign a ring homomorphism

$$\varphi : A(Y) \to A(X), \quad g \mapsto g \circ f = g(f_1, \dots, f_m)$$

given by composing a polynomial function on Y with f to obtain a polynomial function on X. Note that this ring homomorphism φ ...

(a) reverses the roles of source and target compared to the original map $f : X \to Y$; and

(b) is enough to recover f, since $f_i = \varphi(y_i) \in A(X)$ if $y_1, \dots, y_m$ denote the coordinates of $\mathbb{A}^m_K$.

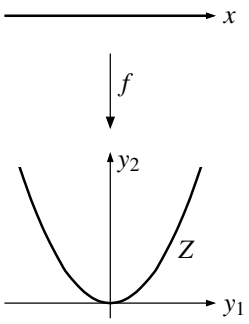
Example 0.12. Let $X = \mathbb{A}^1_{\mathbb{R}}$ (with coordinate x) and $Y = \mathbb{A}^2_{\mathbb{R}}$ (with coordinates y_1 and y_2), so that $A(X) = \mathbb{R}[x]$ and $A(Y) = \mathbb{R}[y_1, y_2]$ by Remark 0.6. Consider the morphism of varieties

$$f : X \to Y, \quad x \mapsto (y_1, y_2) := (x, x^2)$$

whose image is obviously the standard parabola $Z = V(y_2 - y_1^2)$ shown in the picture on the right. Then the associated ring homomorphism $A(Y) = \mathbb{R}[y_1, y_2] \to \mathbb{R}[x] = A(X)$ of Construction 0.11 is given by composing a polynomial function in y_1 and y_2 with f, i. e. by plugging in x and x^2 for y_1 and y_2, respectively:

$$\mathbb{R}[y_1, y_2] \to \mathbb{R}[x], \quad g \mapsto g(x, x^2).$$

Note that with the images of $g = y_1$ and $g = y_2$ under this homomorphism we just recover the polynomials x and x^2 defining the map f.



If we had considered f as a morphism from X to Z (i. e. restricted the target space to the actual image of f) we would have obtained $A(Z) = K[y_1, y_2]/(y_2 - y_1^2)$ and thus the ring homomorphism

$$\mathbb{R}[y_1, y_2]/(y_2 - y_1^2) \to \mathbb{R}[x], \quad g \mapsto g(x, x^2)$$

instead (which is obviously well-defined).

Remark 0.13 (Correspondence between geometry and algebra)**.** Summarizing what we have seen so far, we get the following first version of a dictionary between geometry and algebra:

GEOMETRY	$\longrightarrow$	ALGEBRA
variety X		*ring* $A(X)$ of (polynomial) functions on X
subvariety Y of X		*ideal* $I(Y) \trianglelefteq A(X)$ of functions on X vanishing on Y
morphism $f : X \to Y$ of varieties		*ring homomorphism* $A(Y) \to A(X)$, $g \mapsto g \circ f$

Moreover, passing from ideals to subvarieties reverses inclusions as in Lemma 0.9 (a), and we have $A(X)/I(Y) \cong A(Y)$ for any subvariety Y of X by Lemma 0.9 (d) (with the isomorphism given by restricting functions from X to Y).

We have also seen already that this assignment of algebraic to geometric objects is injective in the sense of Remark 0.10 and Construction 0.11 (b). However, not all rings, ideals, and ring homomorphisms arise from this correspondence with geometry, as we will see in Remark 1.10, Example 1.25 (b), and Remark 1.31. So although the geometric picture is very useful to visualize algebraic statements, it can usually not be used to actually prove them in the case of general rings.

Remark 0.14 (Outline of this class)**.** In order to get an idea of the sort of problems considered in commutative algebra, let us quickly list some of the main topics that we will discuss in this class.

- *Modules.* From linear algebra you know that one of the most important structures related to a field K is that of a vector space over K. If we write down the same axioms as for a vector space but relax the condition on K to allow an arbitrary ring, we obtain the algebraic structure of a module, which is equally important in commutative algebra as that of a vector space in linear algebra. We will study this in Chapter 3.
- *Localization.* If we have a ring R that is not a field, an important construction discussed in Chapter 6 is to make more elements invertible by allowing "fractions" — in the same way as one can construct the rational numbers $\mathbb{Q}$ from the integers $\mathbb{Z}$. Geometrically, we will see that this process corresponds to studying a variety locally around a point, which is why it is called "localization".
- *Decomposition into primes.* In a principal ideal domain R like the integers or a polynomial ring in one variable over a field, an important algebraic tool is the unique prime factorization of elements of R [G1, Proposition 11.9]. We will extend this concept in Chapter 8 to more general rings, and also to a "decomposition of ideals into primes". In terms of geometry, this corresponds to a decomposition of a variety into pieces that cannot be subdivided any further — e. g. writing the variety in Example 0.4 (e) as a union of a line and a plane.

- *Dimension.* Looking at Example 0.4 again it seems obvious that we should be able to assign a dimension to each variety X. We will do this by assigning a dimension to each commutative ring so that the dimension of the coordinate ring $A(X)$ can be interpreted as the geometric dimension of X (see Chapter 11). With this definition of dimension we can then prove its expected properties, e. g. that cutting down a variety by n more equations reduces its dimension by at most n (Remark 11.18).
- *Orders of vanishing.* For a polynomial $f \in K[x]$ in one variable you all know what it means that it has a zero of a certain order at a point. If we now have a different variety, say still locally diffeomorphic to a line such as e. g. the circle $X = V(x_1^2 + x_2^2 - 1) \subset \mathbb{A}^2_{\mathbb{R}}$ in Example 0.4 (a), it seems geometrically reasonable that we should still be able to define such vanishing orders of functions on X at a given point. This is in fact possible, but algebraically more complicated — we will do this in Chapter 12 and study the consequences in Chapter 13.

But before we can discuss these main topics of the class we have to start now by developing more tools to work with ideals than what you know from earlier classes.

Exercise 0.15. Show that the following subsets X of $\mathbb{A}^n_K$ are *not* varieties over K:

(a) $X = \mathbb{Z} \subset \mathbb{A}^1_{\mathbb{R}}$;

(b) $X = \mathbb{A}^1_{\mathbb{R}} \setminus \{0\} \subset \mathbb{A}^1_{\mathbb{R}}$;

(c) $X = \{(x_1, x_2) \in \mathbb{A}^2_{\mathbb{R}} : x_2 = \sin(x_1)\} \subset \mathbb{A}^2_{\mathbb{R}}$;

(d) $X = \{x \in \mathbb{A}^1_{\mathbb{C}} : |x| = 1\} \subset \mathbb{A}^1_{\mathbb{C}}$;

(e) $X = f(Y) \subset \mathbb{A}^n_{\mathbb{R}}$ for an arbitrary variety Y and a morphism of varieties $f : Y \to X$ over $\mathbb{R}$.

Exercise 0.16 (Degree of polynomials)**.** Let R be a ring. Recall that an element $a \in R$ is called a *zero-divisor* if there exists an element $b \neq 0$ with $ab = 0$ [G1, Definition 7.6 (c)], and that R is called an *(integral) domain* if no non-zero element is a zero-divisor, i. e. if $ab = 0$ for $a, b \in R$ implies $a = 0$ or $b = 0$ [G1, Definition 7.6 (d)].

We define the *degree* of a non-zero polynomial $f = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} x_1^{i_1} \cdot \dots \cdot x_n^{i_n} \in R[x_1, \dots, x_n]$ to be

$$\deg f := \max\{i_1 + \dots + i_n : a_{i_1,\dots,i_n} \neq 0\}.$$

Moreover, the degree of the zero polynomial is formally set to $-\infty$. Show that:

(a) $\deg(f \cdot g) \leq \deg f + \deg g$ for all $f, g \in R[x_1, \dots, x_n]$.

(b) Equality holds in (a) for all polynomials f and g if and only if R is an integral domain.

1. Ideals

From the "Algebraic Structures" class you already know the basic constructions and properties concerning ideals and their quotient rings [G1, Chapter 8]. For our purposes however we have to study ideals in much more detail — so this will be our goal for this and the next chapter. Let us start with some general constructions to obtain new ideals from old ones. The ideal generated by a subset M of a ring [G1, Definition 8.5] will be written as (M).

Construction 1.1 (Operations on ideals)**.** Let I and J be ideals in a ring R.

(a) The **sum** of the two given ideals is defined as usual by

$$I+J := \{a+b : a \in I \text{ and } b \in J\}.$$

It is easy to check that this is an ideal — in fact, it is just the ideal generated by $I \cup J$.

(b) It is also obvious that the **intersection** $I \cap J$ is again an ideal of R.

(c) We define the **product** of I and J as the ideal generated by all products of elements of I and J, i. e.

$$I \cdot J := (\{ab : a \in I \text{ and } b \in J\}).$$

Note that just the set of products of elements of I and J would in general not be an ideal: if we take $R = \mathbb{R}[x,y]$ and $I = J = (x,y)$, then obviously x^2 and y^2 are products of an element of I with an element of J, but their sum x^2+y^2 is not.

(d) The **quotient** of I by J is defined to be

$$I:J := \{a \in R : aJ \subset I\}.$$

Again, it is easy to see that this is an ideal.

(e) We call

$$\sqrt{I} := \{a \in R : a^n \in I \text{ for some } n \in \mathbb{N}\}$$

the **radical** of I. Let us check that this an ideal of R:

- We have $0 \in \sqrt{I}$, since $0 \in I$.
- If $a, b \in \sqrt{I}$, i. e. $a^n \in I$ and $b^m \in I$ for some $n, m \in \mathbb{N}$, then

$$(a+b)^{n+m} = \sum_{k=0}^{n+m} \binom{n+m}{k} a^k b^{n+m-k}$$

is again an element of I, since in each summand we must have that the power of a is at least n (in which case $a^k \in I$) or the power of b is at least m (in which case $b^{n+m-k} \in I$). Hence $a+b \in \sqrt{I}$.

- If $r \in R$ and $a \in \sqrt{I}$, i. e. $a^n \in I$ for some $n \in \mathbb{N}$, then $(ra)^n = r^n a^n \in I$, and hence $ra \in \sqrt{I}$.

Note that we certainly have $\sqrt{I} \supset I$. We call I a **radical ideal** if $\sqrt{I} = I$, i. e. if for all $a \in R$ and $n \in \mathbb{N}$ with $a^n \in I$ it follows that $a \in I$. This is a natural definition since the radical $\sqrt{I}$ of an arbitrary ideal I is in fact a radical ideal in this sense: if $a^n \in \sqrt{I}$ for some n, so $a^{nm} \in I$ for some m, then this obviously implies $a \in \sqrt{I}$. 01

Whether an ideal I is radical can also easily be seen from its quotient ring R/I as follows.

Definition 1.2 (Nilradical, nilpotent elements, and reduced rings)**.** Let R be a ring. The ideal

$$\sqrt{(0)} = \{a \in R : a^n = 0 \text{ for some } n \in \mathbb{N}\}$$

is called the **nilradical** of R; its elements are called **nilpotent**. If R has no nilpotent elements except 0, i. e. if the zero ideal is radical, then R is called **reduced**.

Lemma 1.3. *An ideal $I \trianglelefteq R$ is radical if and only if R/I is reduced.*

Proof. By Construction 1.1 (e), the ideal I is radical if and only if for all $a \in R$ and $n \in \mathbb{N}$ with $a^n \in I$ it follows that $a \in I$. Passing to the quotient ring R/I, this is obviously equivalent to saying that $\overline{a}^n = \overline{0}$ implies $\overline{a} = \overline{0}$, i. e. that R/I has no nilpotent elements except $\overline{0}$. □

Example 1.4 (Operations on ideals in principal ideal domains)**.** Recall that a *principal ideal domain* (or short: *PID*) is an integral domain in which every ideal is *principal*, i. e. can be generated by one element [G1, Definition 10.11]. The most prominent examples of such rings are probably *Euclidean domains*, i. e. integral domains admitting a division with remainder [G1, Definition 10.17 and Proposition 10.21], such as $\mathbb{Z}$ or $K[x]$ for a field K [G1, Example 10.18 and Proposition 10.19].

We know that any principal ideal domain R admits a unique prime factorization of its elements [G1, Proposition 11.9] — a concept that we will discuss in more detail in Chapter 8. As a consequence, all operations of Construction 1.1 can then be computed easily: if I and J are not the zero ideal we can write $I = (a)$ and $J = (b)$ for $a = p_1^{a_1} \cdot \dots \cdot p_n^{a_n}$ and $b = p_1^{b_1} \cdot \dots \cdot p_n^{b_n}$ with distinct prime elements $p_1, \dots, p_n$ and $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{N}$. Then we obtain:

(a) $I + J = (p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i = \min(a_i, b_i)$ for $i = 1, \dots, n$: another (principal) ideal contains I (resp. J) if and only if it is of the form $(p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i \leq a_i$ (resp. $c_i \leq b_i$) for all i, so the smallest ideal $I + J$ containing I and J is obtained for $c_i = \min(a_i, b_i)$;

(b) $I \cap J = (p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i = \max(a_i, b_i)$;

(c) $I \cdot J = (ab) = (p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i = a_i + b_i$;

(d) $I : J = (p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i = \max(a_i - b_i, 0)$;

(e) $\sqrt{I} = (p_1^{c_1} \cdot \dots \cdot p_n^{c_n})$ with $c_i = \min(a_i, 1)$.

In particular, we have $I + J = (1) = R$ if and only if a and b have no common prime factor, i. e. if a and b are coprime. We use this observation to define the notion of coprime ideals in general rings:

Definition 1.5 (Coprime ideals)**.** Two ideals I and J in a ring R are called **coprime** if $I + J = R$.

Example 1.6 (Operations on ideals in polynomial rings with SINGULAR)**.** In more general rings, the explicit computation of the operations of Construction 1.1 is quite complicated and requires advanced algorithmic methods that you can learn about in the "Computer Algebra" class. We will not need this here, but if you want to compute some examples in polynomial rings you can use e. g. the computer algebra system SINGULAR [S]. For example, for the ideals $I = (x^2y, xy^3)$ and $J = (x + y)$ in $\mathbb{Q}[x, y]$ the following SINGULAR code computes that $I : J = (x^2y, xy^2)$ and $\sqrt{I \cdot J} = \sqrt{I \cap J} = (x^2y + xy^2)$, and checks that $y^3 \in I + J$:

```
> LIB "primdec.lib"; // library needed for the radical
> ring R=0,(x,y),dp; // set up polynomial ring Q[x,y]
> ideal I=x2y,xy3; // means I=(x^2*y,x*y^3)
> ideal J=x+y;
> quotient(I,J); // compute (generators of) I:J
_[1]=xy2
_[2]=x2y
> radical(I*J); // compute radical of I*J
_[1]=x2y+xy2
> radical(intersect(I,J)); // compute radical of intersection
_[1]=x2y+xy2
> reduce(y3,std(I+J)); // gives 0 if and only if y^3 in I+J
0
```

In this example it turned out that $\sqrt{I \cdot J} = \sqrt{I \cap J}$. In fact, this is not a coincidence — the following lemma and exercise show that the product and the intersection of ideals are very closely related.

Lemma 1.7 (Product and intersection of ideals)**.** *For any two ideals I and J in a ring R we have*

(a) $I \cdot J \subset I \cap J$;

(b) $\sqrt{I \cdot J} = \sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}$.

Proof.

(a) It suffices to check that all generators of $I \cdot J$ lie in $I \cap J$. But for $a \in I$ and $b \in J$ it is obvious that $ab \in I \cap J$, so the result follows.

(b) We show a circular inclusion, with $\sqrt{I \cdot J} \subset \sqrt{I \cap J}$ following from (a).

If $a \in \sqrt{I \cap J}$ then $a^n \in I \cap J$ for some $n \in \mathbb{N}$, so $a^n \in I$ and $a^n \in J$, and hence $a \in \sqrt{I} \cap \sqrt{J}$. Finally, if $a \in \sqrt{I} \cap \sqrt{J}$ then $a^m \in I$ and $a^n \in J$ for some $m, n \in \mathbb{N}$, therefore $a^{m+n} \in I \cdot J$ and thus $a \in \sqrt{I \cdot J}$. □

Exercise 1.8. Let $I_1, \ldots, I_n$ be pairwise coprime ideals in a ring R. Prove that $I_1 \cdot \cdots \cdot I_n = I_1 \cap \cdots \cap I_n$.

Exercise 1.9. Show that the ideal $(x_1, \ldots, x_n) \trianglelefteq K[x_1, \ldots, x_n]$ cannot be generated by fewer than n elements. Hence in particular, the polynomial ring $K[x_1, \ldots, x_n]$ is not a principal ideal domain for $n \geq 2$.

We will see however in Remark 8.6 that these polynomial rings still admit unique prime factorizations of its elements, so that the results of Example 1.4 continue to hold for principal ideals in these rings.

Remark 1.10 (Ideals of subvarieties = radical ideals)**.** Radical ideals play an important role in geometry: if Y is a subvariety of a variety X and $f \in A(X)$ with $f^n \in I(Y)$, then $(f(x))^n = 0$ for all $x \in Y$ — but this obviously implies $f(x) = 0$ for all $x \in Y$, and hence $f \in I(Y)$. So ideals of subvarieties are always radical.

In fact, if the ground field K is *algebraically closed*, i. e. if every non-constant polynomial over K has a zero (as e. g. for $K = \mathbb{C}$), we will see in Corollary 10.14 that it is *exactly* the radical ideals in $A(X)$ that are ideals of subvarieties. So in this case there is a one-to-one correspondence

$$\begin{array}{rcl} \{\text{subvarieties of } X\} & \overset{1:1}{\longleftrightarrow} & \{\text{radical ideals in } A(X)\} \\ Y & \longmapsto & I(Y) \\ V(I) & \longmapsfrom & I. \end{array}$$

In other words, we have $V(I(Y)) = Y$ for every subvariety Y of X (which we have already seen in Lemma 0.9 (c)), and $I(V(I)) = I$ for every radical ideal $I \trianglelefteq A(X)$. In order to simplify our geometric interpretations we will therefore usually assume from now on in our geometric examples that the ground field is algebraically closed and the above one-to-one correspondence holds. Note that this will not lead to circular reasoning as we will never use these geometric examples to prove anything.

Exercise 1.11.

(a) Give a rigorous proof of the one-to-one correspondence of Remark 1.10 in the case of the ambient variety $\mathbb{A}^1_{\mathbb{C}}$, i. e. between subvarieties of $\mathbb{A}^1_{\mathbb{C}}$ and radical ideals in $A(\mathbb{A}^1_{\mathbb{C}}) = \mathbb{C}[x]$.

(b) Show that this one-to-one correspondence does not hold in the case of the ground field $\mathbb{R}$, i. e. between subvarieties of $\mathbb{A}^1_{\mathbb{R}}$ and radical ideals in $A(\mathbb{A}^1_{\mathbb{R}}) = \mathbb{R}[x]$.

Remark 1.12 (Geometric interpretation of operations on ideals)**.** Let X be a variety over an algebraically closed field, and let $A(X)$ be its coordinate ring. Assuming the one-to-one correspondence of Remark 1.10 between subvarieties of X and radical ideals in $A(X)$ we can now give a geometric interpretation of the operations of Construction 1.1:

(a) As $I + J$ is the ideal generated by $I \cup J$, we have for any two (radical) ideals $I, J \trianglelefteq A(X)$

$$\begin{aligned} V(I+J) &= \{x \in X : f(x) = 0 \text{ for all } f \in I \cup J\} \\ &= \{x \in X : f(x) = 0 \text{ for all } f \in I\} \cap \{x \in X : f(x) = 0 \text{ for all } f \in J\} \\ &= V(I) \cap V(J). \end{aligned}$$

So the intersection of subvarieties corresponds to the sum of ideals. (Note however that the sum of two radical ideals may not be radical, so strictly speaking the algebraic operation corresponding to the intersection of subvarieties is taking the sum of the ideals and then its radical.)

Moreover, as the whole space X and the empty set $\emptyset$ obviously correspond to the zero ideal (0) resp. the whole ring $(1) = A(X)$, the condition $I + J = A(X)$ that I and J are coprime translates into the intersection of $V(I)$ and $V(J)$ being empty.

(b) For any two subvarieties Y, Z of X

$$\begin{aligned} I(Y \cup Z) &= \{f \in A(X) : f(x) = 0 \text{ for all } x \in Y \cup Z\} \\ &= \{f \in A(X) : f(x) = 0 \text{ for all } x \in Y\} \cap \{f \in A(X) : f(x) = 0 \text{ for all } x \in Z\} \\ &= I(Y) \cap I(Z), \end{aligned}$$

and thus the union of subvarieties corresponds to the intersection of ideals. As the product of ideals has the same radical as the intersection by Lemma 1.7 (b), the union of subvarieties also corresponds to taking the product of the ideals (and then its radical).

(c) Again for two subvarieties Y, Z of X we have

$$\begin{aligned} I(Y \backslash Z) &= \{f \in A(X) : f(x) = 0 \text{ for all } x \in Y \backslash Z\} \\ &= \{f \in A(X) : f(x) \cdot g(x) = 0 \text{ for all } x \in Y \text{ and } g \in I(Z)\} \\ &= \{f \in A(X) : f \cdot I(Z) \subset I(Y)\} \\ &= I(Y) : I(Z), \end{aligned}$$

so taking the set-theoretic difference $Y \backslash Z$ corresponds to quotient ideals. (Strictly speaking, the difference $Y \backslash Z$ is in general not a variety, so the exact geometric operation corresponding to quotient ideals is taking the smallest subvariety containing $Y \backslash Z$.)

Summarizing, we obtain the following translation between geometric and algebraic terms:

SUBVARIETIES	$\longleftrightarrow$	IDEALS
full space		*(0)*
empty set		*(1)*
intersection		*sum*
union		*product / intersection*
difference		*quotient*
disjoint subvarieties		*coprime ideals*

Exercise 1.13. Show that the equation of ideals

$$(x^3 - x^2, x^2 y - x^2, xy - y, y^2 - y) = (x^2, y) \cap (x - 1, y - 1)$$

holds in the polynomial ring $\mathbb{C}[x, y]$. Is this a radical ideal? What is its zero locus in $\mathbb{A}^2_{\mathbb{C}}$?

As an example that links the concepts introduced so far, let us now consider the Chinese Remainder Theorem that you already know for the integers [G1, Proposition 11.22] and generalize it to arbitrary rings.

Proposition 1.14 (Chinese Remainder Theorem)**.** *Let $I_1, \ldots, I_n$ be ideals in a ring R, and consider the ring homomorphism*

$$\varphi : R \to R/I_1 \times \cdots \times R/I_n, \quad a \mapsto (\overline{a}, \ldots, \overline{a}).$$

(a) *φ is injective if and only if $I_1 \cap \cdots \cap I_n = (0)$.*

(b) *φ is surjective if and only if $I_1, \ldots, I_n$ are pairwise coprime.*

Proof.

(a) This follows immediately from $\ker \varphi = I_1 \cap \cdots \cap I_n$.

(b) "$\Rightarrow$" If φ is surjective then $(1,0,\ldots,0) \in \operatorname{im}\varphi$. In particular, there is an element $a \in R$ with $a = 1 \mod I_1$ and $a = 0 \mod I_2$. But then $1 = (1-a) + a \in I_1 + I_2$, and hence $I_1 + I_2 = R$. In the same way we see $I_i + I_j = R$ for all $i \neq j$.

"$\Leftarrow$" Let $I_i + I_j = R$ for all $i \neq j$. In particular, for $i = 2,\ldots,n$ there are $a_i \in I_1$ and $b_i \in I_i$ with $a_i + b_i = 1$, so that $b_i = 1 - a_i = 1 \mod I_1$ and $b_i = 0 \mod I_i$. If we then set $b := b_2 \cdot \cdots \cdot b_n$ we get $b = 1 \mod I_1$ and $b = 0 \mod I_i$ for all $i = 2,\ldots,n$. So $(1,0,\ldots,0) = \varphi(b) \in \operatorname{im}\varphi$. In the same way we see that the other unit generators are in the image of φ, and hence φ is surjective. □

Example 1.15.

(a) Consider the case $R = \mathbb{Z}$, and let $a_1,\ldots,a_n \in \mathbb{Z}$ be pairwise coprime. Then the residue class map

$$\varphi : \mathbb{Z} \to \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_n},\ x \mapsto (\bar{x},\ldots,\bar{x})$$

is surjective by Proposition 1.14 (b). Its kernel is $(a_1) \cap \cdots \cap (a_n) = (a)$ with $a := a_1 \cdot \cdots \cdot a_n$ by Exercise 1.8, and so by the homomorphism theorem [G1, Proposition 8.12] we obtain an isomorphism

$$\mathbb{Z}_a \to \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_n},\ \bar{x} \mapsto (\bar{x},\ldots,\bar{x}),$$

which is the well-known form of the Chinese Remainder Theorem for the integers [G1, Proposition 11.22].

(b) Let X be a variety, and let $Y_1,\ldots,Y_n$ be subvarieties of X. Recall from Remark 0.13 that for $i = 1,\ldots,n$ we have isomorphisms $A(X)/I(Y_i) \cong A(Y_i)$ by restricting functions from X to Y_i. Using the translations from Remark 1.12, Proposition 1.14 therefore states that the combined restriction map $\varphi : A(X) \to A(Y_1) \times \cdots \times A(Y_n)$ to all given subvarieties is ...

- injective if and only if the subvarieties $Y_1,\ldots,Y_n$ cover all of X;
- surjective if and only if the subvarieties $Y_1,\ldots,Y_n$ are disjoint.

In particular, if X is the disjoint union of the subvarieties $Y_1,\ldots,Y_n$, then the Chinese Remainder Theorem says that φ is an isomorphism, i. e. that giving a function on X is the same as giving a function on each of the subvarieties — which seems obvious from geometry.

In our study of ideals, let us now consider their behavior under ring homomorphisms.

Definition 1.16 (Contraction and extension)**.** Let $\varphi : R \to R'$ be a ring homomorphism.

(a) For any ideal $I \trianglelefteq R'$ the inverse image $\varphi^{-1}(I)$ is an ideal of R. We call it the **inverse image ideal** or **contraction** of I by φ, sometimes denoted I^c if it is clear from the context which morphism we consider.

(b) For $I \trianglelefteq R$ the ideal generated by the image $\varphi(I)$ is called the **image ideal** or **extension** of I by φ. It is written as $\varphi(I) \cdot R'$, or I^e if the morphism is clear from the context.

Remark 1.17.

(a) Note that for the construction of the image ideal of an ideal $I \trianglelefteq R$ under a ring homomorphism $\varphi : R \to R'$ we have to take the ideal generated by $\varphi(I)$, since $\varphi(I)$ itself is in general not yet an ideal: take e. g. $\varphi : \mathbb{Z} \to \mathbb{Z}[x]$ to be the inclusion and $I = \mathbb{Z}$. But if φ is surjective then $\varphi(I)$ is already an ideal and thus $I^e = \varphi(I)$:

- for $b_1, b_2 \in \varphi(I)$ we have $a_1, a_2 \in I$ with $b_1 = \varphi(a_1)$ and $b_2 = \varphi(a_2)$, and so $b_1 + b_2 = \varphi(a_1 + a_2) \in \varphi(I)$;
- for $b \in \varphi(I)$ and $s \in R'$ we have $a \in I$ and $r \in R$ with $\varphi(a) = b$ and $\varphi(r) = s$, and thus $sb = \varphi(ra) \in \varphi(I)$.

(b) If R is a field and $R' \neq \{0\}$ then any ring homomorphism $\varphi : R \to R'$ is injective: its kernel is 0 since an element $a \in R \backslash \{0\}$ with $\varphi(a) = 0$ would lead to the contradiction

$$1 = \varphi(1) = \varphi(a^{-1}a) = \varphi(a^{-1}) \cdot \varphi(a) = 0.$$

This is the origin of the names "contraction" and "extension", since in this case these two operations really make the ideal "smaller" and "bigger", respectively.

Remark 1.18 (Geometric interpretation of contraction and extension)**.** As in Construction 0.11, let $f : X \to Y$ be a morphism of varieties, and let $\varphi : A(Y) \to A(X), \ g \mapsto g \circ f$ be the associated map between the coordinate rings.

(a) For any subvariety $Z \subset X$ we have

$$\begin{aligned} I(f(Z)) &= \{g \in A(Y) : g(f(x)) = 0 \text{ for all } x \in Z\} \\ &= \{g \in A(Y) : \varphi(g) \in I(Z)\} \\ &= \varphi^{-1}(I(Z)), \end{aligned}$$

so taking images of varieties corresponds to the contraction of ideals.

(b) For a subvariety $Z \subset Y$ the zero locus of the extension $I(Z)^e$ by φ is

$$\begin{aligned} V(\varphi(I(Z))) &= \{x \in X : g(f(x)) = 0 \text{ for all } g \in I(Z)\} \\ &= f^{-1}(\{y \in Y : g(y) = 0 \text{ for all } g \in I(Z)\}) \\ &= f^{-1}(V(I(Z))) \\ &= f^{-1}(Z) \end{aligned}$$

by Lemma 0.9 (c). Hence, taking inverse images of subvarieties corresponds to the extension of ideals.

So we can add the following two entries to our dictionary between geometry and algebra:

SUBVARIETIES	$\longleftrightarrow$	IDEALS
image		*contraction*
inverse image		*extension*

Exercise 1.19. Let $\varphi : R \to R'$ a ring homomorphism. Prove:

(a) $I \subset (I^e)^c$ for all $I \trianglelefteq R$;

(b) $I \supset (I^c)^e$ for all $I \trianglelefteq R'$;

(c) $(IJ)^e = I^e J^e$ for all $I, J \trianglelefteq R$;

(d) $(I \cap J)^c = I^c \cap J^c$ for all $I, J \trianglelefteq R'$.

Exercise 1.20. Let $f : X \to Y$ be a morphism of varieties, and let Z and W be subvarieties of X. The geometric statements below are then obvious. Find and prove corresponding algebraic statements for ideals in rings.

(a) $f(Z \cup W) = f(Z) \cup f(W)$;

(b) $f(Z \cap W) \subset f(Z) \cap f(W)$;

(c) $f(Z \backslash W) \supset f(Z) \backslash f(W)$.

An important application of contraction and extension is that it allows an easy explicit description of ideals in quotient rings.

Lemma 1.21 (Ideals in quotient rings)**.** *Let I be an ideal in a ring R. Then contraction and extension by the quotient map $\varphi : R \to R/I$ give a one-to-one correspondence*

$$\begin{array}{ccc} \{\text{ideals in } R/I\} & \overset{1:1}{\longleftrightarrow} & \{\text{ideals } J \text{ in } R \text{ with } J \supset I\} \\ J & \longmapsto & J^c \\ J^e & \longmapsfrom & J. \end{array}$$

Proof. As the quotient map φ is surjective, we know by Remark 1.17 (a) that contraction and extension are just the inverse image and image of an ideal, respectively. Moreover, it is clear that the contraction of an ideal in R/I yields an ideal of R that contains I, and that the extension of an ideal in R gives an ideal in R/I. So we just have to show that contraction and extension are inverse to each other on the sets of ideals given in the lemma. But this is easy to check:

- For any ideal $J \trianglelefteq R/I$ we have $(J^c)^e = \varphi(\varphi^{-1}(J)) = J$ since φ is surjective.
- For any ideal $J \trianglelefteq R$ with $J \supset I$ we get
$$(J^e)^c = \varphi^{-1}(\varphi(J)) = \{a \in R : \varphi(a) \in \varphi(J)\} = J + I = J.$$ □

Exercise 1.22. Let $I \subset J$ be ideals in a ring R. By Lemma 1.21, the extension J/I of J by the quotient map $R \to R/I$ is an ideal in R/I. Prove that

$$(R/I)\,/\,(J/I) \cong R/J.$$

02

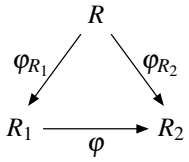
At the end of this chapter, let us now consider ring homomorphisms from a slightly different point of view that will also tell us which rings "come from geometry", i. e. can be written as coordinate rings of varieties.

Definition 1.23 (Algebras and algebra homomorphisms)**.** Let R be a ring.

(a) An **R-algebra** is a ring R' together with a ring homomorphism $\varphi_{R'} : R \to R'$.

(b) Let R_1 and R_2 be R-algebras with corresponding ring homomorphisms $\varphi_{R_1} : R \to R_1$ and $\varphi_{R_2} : R \to R_2$. A **morphism** or **$R$-algebra homomorphism** from R_1 to R_2 is a ring homomorphism $\varphi : R_1 \to R_2$ with $\varphi \circ \varphi_{R_1} = \varphi_{R_2}$.

It is often helpful to draw these maps in a diagram as shown on the right. Then the condition $\varphi \circ \varphi_{R_1} = \varphi_{R_2}$ just states that this diagram *commutes*, i. e. that any two ways along the arrows in the diagram having the same source and target — in this case the two ways to go from R to R_2 — will give the same map.



(c) Let R' be an R-algebra with corresponding ring homomorphism $\varphi_{R'} : R \to R'$. An **R-subalgebra** of R' is a subring $\tilde{R}$ of R' containing the image of φ. Note that $\tilde{R}$ is then an R-algebra using the ring homomorphism $\varphi_{\tilde{R}} : R \to \tilde{R}$ given by $\varphi_{R'}$ with the target restricted to $\tilde{R}$. Moreover, the inclusion $\tilde{R} \to R'$ is an R-algebra homomorphism in the sense of (b).

In most of our applications, the ring homomorphism $\varphi_{R'} : R \to R'$ needed to define an R-algebra R' will be clear from the context, and we will write the R-algebra simply as R'. In fact, in many cases it will even be injective. In this case we usually consider R as a subring of R', drop the homomorphism $\varphi_{R'}$ in the notation completely, and say that $R \subset R'$ is a *ring extension*. We will consider these ring extensions in detail in Chapter 9.

Remark 1.24. The ring homomorphism $\varphi_{R'} : R \to R'$ associated to an R-algebra R' can be used to define a "scalar multiplication" of R on R' by

$$R \times R' \to R', \quad (a,c) \mapsto a \cdot c := \varphi_{R'}(a) \cdot c.$$

Note that by setting $c = 1$ this scalar multiplication determines $\varphi_{R'}$ back. So one can also think of an R-algebra as a ring together with a scalar multiplication with elements of R that has the expected compatibility properties. In fact, one could also define R-algebras in this way.

Example 1.25.

(a) Without doubt the most important example of an algebra over a ring R is the polynomial ring $R[x_1,\dots,x_n]$, together with the obvious injective ring homomorphism $R \to R[x_1,\dots,x_n]$ that embeds R into the polynomial ring as constant polynomials. In the same way, any quotient $R[x_1,\dots,x_n]/I$ of the polynomial ring by an ideal I is an R-algebra as well.

(b) Let $X \subset \mathbb{A}^n_K$ be a variety over a field K. Then its coordinate ring $A(X) = K[x_1,\dots,x_n]/I(X)$ is a K-algebra by (a), with K mapping to $A(X)$ as the constant functions. Moreover, the ring homomorphism $A(Y) \to A(X)$ of Construction 0.11 corresponding to a morphism $f : X \to Y$ to another variety Y is a K-algebra homomorphism, since composing a constant function with f gives again a constant function. In fact, one can show that these are precisely the maps between the coordinate rings coming from morphisms of varieties, i. e. that Construction 0.11 gives a one-to-one correspondence

$$\{\text{morphisms } X \to Y\} \overset{1:1}{\longleftrightarrow} \{K\text{-algebra homomorphisms } A(Y) \to A(X)\}.$$

Definition 1.26 (Generated subalgebras)**.** Let R' be an R-algebra.

(a) For any subset $M \subset R'$ let

$$R[M] := \bigcap_{\substack{T \supset M \\ R\text{-subalgebra of } R'}} T$$

be the smallest R-subalgebra of R' that contains M. We call it the R-subalgebra **generated by** M. If $M = \{c_1,\dots,c_n\}$ is finite, we write $R[M] = R[\{c_1,\dots,c_n\}]$ also as $R[c_1,\dots,c_n]$.

(b) We say that R' is a **finitely generated** R-algebra if there are finitely many $c_1,\dots,c_n$ with $R[c_1,\dots,c_n] = R'$.

Remark 1.27. Note that the square bracket notation in Definition 1.26 is ambiguous: $R[x_1,\dots,x_n]$ can either mean the polynomial ring over R as in Definition 0.2 (if $x_1,\dots,x_n$ are formal variables), or the subalgebra of an R-algebra R' generated by $x_1,\dots,x_n$ (if $x_1,\dots,x_n$ are elements of R'). Unfortunately, the usage of the notation $R[x_1,\dots,x_n]$ for both concepts is well-established in the literature, so we will adopt it here as well. Its origin lies in the following lemma, which shows that the elements of an R-subalgebra generated by a set M are just the polynomial expressions in elements of M with coefficients in R.

Lemma 1.28 (Explicit description of $R[M]$)**.** *Let M be a subset of an R-algebra R'. Then*

$$R[M] = \left\{ \sum_{i_1,\dots,i_n \in \mathbb{N}} a_{i_1,\dots,i_n}\, c_1^{i_1} \cdot \dots \cdot c_n^{i_n} : a_{i_1,\dots,i_n} \in R,\, c_1,\dots,c_n \in M, \text{ only finitely many } a_{i_1,\dots,i_n} \neq 0 \right\},$$

where multiplication in R' with elements of R is defined as in Remark 1.24.

Proof. It is obvious that this set of polynomial expressions is an R-subalgebra of R'. Conversely, every R-subalgebra of R' containing M must also contain these polynomial expressions, so the result follows. □

Example 1.29. In the field $\mathbb{C}$ of complex numbers the $\mathbb{Z}$-algebra generated by the imaginary unit i is

$$\mathbb{Z}[i] = \{f(i) : f \in \mathbb{Z}[x]\} = \{a + b\,i : a,b \in \mathbb{Z}\} \quad \subset \mathbb{C}$$

by Lemma 1.28. (Note again the double use of the square bracket notation: $\mathbb{Z}[x]$ is the polynomial ring over $\mathbb{Z}$, whereas $\mathbb{Z}[i]$ is the $\mathbb{Z}$-subalgebra of $\mathbb{C}$ generated by i.)

Lemma 1.30 (Finitely generated R-algebras)**.** *An algebra R' over a ring R is finitely generated if and only if $R' \cong R[x_1,\dots,x_n]/I$ for some $n \in \mathbb{N}$ and an ideal I in the polynomial ring $R[x_1,\dots,x_n]$.*

Proof. Certainly, $R[x_1,\dots,x_n]/I$ is a finitely generated R-algebra since it is generated by the classes of $x_1,\dots,x_n$. Conversely, let R' be an R-algebra generated by $c_1,\dots,c_n \in S$. Then

$$\varphi : R[x_1,\dots,x_n] \to R',\ \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n}\, x_1^{i_1}\cdot\dots\cdot x_n^{i_n} \mapsto \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n}\, c_1^{i_1}\cdot\dots\cdot c_n^{i_n}$$

is a ring homomorphism, and its image is precisely $R[c_1,\dots,c_n] = R'$ by Lemma 1.28. So by the homomorphism theorem [G1, Proposition 8.12] φ induces a ring isomorphism $R[x_1,\dots,x_n]/\ker\varphi \cong R'$, which by construction is also an R-algebra isomorphism. □

Remark 1.31 (Coordinate rings = reduced finitely generated K-algebras)**.** Let K be an algebraically closed field. Then by Remark 1.10 the coordinate rings of varieties over K are exactly the rings of the form $K[x_1,\dots,x_n]/I$ for a radical ideal $I \trianglelefteq K[x_1,\dots,x_n]$, so by Lemma 1.3 and Lemma 1.30 precisely the reduced finitely generated K-algebras.

2. Prime and Maximal Ideals

There are two special kinds of ideals that are of particular importance, both algebraically and geometrically: the so-called prime and maximal ideals. Let us start by defining these concepts.

Definition 2.1 (Prime and maximal ideals)**.** Let I be an ideal in a ring R with $I \neq R$.

(a) I is called a **prime ideal** if for all $a, b \in R$ with $ab \in I$ we have $a \in I$ or $b \in I$. By induction, this is obviously the same as saying that for all $a_1, \dots, a_n \in R$ with $a_1 \cdot \dots \cdot a_n \in I$ one of the a_i must be in I.

(b) I is called a **maximal ideal** if there is no ideal J with $I \subsetneq J \subsetneq R$.

(c) The set of all prime ideals of R is called the **spectrum**, the set of all maximal ideals the **maximal spectrum** of R. We denote these sets by $\operatorname{Spec} R$ and $\operatorname{mSpec} R$, respectively.

Remark 2.2. Let $R \neq \{0\}$ be a ring. Note that its two trivial ideals R and (0) are treated differently in Definition 2.1:

(a) The whole ring R is by definition never a prime or maximal ideal. In fact, maximal ideals are just defined to be the inclusion-maximal ones among all ideals that are not equal to R.

(b) The zero ideal (0) may be prime or maximal if the corresponding conditions are satisfied. More precisely, by definition (0) is a prime ideal if and only if R is an integral domain [G1, Definition 7.6 (d)], and it is maximal if and only if there are no ideals except (0) and R, i. e. if R is a field [G1, Example 8.8 (c)].

In fact, there is a similar criterion for arbitrary ideals if one passes to quotient rings:

Lemma 2.3. *Let I be an ideal in a ring R with $I \neq R$.*

(a) *I is a prime ideal if and only if R/I is an integral domain.*

(b) *I is a maximal ideal if and only if R/I is a field.*

Proof.

(a) Passing to the quotient ring R/I, the condition of Definition 2.1 (a) says that I is prime if and only if for all $\overline{a}, \overline{b} \in R/I$ with $\overline{a}\,\overline{b} = \overline{0}$ we have $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$, i. e. if and only if R/I is an integral domain.

(b) By Lemma 1.21, the condition of Definition 2.1 (b) means exactly that the ring R/I has only the trivial ideals I/I and R/I, which is equivalent to R/I being a field [G1, Example 8.8 (c)]. □

We can view this lemma as being analogous to Lemma 1.3, which asserted that I is a radical ideal if and only if R/I is reduced. The fact that these properties of ideals are reflected in their quotient rings has the immediate consequence that they are preserved under taking quotients as in Lemma 1.21:

Corollary 2.4. *Let $I \subset J$ be ideals in a ring R. Then J is radical / prime / maximal in R if and only if J/I is radical / prime / maximal in R/I.*

Proof. By Lemma 1.3, the ideal J is radical in R if and only if R/J is reduced, and J/I is radical in R/I if and only if $(R/I)/(J/I)$ is reduced. But these two rings are isomorphic by Exercise 1.22, so the result follows.

The statement about prime and maximal ideals follows in the same way, using Lemma 2.3 instead of Lemma 1.3. □

Corollary 2.5. *Every maximal ideal in a ring is prime, and every prime ideal is radical.*

Proof. Passing to the quotient ring, this follows immediately from Lemma 1.3 and Lemma 2.3 since a field is an integral domain and an integral domain is reduced. □

Example 2.6.

(a) Let R be an integral domain, and let $p \in R\backslash\{0\}$ not be a *unit*, i. e. it does not have a multiplicative inverse in R [G1, Definition 7.6 (a)]. Then by definition the ideal (p) is prime if and only if for all $a, b \in R$ with $p \mid ab$ we have $p \mid a$ or $p \mid b$, i. e. by definition if and only if p is a *prime element* of R [G1, Definition 11.1 (b)]. Of course, this is the origin of the name "prime ideal".

(b) We claim that for non-zero ideals in a principal ideal domain R the notions of prime and maximal ideals agree. To see this, is suffices by Corollary 2.5 to show that every non-zero prime ideal is maximal. So let $I \trianglelefteq R$ be prime. Of course, we have $I = (p)$ for some $p \in R$ as R is a principal ideal domain, and since $I \neq R$ by definition and $I \neq 0$ by assumption we know by (a) that p is prime. Now if $J \supset I$ is another ideal we must have $J = (q)$ for some $q \mid p$. But p is prime and thus *irreducible*, i. e. it cannot be written as a product of two non-units in R [G1, Definition 11.1 (a) and Lemma 11.3]. So up to a unit q must be 1 or p. But then $J = R$ or $J = I$, respectively, which means that I must have been maximal.

(c) Let K be a field, and consider the ideal $I(a) = (x_1 - a_1, \ldots, x_n - a_n) \trianglelefteq K[x_1, \ldots, x_n]$ of a for a given point $a = (a_1, \ldots, a_n) \in \mathbb{A}^n_K$ as in Example 0.7. Then the ring homomorphism

$$K[x_1, \ldots, x_n]/I(a) \to K, \ \overline{f} \mapsto f(a)$$

is obviously an isomorphism, since $f \in I(a)$ is by definition equivalent to $f(a) = 0$. So $K[x_1, \ldots, x_n]/I(a) \cong K$ is a field, and thus by Lemma 2.3 (b) the ideal $I(a)$ is maximal.

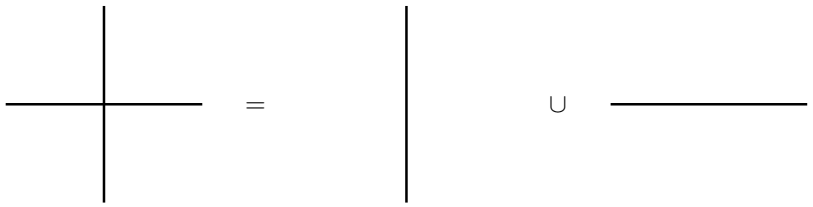
For general fields, not all maximal ideals of $K[x_1, \ldots, x_n]$ have to be of this form. For example, the ideal $(x^2 + 1) \trianglelefteq \mathbb{R}[x]$ is also maximal by (a) and (b), since the real polynomial $x^2 + 1$ is irreducible and thus prime in $\mathbb{R}[x]$ [G1, Proposition 11.5]. But if K is algebraically closed, we will see in Corollary 10.10 that the ideals considered above are the only maximal ideals in the polynomial ring. In fact, it is easy to see that we would expect this if we look at the following geometric interpretation of maximal ideals.

Remark 2.7 (Geometric interpretation of prime and maximal ideals)**.** Let X be a variety over an algebraically closed field K, so that we have a one-to-one correspondence between subvarieties of X and radical ideals in $A(X)$ by Remark 1.10.

(a) As the correspondence between subvarieties and ideals reverses inclusions, the maximal ideals of $A(X)$ correspond to minimal subvarieties of X, i. e. to points of X. For example, we have just seen in Example 2.6 (c) that the maximal ideal $(x_1 - a_1, \ldots, x_n - a_n) \trianglelefteq K[x_1, \ldots, x_n]$ is the ideal $I(a)$ of the point $a = (a_1, \ldots, a_n) \in \mathbb{A}^n_K$.

(b) Let Y be a non-empty subvariety of X, corresponding to a proper ideal $I(Y) \trianglelefteq A(X)$.

If $I(Y)$ is *not* prime then there are functions $f_1, f_2 \in A(X)$ such that $f_1 \cdot f_2$ vanishes on Y, but f_1 and f_2 do not. Hence the zero loci $Y_1 := V_Y(f_1)$ and $Y_2 := V_Y(f_2)$ of f_1 and f_2 on Y are not all of Y, but their union is $Y_1 \cup Y_2 = V_Y(f_1 f_2) = Y$. So we can write Y as a non-trivial union of two subvarieties. If a variety has this property we call it *reducible*, otherwise *irreducible*. As shown in the picture below, the union $Y = V(x_1 x_2)$ of the two coordinate axes in $\mathbb{A}^2_{\mathbb{R}}$ is a typical example of a reducible variety, with $f_1 = x_1$ and $f_2 = x_2$ in the notation above.



$Y = V(x_1 x_2)$ $Y_1 = V(x_1)$ $Y_2 = V(x_2)$

Conversely, if Y is reducible with $Y = Y_1 \cup Y_2$ for two proper subvarieties Y_1 and Y_2, we can find $f_1 \in I(Y_1) \backslash I(Y)$ and $f_2 \in I(Y_2) \backslash I(Y)$. Then $f_1 f_2 \in A(X)$ vanishes on Y although f_1 and f_2 do not, and thus $I(Y)$ is not a prime ideal.

Summarizing, we get the following correspondence:

SUBVARIETIES	$\longleftrightarrow$	IDEALS
irreducible		*prime ideal*
point		*maximal ideal*

Exercise 2.8. Which of the following ideals are prime, which ones are maximal in $\mathbb{Z}[x]$?

$$I = (5, x^3 + 2x + 3) \qquad J = (4, x^2 + x + 1, x^2 + x - 1)$$

Exercise 2.9. Let $\varphi : R \to S$ be a ring homomorphism, and let $I \trianglelefteq S$. Show that:

(a) If I is radical, then so is $\varphi^{-1}(I)$.

(b) If I is prime, then so is $\varphi^{-1}(I)$.

(c) If I is maximal, then $\varphi^{-1}(I)$ need not be maximal.

Exercise 2.10. Let R be a ring.

(a) Let $I_1, \dots, I_n \trianglelefteq R$, and let $P \trianglelefteq R$ be a prime ideal. If $P \supset I_1 \cap \dots \cap I_n$, prove that $P \supset I_k$ for some $k = 1, \dots, n$.

(b) Let $I \trianglelefteq R$, and let $P_1, \dots, P_n \trianglelefteq R$ be prime ideals. If $I \subset P_1 \cup \dots \cup P_n$, prove that $I \subset P_k$ for some $k = 1, \dots, n$.

(c) Show that the statement of (b) still holds if P_1 is not necessarily prime (but $P_2, \dots, P_n$ still are).

Can you give a geometric interpretation of these statements?

Exercise 2.11. Let R be the ring of all continuous real-valued functions on the unit interval $[0,1]$. Similarly to Definition 0.3 (c), for any subset S of R we denote by

$$V(S) := \{a \in [0,1] : f(a) = 0 \text{ for all } f \in S\} \quad \subset [0,1]$$

the zero locus of S. Prove:

(a) For all $a \in [0,1]$ the ideal $I_a := \{f \in R : f(a) = 0\}$ is maximal.

(b) If $f_1, \dots, f_n \in R$ with $V(f_1, \dots, f_n) = \emptyset$, then $f_1^2 + \dots + f_n^2$ is invertible in R.

(c) For any ideal $I \trianglelefteq R$ with $I \neq R$ we have $V(I) \neq \emptyset$.

(d) The assignment $[0,1] \to \operatorname{mSpec} R$, $a \mapsto I_a$ gives a one-to-one correspondence between points in the unit interval and maximal ideals of R (compare this to Remark 2.7 (a)).

Exercise 2.12. Let R be a ring such that for all $a \in R$ there is a natural number $n > 1$ with $a^n = a$.

(a) Show that every prime ideal of R is maximal.

(b) Give an example of such a ring which is neither a field nor the zero ring.

We have now studied prime and maximal ideals in some detail, but left out one important question: whether such ideals actually always exist. More precisely, if I is an ideal in a ring R with $I \neq R$, is it guaranteed that there is a maximal (and hence also prime) ideal that contains I? In particular, taking $I = (0)$, is it clear that there is any maximal ideal in R at all? From a geometric point of view this seems to be trivial: using the translations of Remark 2.7 (a) we are just asking whether a non-empty variety always contains a point — which is of course true. But we know already that not every ring is the coordinate ring of a variety, so for general rings we have to find an algebraic argument that ensures the existence of maximal ideals. It turns out that this is rather tricky, so let us start by giving the idea how such maximal ideals could be found.

Remark 2.13 (Naive strategy to find maximal ideals)**.** Let I be an ideal in a ring R with $I \neq R$; we want to find a maximal ideal that contains I. Set $I_0 := I$. Of course, there is nothing to be done if I_0 is already maximal, so let us assume that it is not. Then there is an ideal I_1 with $I_0 \subsetneq I_1 \subsetneq R$. Now if I_1 is maximal we are done, otherwise we find an ideal I_2 with $I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq R$. Continuing this process, we either find a maximal ideal containing I after a finite number of steps, or arrive at an infinite strictly increasing chain

$$I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq \cdots$$

of ideals in R.

Is it possible that such an infinite chain of ideals exists? In general, the answer to this question is yes (see e. g. Example 7.2 (c)). There is a special and very important class of rings however — the *Noetherian rings* that we will discuss in Chapter 7 — that do not admit such infinite increasing chains of ideals. In these rings the above process must terminate, and so we can easily find a maximal ideal containing I. In fact, all coordinate rings of varieties turn out to be Noetherian (see Remark 7.15), which is why our geometric picture above suggested that the existence of maximal ideals might be trivial. 03

But what can we do if the chain above does not terminate? The main observation is that we can then form a "limit"

$$I_\infty := \bigcup_{n \in \mathbb{N}} I_n$$

over all previous ideals (it is easy to see that this is in fact an ideal which is not equal to R; we will check this in the proof of Corollary 2.17). Of course, I_∞ is strictly bigger than every I_n. So it is an ideal that we have not seen before in the chain, and consequently we can continue the above process with this limit ideal: if it is maximal we are finished, otherwise choose a bigger one which we might call $I_{\infty+1}$, and so on. Unless we find a maximal ideal at some point, we can thus construct another infinite increasing chain starting with I_∞, take the limit over this chain again, repeat the whole process of constructing a chain and taking its limit infinitely many times, take the limit over all these infinitely many limit ideals, and so on. Intuitively speaking, we obtain an increasing "chain" of ideals in this way that is *really* long, certainly not indexed any more by the natural numbers, and not even countable. Continuing our sloppy wording, we could hope that eventually we would have to obtain even more ideals in this way than there are subsets of R at all. This would of course be a contradiction, suggesting that the process must stop at some point and give us a maximal ideal.

In fact, we will now make this argument precise. The key step in the proof is *Zorn's Lemma*, which abstracts from the above situation of rings and ideals and roughly states that in any sort of structure where you can compare objects and the above "limiting process over chains" works to get something bigger than what you had before, you will find a maximal object. To be able to formulate this rigorously we need some definitions regarding orders on sets.

Definition 2.14 (Orders)**.** Let $\leq$ be a relation on a set M.

(a) We call $\leq$ a **partial order** on M if:
  (i) $a \leq a$ for all $a \in M$ (we say $\leq$ is *reflexive*);
  (ii) for all $a, b, c \in M$ with $a \leq b$ and $b \leq c$ we have $a \leq c$ (we say $\leq$ is *transitive*);
  (iii) for all $a, b \in M$ with $a \leq b$ and $b \leq a$ we have $a = b$ (we say $\leq$ is *antisymmetric*).

(b) If in addition $a \leq b$ or $b \leq a$ for all $a, b \in M$, then $\leq$ is called a **total order** on M.

(c) An element $b \in M$ is an **upper bound** for a subset $A \subset M$ if $a \leq b$ for all $a \in A$.

(d) An element $a \in M$ is called **maximal** if for all $b \in M$ with $a \leq b$ we have $b = a$.

For a partial order $\leq$ we write $a < b$ if $a \leq b$ and $a \neq b$. A set M together with a partial or total order is called a **partially** or **totally ordered set**, respectively.

Example 2.15.

(a) The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ with their standard order are all totally ordered sets.

(b) Certainly the most important example of a partial order (in fact, probably our only example) is the set-theoretic inclusion, where M is any family of sets. Note that this is in general not a total order since for two sets $A, B \in M$ it might of course happen that neither $A \subset B$ nor $B \subset A$ holds. But if we have a chain $A_0 \subsetneq A_1 \subsetneq A_2 \subsetneq \cdots$ of sets (or even a "longer chain" as in Remark 2.13) then the family $\{A_0, A_1, A_2, \dots\}$ of all sets in this chain is totally ordered by inclusion. So totally ordered sets will be our generalization of chains to families that are not necessarily indexed by the natural numbers.

(c) If M is the set of all ideals $I \neq R$ in a ring R, partially ordered by inclusion as in (b), then the maximal elements of M are by definition exactly the maximal ideals of R.

(d) In contrast to the case of total orders, a maximal element a in a partially ordered set M need not be an upper bound for M, because a might not be comparable to some of the elements of M.

With this notation we can thus translate our informal condition in Remark 2.13 that "the limiting process works" into the rigorous requirement that every totally ordered set should have an upper bound. If this condition is met, Zorn's Lemma guarantees the existence of a maximal element and thus makes the intuitive argument of Remark 2.13 precise:

Proposition 2.16 (Zorn's Lemma)**.** *Let M be a partially ordered set in which every totally ordered subset has an upper bound. Then M has a maximal element.*

Before we prove Zorn's Lemma, let us first apply it to the case of ideals in rings to see that it really solves our existence problem for maximal ideals.

Corollary 2.17 (Existence of maximal ideals)**.** *Let I be an ideal in a ring R with $I \neq R$. Then I is contained in a maximal ideal of R.*

In particular, every ring $R \neq 0$ has a maximal ideal.

Proof. Let M be the set of all ideals $J \trianglelefteq R$ with $J \supset I$ and $J \neq R$. By Example 2.15 (c), the maximal ideals of R containing I are exactly the maximal elements of M, and hence by Zorn's Lemma it suffices to show that every totally ordered subset of M has an upper bound.

So let $A \subset M$ be a totally ordered subset, i. e. a family of proper ideals of R containing I such that, for any two of these ideals, one is contained in the other. If $A = \emptyset$ then we can just take $I \in M$ as an upper bound for A. Otherwise, let

$$J' := \bigcup_{J \in A} J$$

be the union of all ideals in A. We claim that this is an ideal:

- $0 \in J'$, since 0 is contained in each $J \in A$, and A is non-empty.
- If $a_1, a_2 \in J'$, then $a_1 \in J_1$ and $a_2 \in J_2$ for some $J_1, J_2 \in A$. But A is totally ordered, so without loss of generality we can assume that $J_1 \subset J_2$. It follows that $a_1 + a_2 \in J_2 \subset J'$.
- If $a \in J'$, i. e. $a \in J$ for some $J \in A$, then $ra \in J \subset J'$ for any $r \in R$.

Moreover, J' certainly contains I, and we must have $J' \neq R$ since $1 \notin J$ for all $J \in A$, so that $1 \notin J'$. Hence $J' \in M$, and it is certainly an upper bound for A. Thus, by Zorn's Lemma, M has a maximal element, i. e. there is a maximal ideal in R containing I. □

So to complete our argument we have to give a proof of Zorn's Lemma. However, as most textbooks using Zorn's Lemma do not prove it but rather say that it is simply an axiom of set theory, let us first explain shortly in what sense we can prove it.

Remark 2.18 (Zorn's Lemma and the Axiom of Choice)**.** As you know, essentially all of mathematics is built up on the notion of sets. Nevertheless, if you remember your first days at university, you were not given precise definitions of what sets actually are and what sort of operations you can do

with them. One usually just uses the informal statement that a set is a "collection of distinct objects" and applies common sense when dealing with them.

Although this approach is good to get you started, it is certainly not satisfactory from a mathematically rigorous point of view. In fact, it is even easy to obtain contradictions (such as Russell's Paradox [G2, Remark 1.14]) if one applies common sense too naively when working with sets. So one needs strict axioms for set theory — the ones used today were set up by Zermelo and Fraenkel around 1930 — that state exactly which operations on sets are allowed. We do not want to list all these axioms here, but as a first approximation one can say that one can always construct new sets from old ones, whereas "circular definitions" (that try e. g. to construct a set that contains itself as an element) are forbidden.

Of course, the idea of these axioms is that they are all "intuitively obvious", so that nobody will have problems to accept them as the foundation for all of mathematics. One of them is the so-called *Axiom of Choice*; it states that if you have a collection of non-empty sets you can simultaneously choose an element from each of them (even if you do not have a specific rule to make your choice). For example, if you want to prove that a surjective map $f : A \to B$ has a right-sided inverse, i. e. a map $g : B \to A$ with $f \circ g = \mathrm{id}_B$, you need to apply the Axiom of Choice since you have to construct g by simultaneously choosing an inverse image of every element of B under f. In a similar way you have probably used the Axiom of Choice many times already without knowing about it — simply because it seems intuitively obvious.

Now it happens that Zorn's Lemma is in fact equivalent to the Axiom of Choice. In other words, if we removed the Axiom of Choice from the axioms of set theory we could actually prove it if we took Zorn's Lemma as an axiom instead. But nobody would want to do this since the statement of the Axiom of Choice is intuitively clear, whereas Zorn's Lemma is certainly not. So it seems a bit cheated not to prove Zorn's Lemma because it could be taken as an axiom.

Having said all this, what we want to do now is to assume the Axiom of Choice (which you have done in all your mathematical life anyway) and to prove Zorn's Lemma with it. To do this, we need one more notion concerning orders.

Definition 2.19 (Well-ordered sets)**.** A totally ordered set M is called **well-ordered** if every non-empty subset A of M has a minimum, i. e. an element $a \in A$ such that $a \leq b$ for all $b \in A$.

Example 2.20.

(a) Any finite totally ordered set is well-ordered. Every subset of a well-ordered set is obviously well-ordered, too.

(b) The set $\mathbb{N}$ of natural numbers is well-ordered with its standard order, whereas $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are not. Instead, to construct "bigger" well-ordered sets than $\mathbb{N}$ one has to add "infinite elements": the set $\mathbb{N} \cup \{\infty\}$ (obtained from $\mathbb{N}$ by adding one more element which is defined to be bigger than all the others) is well-ordered. One can go on like this and obtain well-ordered sets $\mathbb{N} \cup \{\infty, \infty + 1\}$, $\mathbb{N} \cup \{\infty, \infty + 1, \infty + 2\}$, and so on.

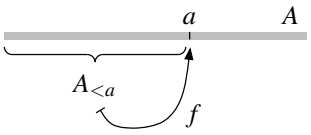
It can be seen from these examples already that well-ordered sets are quite similar to the chains of ideals constructed in Remark 2.13. In fact, the idea of Remark 2.13 should be clearly visible in the following proof of Zorn's Lemma, in which our chains of ideals correspond to the f-sets introduced below, and the choice of a new bigger ideal that extends a previously constructed chain is given by the function f.

Proof of Proposition 2.16 (Zorn's Lemma). Let M be a partially ordered set in which every *well-ordered* subset has an upper bound (this is all we will need — so we could in fact weaken the assumption of Proposition 2.16 in this way). We will prove Zorn's Lemma by contradiction, so assume that M has no maximal element.

For any well-ordered subset $A \subset M$ there is then an upper bound which cannot be maximal, and so we can find an element of M which is even bigger, and thus *strictly* bigger than all elements of A. Choose such an element and call it $f(A)$ — we can thus consider f as a function from the set of all

well-ordered subsets of M to M. (Actually, this is the point where we apply the Axiom of Choice as explained in Remark 2.18.)

Let us call a subset $A \subset M$ an *f-set* (we choose this name to indicate that this notion depends on the choice of f made above) if it is well-ordered and satisfies $a = f(A_{<a})$ for all $a \in A$, where we used the obvious notation $A_{<a} := \{b \in A : b < a\}$. Drawing A symbolically as a line (since it is a totally ordered set after all), one can visualize this condition as in the picture on the right.



Intuitively, an f-set is thus determined at each point $a \in A$ by its predecessors in A by applying f. For example:

- If an f-set A has only finitely many elements $a_1 < \cdots < a_n$, we must have $a_1 = f(\emptyset)$ and $a_i = f(\{a_1, \ldots, a_{i-1}\})$ for $i = 2, \ldots, n$.
- If A is an f-set, then $A \cup \{f(A)\}$ is also an f-set, obtained by "adding one element at the end" — and this is in fact the only element we could add at the end to obtain a new f-set.

In particular, we would expect that, although any two f-sets A and B might have different lengths, they should contain the same elements up to the point where one of them ends — so they should look like the picture below on the left, and not like the one on the right. Let us prove this rigorously:

Claim: If A and B are two f-sets and there is an element $b_0 \in B \backslash A$, then $A \subset B$ and b_0 is bigger than all elements of A.

To prove this, let C be the union of all subsets of $A \cap B$ with the property that with any element they also contain all smaller elements in $A \cup B$ — let us call such subsets of $A \cap B$ *saturated*. Of course, C is then saturated as well, so it is obviously the biggest saturated set. We can think of it as the part where A and B still coincide, as in the picture above.

By construction, it is clear that $C \subset A$. If we had $C \neq A$, then $A \backslash C$ and $B \backslash C$ would be non-empty (the latter because it contains b_0), and so there are $a = \min(A \backslash C)$ and $b = \min(B \backslash C)$ since A and B are well-ordered. Then $A_{<a} = B_{<b} = C$ by construction, and so $a = f(A_{<a}) = f(B_{<b}) = b$ as A and B are f-sets. But this means that $C \cup \{a\}$ is a bigger saturated set than C, which is a contradiction and shows that we must have $C = A$. So $A = C$ is a subset of B, and b_0 is bigger than all elements of $C = A$, proving our claim.

Now let D be the union of all f-sets. Then every $a \in D$ is contained in an f-set A, and by our claim all elements of $D \backslash A$ (which must be contained in another f-set) are bigger than a. Hence D is an f-set as well:

- D is totally ordered (an element a of an f-set A is smaller than all elements of $D \backslash A$, and can be compared to all other elements of A since A is totally ordered);
- a minimum of any non-empty subset $D' \subset D$ can be found in any f-set A with $A \cap D' \neq \emptyset$, since the other elements of D' are bigger anyway — so D is well-ordered;
- for any $a \in D$ in an f-set A we have $f(D_{<a}) = f(A_{<a}) = a$.

So D is the biggest f-set of M. But $D \cup \{f(D)\}$ is an even bigger f-set, which is a contradiction. Hence M must have a maximal element. □

As another application of Zorn's lemma, let us prove a formula for the radical of an ideal in terms of prime ideals.

Lemma 2.21. *For every ideal I in a ring R we have*

$$\sqrt{I} = \bigcap_{\substack{P \text{ prime} \\ P \supset I}} P.$$

Proof.

"$\subset$" If $a \in \sqrt{I}$ then $a^n \in I$ for some n. But then also $a^n \in P$ for every prime ideal $P \supset I$, which implies $a \in P$ by Definition 2.1 (a).

"$\supset$" Let $a \in R$ with $a \notin \sqrt{I}$, i. e. $a^n \notin I$ for all $n \in \mathbb{N}$. Consider the set

$$M = \{J : J \trianglelefteq R \text{ with } J \supset I \text{ and } a^n \notin J \text{ for all } n \in \mathbb{N}\}.$$

In the same way as in the proof of Corollary 2.17 we see that every totally ordered subset of M has an upper bound (namely I if the subset is empty, and the union of all ideals in the subset otherwise). Hence by Proposition 2.16 there is a maximal element P of M. It suffices to prove that P is prime, for then we have found a prime ideal $P \supset I$ with $a \notin P$, so that a does not lie in the right hand side of the equation of the lemma.

So assume that we have $b, c \in R$ with $bc \in P$, but $b \notin P$ and $c \notin P$. Then $P + (b)$ and $P + (c)$ are strictly bigger than P, and thus by maximality cannot lie in M. This means that there are $n, m \in \mathbb{N}$ such that $a^n \in P + (b)$ and $a^m \in P + (c)$, from which we obtain

$$a^{n+m} \in (P + (b)) \cdot (P + (c)) \subset P + (bc) = P,$$

in contradiction to $P \in M$. Hence P must be prime, which proves the lemma. □

Remark 2.22. Let Y be a subvariety of a variety X. Then the statement of Lemma 2.21 for the (already radical) ideal $I(Y)$ in the ring $A(X)$ corresponds to the geometrically obvious statement that the variety Y is the union of its irreducible subvarieties (see Remark 2.7).

Exercise 2.23 (Minimal primes)**.** Let I be an ideal in a ring R with $I \neq R$. We say that a prime ideal $P \trianglelefteq R$ is *minimal over* I if $I \subset P$ and there is no prime ideal Q with $I \subset Q \subsetneq P$.

(a) Prove that there is always a minimal prime ideal over I.

(b) Determine all minimal prime ideals over (x^2y, xy^2) in $\mathbb{R}[x, y]$.

If R is the coordinate ring of a variety and I the ideal of a subvariety, what is the geometric interpretation of a prime ideal minimal over I?

Exercise 2.24. Let P and Q be two distinct maximal ideals in a ring R, and let $n, m \in \mathbb{N}$. Show that P^n and Q^m are coprime.

Exercise 2.25. Show that for any ring R the following three statements are equivalent:

(a) R has exactly one prime ideal.

(b) Every element of R is either a unit or nilpotent.

(c) $\sqrt{(0)}$ is a maximal ideal.

Give an example for such a ring which is not a field.

Exercise 2.26. Let M be an infinite set.

(a) Use Zorn's Lemma to prove that M can be written as a disjoint union of sets that are all countably infinite.

(b) Show that for any non-empty and at most countable set A there is a bijection between M and $M \times A$.

Exercise 2.27.

(a) Show that every vector space has a basis (even if it is not finitely generated).

(b) Let $n, m \in \mathbb{N}_{>0}$ with $n \neq m$. Of course, you know that $\mathbb{R}^n$ and $\mathbb{R}^m$ are not isomorphic as $\mathbb{R}$-vector spaces. However, prove now that $\mathbb{R}^n$ and $\mathbb{R}^m$ are isomorphic as groups (with the standard addition).

(Hint: Consider $\mathbb{R}^n$ and $\mathbb{R}^m$ as $\mathbb{Q}$-vector spaces, and use Exercise 2.26 (b).)

04

3. Modules

In linear algebra, the most important structure is that of a vector space over a field. For commutative algebra it is therefore useful to consider the generalization of this concept to the case where the underlying space of scalars is a commutative ring R instead of a field. The resulting structure is called a *module*; we will introduce and study it in this chapter.

In fact, there is another more subtle reason why modules are very powerful: they unify many other structures that you already know. For example, when you first heard about quotient rings you were probably surprised that in order to obtain a quotient *ring* R/I one needs an *ideal* I of R, i. e. a structure somewhat different from that of a (sub-)ring. In contrast, we will see in Example 3.4 (a) that ideals as well as quotient rings of R are just special cases of modules over R, so that one can deal with both these structures in the same way. Even more unexpectedly, it turns out that modules over certain rings allow a special interpretation: modules over $\mathbb{Z}$ are nothing but Abelian groups, whereas a module over the polynomial ring $K[x]$ over a field K is exactly the same as a K-vector space V together with a linear map $\varphi : V \to V$ (see Examples 3.2 (d) and 3.8, respectively). Consequently, general results on modules will have numerous consequences in many different setups.

So let us now start with the definition of modules. In principle, their theory that we will then quickly discuss in this chapter is entirely analogous to that of vector spaces [G2, Chapters 13 to 18]. However, although many properties just carry over without change, others will turn out to be vastly different. Of course, proofs that are literally the same as for vector spaces will not be repeated here; instead we will just give references to the corresponding well-known linear algebra statements in these cases.

Definition 3.1 (Modules)**.** Let R be a ring. An **R-module** is a set M together with two operations

$$+ : M \times M \to M \qquad \text{and} \qquad \cdot : R \times M \to M$$

(an "addition" in M and a "scalar multiplication" with elements of R) such that for all $m, n \in M$ and $a, b \in R$ we have:

(a) $(M, +)$ is an Abelian group;

(b) $(a+b) \cdot m = a \cdot m + b \cdot m$ and $a \cdot (m+n) = a \cdot m + a \cdot n$;

(c) $(a \cdot b) \cdot m = a \cdot (b \cdot m)$;

(d) $1 \cdot m = m$.

We will also call M a *module over* R, or just a module if the base ring is clear from the context.

Example 3.2.

(a) For a field R, an R-module is by definition exactly the same as an R-vector space [G2, Definition 13.1].

(b) Of course, the zero set $\{0\}$ is a module, which we often simply write as 0.

(c) For $n \in \mathbb{N}_{>0}$ the set $R^n = \{(a_1, \ldots, a_n) : a_1, \ldots, a_n \in R\}$ is an R-module with componentwise addition and scalar multiplication. More generally, for two R-modules M and N the product $M \times N$ with componentwise addition and scalar multiplication is an R-module again.

(d) A $\mathbb{Z}$-module is just the same as an Abelian group. In fact, any $\mathbb{Z}$-module is an Abelian group by definition 3.1 (a), and in any Abelian group $(M, +)$ we can define a multiplication with integers in the usual way by $(-1) \cdot m := -m$ and $a \cdot m := m + \cdots + m$ (a times) for $a \in \mathbb{N}$ and $m \in M$.

(e) Any R-algebra M is also an R-module by Remark 1.24, if we just forget about the possibility to multiply two elements of M.

Definition 3.3 (Submodules, sums, and quotients)**.** Let M be an R-module.

(a) A **submodule** of M is a non-empty subset $N \subset M$ satisfying $m+n \in N$ and $am \in N$ for all $m,n \in N$ and $a \in R$. We write this as $N \leq M$. Of course, N is then an R-module itself, with the same addition and scalar multiplication as in M.

(b) For any subset $S \subset M$ the set

$$\langle S \rangle := \{a_1 m_1 + \cdots + a_n m_n : n \in \mathbb{N}, a_1, \ldots, a_n \in R, m_1, \ldots, m_n \in S\} \quad \subset M$$

of all R-linear combinations of elements of S is the smallest submodule of M that contains S. It is called the submodule **generated by** S. If $S = \{m_1, \ldots, m_n\}$ is finite, we write $\langle S \rangle = \langle \{m_1, \ldots, m_n\} \rangle$ also as $\langle m_1, \ldots, m_n \rangle$. The module M is called **finitely generated** if $M = \langle S \rangle$ for a finite set $S \subset M$.

(c) For submodules $N_1, \ldots, N_n \leq M$ their **sum**

$$N_1 + \cdots + N_n = \{m_1 + \cdots + m_n : m_i \in N_i \text{ for all } i = 1, \ldots, n\}$$

is obviously a submodule of M again. If moreover every element $m \in N_1 + \cdots + N_n$ has a *unique* representation as $m = m_1 + \cdots + m_n$ with $m_i \in N_i$ for all i, we call $N_1 + \cdots + N_n$ a **direct sum** and write it also as $N_1 \oplus \cdots \oplus N_n$.

(d) If $N \leq M$ is a submodule, the set

$$M/N := \{\bar{x} : x \in M\} \quad \text{with} \quad \bar{x} := x + N$$

of equivalence classes modulo N is again a module [G2, Proposition 15.15], the so-called **quotient module** of M modulo N.

Example 3.4.

(a) Let R be a ring. If we consider R itself as an R-module, a submodule of R is by definition the same as an ideal I of R. Moreover, the quotient ring R/I is then by Definition 3.3 (d) an R-module again.

Note that this is the first case where modules and vector spaces behave in a slightly different way: if K is a field then the K-vector space K has no non-trivial subspaces.

(b) The polynomial ring $K[x_1, \ldots, x_n]$ over a field K is finitely generated as a K-algebra (by $x_1, \ldots, x_n$), but not finitely generated as a K-module, i. e. as a K-vector space (the monomials $1, x_1, x_1^2, \ldots$ are linearly independent). So if we use the term "finitely generated" we always have to make sure to specify whether we mean "finitely generated as an algebra" or "finitely generated as a module", as these are two different concepts.

Exercise 3.5. Let N be a submodule of a module M over a ring R. Show:

(a) If N and M/N are finitely generated, then so is M.

(b) If M is finitely generated, then so is M/N.

(c) If M is finitely generated, N need not be finitely generated.

Definition 3.6 (Morphisms)**.** Let M and N be R-modules.

(a) A **morphism** of R-modules (or **R-module homomorphism**, or **R-linear map**) from M to N is a map $\varphi : M \to N$ such that

$$\varphi(m+n) = \varphi(m) + \varphi(n) \qquad \text{and} \qquad \varphi(am) = a\,\varphi(m)$$

for all $m,n \in M$ and $a \in R$. The set of all such morphisms from M to N will be denoted $\mathrm{Hom}_R(M,N)$ or just $\mathrm{Hom}(M,N)$; it is an R-module again with pointwise addition and scalar multiplication.

(b) A morphism $\varphi : M \to N$ of R-modules is called an **isomorphism** if it is bijective. In this case, the inverse map $\varphi^{-1} : N \to M$ is a morphism of R-modules again [G2, Lemma 13.25 (a)]. We call M and N **isomorphic** (written $M \cong N$) if there is an isomorphism between them.

Example 3.7.

(a) For any ideal I in a ring R, the quotient map $\varphi : R \to R/I,\ a \mapsto \overline{a}$ is a surjective R-module homomorphism.

(b) Let M and N be Abelian groups, considered as $\mathbb{Z}$-modules as in Example 3.2 (d). Then a $\mathbb{Z}$-module homomorphism $\varphi : M \to N$ is the same as a homomorphism of Abelian groups, since $\varphi(m+n) = \varphi(m) + \varphi(n)$ already implies $\varphi(am) = a\,\varphi(m)$ for all $a \in \mathbb{Z}$.

(c) For any R-module M we have $\mathrm{Hom}_R(R,M) \cong M$: the maps

$$M \to \mathrm{Hom}_R(R,M),\ m \mapsto (R \to M, a \mapsto am) \qquad \text{and} \qquad \mathrm{Hom}_R(R,M) \to M,\ \varphi \mapsto \varphi(1)$$

are obviously R-module homomorphisms and inverse to each other. On the other hand, the module $\mathrm{Hom}_R(M,R)$ is in general not isomorphic to M: for the $\mathbb{Z}$-module $\mathbb{Z}_2$ we have $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_2,\mathbb{Z}) = 0$ by (b), as there are no non-trivial group homomorphisms from $\mathbb{Z}_2$ to $\mathbb{Z}$.

(d) If $N_1,\dots,N_n$ are submodules of an R-module M such that their sum $N_1 \oplus \cdots \oplus N_n$ is direct, the morphism

$$N_1 \times \cdots \times N_n \to N_1 \oplus \cdots \oplus N_n,\ (m_1,\dots,m_n) \mapsto m_1 + \cdots + m_n$$

is bijective, and hence an isomorphism. One therefore often uses the notation $N_1 \oplus \cdots \oplus N_n$ for $N_1 \times \cdots \times N_n$ also in the cases where $N_1,\dots,N_n$ are R-modules that are not necessarily submodules of a given ambient module M.

Example 3.8 (Modules over polynomial rings)**.** Let R be a ring. Then an $R[x]$-module M is the same as an R-module M together with an R-module homomorphism $\varphi : M \to M$:

"$\Rightarrow$" Let M be an $R[x]$-module. Of course, M is then also an R-module. Moreover, multiplication with x has to be R-linear, so $\varphi : M \to M, m \mapsto x \cdot m$ is an R-module homomorphism.

"$\Leftarrow$" If M is an R-module and $\varphi : M \to M$ an R-module homomorphism we can give M the structure of an $R[x]$-module by setting $x \cdot m := \varphi(m)$, or more precisely by defining scalar multiplication by

$$\Big(\sum_{i=0}^{n} a_i x^i\Big) \cdot m := \sum_{i=0}^{n} a_i \varphi^i(m),$$

where φ^i denotes the i-fold composition of φ with itself, and $\varphi^0 := \mathrm{id}_M$.

Remark 3.9 (Images and kernels of morphisms)**.** Let $\varphi : M \to N$ be a homomorphism of R-modules.

(a) For any submodule $M' \leq M$ the image $\varphi(M')$ is a submodule of N [G2, Lemma 13.21 (a)]. In particular, $\varphi(M)$ is a submodule of N, called the **image** of φ.

(b) For any submodule $N' \leq N$ the inverse image $\varphi^{-1}(N')$ is a submodule of M [G2, Lemma 13.21 (b)]. In particular, $\varphi^{-1}(0)$ is a submodule of M, called the **kernel** of φ.

Proposition 3.10 (Isomorphism theorems)**.**

(a) *For any morphism $\varphi : M \to N$ of R-modules there is an isomorphism*

$$M/\ker\varphi \to \mathrm{im}\,\varphi,\ \ \overline{m} \mapsto \varphi(m).$$

(b) *For R-modules $N' \leq N \leq M$ we have*

$$(M/N')\,/\,(N/N') \ \cong\ M/N.$$

(c) *For two submodules N, N' of an R-module M we have*

$$(N+N')/N' \ \cong\ N/(N \cap N').$$

Proof. The proofs of (a) and (b) are the same as in [G2, Proposition 15.22] and Exercise 1.22, respectively. For (c) note that $N \to (N+N')/N',\ m \mapsto \overline{m}$ is a surjective R-module homomorphism with kernel $N \cap N'$, so the statement follows from (a). □

Exercise 3.11. Let N be a proper submodule of an R-module M. Show that the following statements are equivalent:

(a) There is no submodule P of M with $N \subsetneq P \subsetneq M$.

(b) The module M/N has only the trivial submodules 0 and M/N.

(c) $M/N \cong R/I$ for a maximal ideal $I \trianglelefteq R$.

The concepts so far were all entirely analogous to the case of vector spaces. There are a few constructions however that are only useful for modules due to the existence of non-trivial ideals in the base ring. Let us introduce them now.

Definition 3.12 (IM, module quotients, annihilators)**.** Let M be an R-module.

(a) For an ideal $I \trianglelefteq R$ we set

$$\begin{aligned} IM &:= \langle \{am : a \in I, m \in M\} \rangle \\ &= \{a_1m_1 + \cdots + a_nm_n : n \in \mathbb{N}, a_1, \ldots, a_n \in I, m_1, \ldots, m_n \in M\}. \end{aligned}$$

Note that IM is a submodule of M, and M/IM is an R/I-module in the obvious way.

(b) For two submodules $N, N' \leq M$ the **module quotient** (not to be confused with the quotient modules of Definition 3.3 (d)) is defined to be

$$N' : N := \{a \in R : aN \subset N'\} \quad \trianglelefteq R.$$

In particular, for $N' = 0$ we obtain the so-called **annihilator**

$$\operatorname{ann} N := \operatorname{ann}_R N := \{a \in R : aN = 0\} \quad \trianglelefteq R$$

of N. The same definition can also be applied to a single element $m \in M$ instead of a submodule N: we then obtain the ideals

$$N' : m := \{a \in R : am \in N'\} \quad \text{and} \quad \operatorname{ann} m := \{a \in R : am = 0\}$$

of R.

Example 3.13.

(a) If M, N, and N' are submodules of the R-module R, i. e. ideals of R by Example 3.4 (a), the product IM and quotient $N' : N$ of Definition 3.12 are exactly the product and quotient of ideals as in Construction 1.1.

(b) If I is an ideal of a ring R then $\operatorname{ann}_R(R/I) = I$.

Let us recall again the linear algebra of vector spaces over a field K. At the point where we are now, i. e. after having studied subspaces and morphisms in general, one usually restricts to finitely generated vector spaces and shows that every such vector space V has a finite basis. This makes V isomorphic to K^n with $n = \dim_K V \in \mathbb{N}$ [G2, Proposition 14.22]. In other words, we can describe vectors by their coordinates with respect to some basis, and linear maps by matrices — which are then easy to deal with.

For a finitely generated module M over a ring R this strategy unfortunately breaks down. Ultimately, the reason for this is that the lack of a division in R means that a linear relation among generators of M cannot necessarily be used to express one of them in terms of the others (so that it can be dropped from the set of generators). For example, the elements $m = 2$ and $n = 3$ in the $\mathbb{Z}$-module $\mathbb{Z}$ satisfy the linear relation $3m - 2n = 0$, but neither is m an integer multiple of n, nor vice versa. So although $\mathbb{Z} = \langle m, n \rangle$ and these two generators are linearly dependent over $\mathbb{Z}$, neither m nor n alone generates $\mathbb{Z}$.

The consequence of this is that a finitely generated module M need not have a linearly independent set of generators. But this means that M is in general not isomorphic to R^n for some $n \in \mathbb{N}$, and thus there is no obvious well-defined notion of dimension. It is in fact easy to find examples for this: $\mathbb{Z}_2$ as a $\mathbb{Z}$-module is certainly not isomorphic to $\mathbb{Z}^n$ for some n.

So essentially we have two choices if we want to continue to carry over our linear algebra results on finitely generated vector spaces to finitely generated modules:

- restrict to R-modules that are of the form R^n for some $n \in \mathbb{N}$; or
- go on with general finitely generated modules, taking care of the fact that generating systems cannot be chosen to be independent, and thus that the coordinates with respect to such systems are no longer unique.

In the rest of this chapter, we will follow both strategies to some extent, and see what they lead to. Let us start by considering finitely generated modules that do admit a basis.

Definition 3.14 (Bases and free modules)**.** Let M be a finitely generated R-module.

(a) We say that a family $(m_1,\dots,m_n)$ of elements of M is a **basis** of M if the R-module homomorphism
$$R^n \to M,\ \ (a_1,\dots,a_n) \mapsto a_1m_1 + \dots + a_nm_n$$
is an isomorphism.

(b) If M has a basis, i. e. if it is isomorphic to R^n for some n, it is called a **free** R-module.

Example 3.15. If I is a non-trivial ideal in a ring R then R/I is never a free R-module: there can be no isomorphism
$$\varphi : R^n \to R/I,\ \ (a_1,\dots,a_n) \mapsto a_1m_1 + \dots + a_nm_n$$
since in any case $\varphi(0,\dots,0) = \varphi(a,0,\dots,0)$ for every $a \in I$.

Exercise 3.16. Let R be an integral domain. Prove that a non-zero ideal $I \trianglelefteq R$ is a principal ideal if and only if it is a free R-module.

Remark 3.17 (Linear algebra for free modules)**.** Let M and N be finitely generated, free R modules.

(a) Any two bases of M have the same number of elements: assume that we have a basis with n elements, so that $M \cong R^n$ as R-modules. Choose a maximal ideal I of R by Corollary 2.17. Then R/I is a field by Lemma 2.3 (b), and M/IM is an R/I-vector space by Definition 3.12 (a). Its dimension is
$$\dim_{R/I} M/IM = \dim_{R/I} R^n/IR^n = \dim_{R/I}(R/I)^n = n,$$
and so n is uniquely determined by M. We call n the **rank** $\operatorname{rk} M$ of M.

(b) In the same way as for vector spaces, we see that $\operatorname{Hom}_R(R^m,R^n)$ is isomorphic to the R-module $\operatorname{Mat}(n\times m,R)$ of $n\times m$-matrices over R [G2, Proposition 16.11]. Hence, after choosing bases for M and N we also have $\operatorname{Hom}_R(M,N) \cong \operatorname{Mat}(n\times m,R)$ with $m = \operatorname{rk} M$ and $n = \operatorname{rk} N$ [G2, Proposition 16.23].

(c) An R-module homomorphism $\varphi : M \to M$ is an isomorphism if and only if its matrix $A \in \operatorname{Mat}(m\times m,R)$ as in (b) is invertible, i. e. if and only if there is a matrix $A^{-1} \in \operatorname{Mat}(m\times m,R)$ such that $A^{-1}A = AA^{-1} = E$ is the unit matrix. As expected, whether this is the case can be checked with determinants as follows.

(d) For a square matrix $A \in \operatorname{Mat}(m\times m,R)$ the *determinant* $\det A$ is defined in the usual way [G2, Proposition 18.12]. It has all the expected properties; in particular there is an *adjoint matrix* $A^\# \in \operatorname{Mat}(m\times m,R)$ such that $A^\#A = AA^\# = \det A\cdot E$ (namely the matrix with (i,j)-entry $(-1)^{i+j}\det A'_{j,i}$, where $A'_{j,i}$ is obtained from A by deleting row j and column i) [G2, Proposition 18.20 (a)]. With this we can see that A is invertible if and only if $\det A$ is a unit in R:

"$\Rightarrow$" If there is an inverse matrix A^{-1} then $1 = \det E = \det(A^{-1}A) = \det A^{-1}\cdot\det A$, so $\det A$ is a unit in R.

"$\Leftarrow$" If $\det A$ is a unit, we see from the equation $A^\#A = AA^\# = \det A\cdot E$ that $(\det A)^{-1}\cdot A^\#$ is an inverse of A.

05

So all in all finitely generated, free modules behave very much in the same way as vector spaces. However, most modules occurring in practice will not be free — in fact, submodules and quotient modules of free modules, as well as images and kernels of homomorphisms of free modules, will in

general not be free again. So let us now also find out what we can say about more general finitely generated modules.

First of all, the notion of dimension of a vector space, or rank of a free module as in Remark 3.17 (a), is then no longer defined. The following notion of the *length* of a module can often be used to substitute this.

Definition 3.18 (Length of modules)**.** Let M be an R-module.

(a) A **composition series** for M is a finite chain

$$0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots \subsetneq M_n = M$$

of submodules of M that cannot be refined, i. e. such that there is no submodule N of M with $M_{i-1} \subsetneq N \subsetneq M_i$ for any $i = 1, \ldots, n$. (By Exercise 3.11, this is equivalent to M_i/M_{i-1} having no non-trivial submodules for all i, and to M_i/M_{i-1} being isomorphic to R modulo some maximal ideal for all i).

The number n above will be called the length of the series.

(b) If there is a composition series for M, the shortest length of such a series is called the **length** of M and denoted $l_R(M)$ (in fact, we will see in Exercise 3.19 (b) that then all composition series have this length). Otherwise, we set formally $l_R(M) = \infty$.

If there is no risk of confusion about the base ring, we write $l_R(M)$ also as $l(M)$.

Exercise 3.19. Let M be an R-module of finite length, i. e. an R-module that admits a composition series. Show that:

(a) If $N < M$ is a proper submodule of M then $l(N) < l(M)$.

(b) Every composition series for M has length $l(M)$.

(c) Every chain $0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots \subsetneq M_n = M$ of submodules of M can be refined to a composition series for M.

Example 3.20.

(a) Let V be a vector space over a field K. If V has finite dimension n, there is a chain

$$0 = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_n = V$$

of subspaces of V with $\dim_K V_i = i$ for all i. Obviously, this chain cannot be refined. Hence it is a composition series for V, and we conclude by Exercise 3.19 (b) that $l(V) = n = \dim_K V$.

On the other hand, if V has infinite dimension, there are chains of subspaces of V of any length. By Exercise 3.19 this is only possible if $l(V) = \infty$.

So for vector spaces over a field, the length is just the same as the dimension.

(b) There is no statement analogous to (a) for free modules over a ring: already $\mathbb{Z}$ has infinite length over $\mathbb{Z}$, since there are chains

$$0 \subsetneq (2^n) \subsetneq (2^{n-1}) \subsetneq \cdots \subsetneq (2) \subsetneq \mathbb{Z}$$

of ideals in $\mathbb{Z}$ of any length.

(c) Certainly, a module M of finite length must be finitely generated: otherwise there would be infinitely many elements $(m_i)_{i \in \mathbb{N}}$ of M such that all submodules $M_i = \langle m_1, \ldots, m_i \rangle$ are distinct. But then $0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$ is an infinite chain of submodules, which by Exercise 3.19 is impossible for modules of finite length.

On the other hand, a finitely generated module need not have finite length, as we have seen in (b). In fact, we will study the relation between the conditions of finite generation and finite length in more detail in Chapter 7.

Exercise 3.21. What are the lengths of $\mathbb{Z}_8$ and $\mathbb{Z}_{12}$ as $\mathbb{Z}$-modules? Can you generalize this statement to compute the length of any quotient ring R/I as an R-module, where I is an ideal in a principal ideal domain R?

Let us now show that the length of modules satisfies the same relations as the dimension of vector spaces when taking sums, intersections, quotients, or images and kernels [G2, Corollary 15.30, Proposition 15.16, and Corollary 15.25].

Proposition 3.22 (Additivity of the length of modules)**.** *For any submodule N of an R-submodule M we have*

$$l(N)+l(M/N)=l(M).$$

Proof. Let us assume first that $l(M)<\infty$. By Exercise 3.19 (c), the chain $0\le N\le M$ can be refined to a composition series

$$0=N_0\subsetneq\cdots\subsetneq N_n=N=M_0\subsetneq\cdots\subsetneq M_m=M \qquad (*)$$

for M, where $l(M)=n+m$ by Exercise 3.19 (b). Of course, the first part of this chain is then a composition series for N, and so $l(N)=n$. Moreover, setting $P_i:=M_i/N$ for $i=1,\dots,m$ we obtain a chain

$$0=P_0\subsetneq\cdots\subsetneq P_m=M/N$$

in which $P_i/P_{i-1}\cong M_i/M_{i-1}$ by Proposition 3.10 (b). As these modules have no non-trivial submodules, we see that the above chain of length m is a composition series for M/N, so that we get the desired result $l(N)+l(M/N)=n+m=l(M)$.

Conversely, if $l(N)$ and $l(M/N)$ are finite, there are composition series

$$0=N_0\subsetneq\cdots\subsetneq N_n=N \quad\text{and}\quad 0=P_0\subsetneq\cdots\subsetneq P_m=M/N$$

for N and M/N, respectively. Setting $M_i:=q^{-1}(P_i)$ with the quotient map $q:M\to M/N$ for all $i=1,\dots,m$, we have $M_i/N=P_i$. So as above, $M_i/M_{i-1}\cong P_i/P_{i-1}$ has no non-trivial submodules, and we obtain a composition series $(*)$ for M. This means that M has finite length as well, and the above argument can be applied to prove the equation $l(N)+l(M/N)=l(M)$ again.

The only remaining case is that both sides of the equation of the proposition are infinite — but then of course the statement is true as well. □

Corollary 3.23.

(a) *For any two submodules M_1,M_2 of an R-module M we have*

$$l(M_1+M_2)+l(M_1\cap M_2)=l(M_1)+l(M_2).$$

(b) *For any R-module homomorphism $\varphi:M\to N$ we have*

$$l(\ker\varphi)+l(\operatorname{im}\varphi)=l(M).$$

Proof.

(a) By Proposition 3.10 (c) we have $(M_1+M_2)/M_2\cong M_1/(M_1\cap M_2)$. Calling this module Q, we obtain by Proposition 3.22

$$l(M_1+M_2)=l(M_2)+l(Q) \qquad\text{and}\qquad l(M_1)=l(M_1\cap M_2)+l(Q).$$

So if $l(M_2)=\infty$ then $l(M_1+M_2)=\infty$, and the statement of the corollary holds. The same is true if $l(M_1\cap M_2)=\infty$ and thus $l(M_1)=\infty$. Otherwise, we obtain

$$l(Q)=l(M_1+M_2)-l(M_2)=l(M_1)-l(M_1\cap M_2),$$

and hence the corollary holds in this case as well.

(b) This is just Proposition 3.22 applied to the homomorphism theorem $M/\ker\varphi\cong\operatorname{im}\varphi$ of Proposition 3.10 (a). □

Remark 3.24. An easy consequence of Corollary 3.23 (b) is that for a homomorphism $\varphi:M\to M$ from a module of finite length to itself we have

$$\varphi\text{ injective}\quad\Leftrightarrow\quad\varphi\text{ surjective}\quad\Leftrightarrow\quad\varphi\text{ bijective}$$

as in [G2, Corollary 15.26], since φ is injective if and only if $l(\ker\varphi)=0$, and surjective if and only if $l(\operatorname{im}\varphi)=l(M)$ (see Exercise 3.19 (a)).

What happens in this statement if we consider a module M that is only finitely generated, but not necessarily of finite length (see Example 3.20 (c))? It is actually easy to see that in this case an injective morphism $\varphi : M \to M$ need not be bijective: the map $\varphi : \mathbb{Z} \to \mathbb{Z}$, $m \mapsto 2m$ is a simple counterexample. In view of this example it is probably surprising that the statement that a surjective map is also bijective still holds — this is what we want to show in Corollary 3.28 below. The main ingredient in the proof is the following generalization of the Cayley-Hamilton theorem from linear algebra.

Proposition 3.25 (Cayley-Hamilton). *Let M be a finitely generated R-module, I an ideal of R, and $\varphi : M \to M$ an R-module homomorphism with $\varphi(M) \subset IM$. Then there is a monic polynomial (i. e. its leading coefficient is 1)*

$$\chi = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \quad \in R[x]$$

with $a_0, \ldots, a_{n-1} \in I$ and

$$\chi(\varphi) := \varphi^n + a_{n-1}\varphi^{n-1} + \cdots + a_0 \,\mathrm{id} = 0 \quad \in \mathrm{Hom}_R(M,M),$$

where φ^i denotes the i-fold composition of φ with itself.

Proof. Let $m_1, \ldots, m_n$ be generators of M. By assumption we have $\varphi(m_i) \in IM$ for all i, and thus there are $a_{i,j} \in I$ with

$$\varphi(m_i) = \sum_{j=1}^{n} a_{i,j} m_j \quad \text{for all } i = 1, \ldots, n.$$

Considering M as an $R[x]$-module by setting $x \cdot m := \varphi(m)$ for all $m \in M$ as in Example 3.8, we can rewrite this as

$$\sum_{j=1}^{n} (x\delta_{i,j} - a_{i,j}) m_j = 0 \quad \text{with} \quad \delta_{i,j} := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

for all $i = 1, \ldots, n$. Note that the left hand side of this equation, taken for all i, gives us an element of M^n. If we multiply this from the left with the adjoint matrix of $(x\delta_{i,j} - a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, R[x])$ as in Remark 3.17 (d), we get

$$\det((x\delta_{i,k} - a_{i,k})_{i,k}) \cdot m_j = 0$$

for all j. So $\chi := \det((x\delta_{i,k} - a_{i,k})_{i,k})$ acts as the zero homomorphism on M, and expanding the determinant shows that the non-leading coefficients of this polynomial lie in fact in I. □

Remark 3.26. If R is a field and thus M a finitely generated vector space, we can only take $I = R$ in Proposition 3.25. In the proof, we can then choose $m_1, \ldots, m_n$ to be a basis of M, so that $(a_{i,j})_{i,j}$ is the matrix of φ with respect to this basis, and χ is the characteristic polynomial of φ [G2, Definitions 19.16 and Remark 19.21]. So in this case the statement of Proposition 3.25 is just the ordinary Cayley-Hamilton theorem for endomorphisms of finite-dimensional vector spaces [G2, Exercise 20.24]. For general rings however, the generators $m_1, \ldots, m_n$ are no longer independent, and so there are several choices for the matrix $(a_{i,j})_{i,j}$.

The following easy consequence of Proposition 3.25 is usually attributed to Nakayama. In fact, there are many versions of Nakayama's lemma in the literature (we will also meet some other closely related statements in Exercise 6.16), but this one is probably one of the strongest. It concerns the construction IM of Definition 3.12 (a) for an ideal I in a ring R and an R-module M. More precisely, let us assume that $M \neq 0$ and $IM = M$. Of course, if R is a field this is only possible if $I = R$, i. e. if $1 \in I$. If R is a general ring however, it does not necessarily follow that $1 \in I$ — but Nakayama's Lemma states that in the case of a finitely generated module M there is at least an element $a \in I$ that acts as the identity on M, i. e. such that $am = m$ for all $m \in M$.

Corollary 3.27 (Nakayama's Lemma). *Let M be a finitely generated R-module, and I an ideal of R with $IM = M$. Then there is an element $a \in I$ with $am = m$ for all $m \in M$.*

Proof. As $M = IM$ we can apply Proposition 3.25 to $\varphi = \mathrm{id}$ and our given ideal I to obtain $a_0, \dots, a_{n-1} \in I$ such that

$$\mathrm{id}^n + a_{n-1}\,\mathrm{id}^{n-1} + \cdots + a_0\,\mathrm{id} = (1 + a_{n-1} + \cdots + a_0)\,\mathrm{id} = 0 \quad \in \mathrm{Hom}_R(M,M).$$

Setting $a := -a_{n-1} - \cdots - a_0 \in I$, this just means that $(1-a)m = 0$, i. e. $am = m$ for all $m \in M$. □

Corollary 3.28. *If M is a finitely generated R-module, any surjective homomorphism $\varphi : M \to M$ is an isomorphism.*

Proof. As in Example 3.8, consider M as an $R[x]$-module by setting $x \cdot m := \varphi(m)$ for all $m \in M$. Then $x \cdot M = \varphi(M) = M$ since φ is surjective, so we can apply Corollary 3.27 with $I = (x)$ to obtain a polynomial $f \in (x)$, i. e. a polynomial $f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x$ without constant coefficient, such that

$$f \cdot m = a_n \varphi^n(m) + a_{n-1}\varphi^{n-1}(m) + \cdots + a_1 \varphi(m) = m \quad \text{for all } m \in M.$$

But this means that $\varphi(m) = 0$ implies $m = 0$, and so φ is injective. □

Exercise 3.29. For a prime number $p \in \mathbb{N}$ consider the subring $R = \{\frac{a}{b} : a, b \in \mathbb{Z}, p \nmid b\}$ of $\mathbb{Q}$, and let $M = \mathbb{Q}$ as an R-module.

(a) Show that R has exactly one maximal ideal I. Which one?

(In fact, this will be obvious once we have studied localizations — see Example 6.6 and Corollary 6.10.)

(b) For the ideal of (a), prove that $IM = M$, but there is no $a \in I$ with $am = m$ for all $m \in M$.

(c) Find a "small" set of generators for M as an R-module. Can you find a finite one? A minimal one?

4. Exact Sequences

In the last chapter we have studied many structures related to modules, such as submodules, quotient modules, and module homomorphisms together with their images and kernels. We now want to introduce a very useful piece of notation that can be used to deal with all these concepts in a unified way: the so-called *exact sequences*. In many cases they provide an easy and almost "graphical" way to express and prove statements that are in principle elementary, but tedious to write down without this language due to the sheer number of spaces, morphisms, or relations involved.

Definition 4.1 (Exact sequences)**.** Let R be a ring, and let $n \in \mathbb{N}_{\geq 3}$. A sequence

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} \quad \dots \quad \xrightarrow{\varphi_{n-1}} M_n$$

of R-modules $M_1, \dots, M_n$ and R-module homomorphisms $\varphi_i : M_i \to M_{i+1}$ for $i = 1, \dots, n-1$ is called **exact at** position $i \in \{2, \dots, n-1\}$ if $\operatorname{im} \varphi_{i-1} = \ker \varphi_i$. It is called **exact** if it is exact at every position $i \in \{2, \dots, n-1\}$.

In such sequences, we will often not label an arrow with its morphism if it is clear from the context what the morphism is. In particular, this is the case if the source or target of the map is the zero module, so that the map is necessarily the zero morphism.

Example 4.2 (Exact sequences with few modules)**.**

(a) A sequence $0 \longrightarrow M \xrightarrow{\varphi} N$ is exact if and only if $\ker \varphi = 0$, i. e. if and only if φ is injective.

A sequence $M \xrightarrow{\varphi} N \longrightarrow 0$ is exact if and only if $\operatorname{im} \varphi = N$, i. e. if and only if φ is surjective.

(b) The sequence $0 \longrightarrow M \longrightarrow 0$ is exact if and only if $M = 0$.

By (a), the sequence $0 \longrightarrow M \xrightarrow{\varphi} N \longrightarrow 0$ is exact if and only if φ is injective and surjective, i. e. if and only if φ is an isomorphism.

Example 4.3 (Short exact sequences)**.** By Example 4.2, the first interesting case of an exact sequence occurs if it has at least three non-zero terms. Therefore, an exact sequence of the form

$$0 \longrightarrow M_1 \xrightarrow{\varphi} M_2 \xrightarrow{\psi} M_3 \longrightarrow 0 \qquad (*)$$

is called a **short exact sequence**. There are two main sources for such short exact sequences:

(a) For any R-module homomorphism $\psi : M \to N$ the sequence

$$0 \longrightarrow \ker \psi \longrightarrow M \xrightarrow{\psi} \operatorname{im} \psi \longrightarrow 0$$

is exact, where the first map is the inclusion of $\ker \psi$ in M.

(b) For any submodule N of an R-module M the sequence

$$0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$$

is exact, where the first map is again the inclusion, and the second the quotient map.

In fact, up to isomorphisms every short exact sequence is of these forms: in a short exact sequence as in $(*)$ the second map ψ is surjective, and thus $\operatorname{im} \psi = M_3$. Moreover, $\ker \psi = \operatorname{im} \varphi = \varphi(M_1) \cong M_1$, where the last isomorphism follows from the injectivity of φ. So the given sequence is of the type as in (a). If we set $N = \ker \psi \leq M_2$ and use the homomorphism theorem $\operatorname{im} \psi \cong M_2/\ker \psi = M_2/N$ of Proposition 3.10 (a), we can also regard it as a sequence as in (b).

A nice feature of exact sequences is that there are simple rules to create new sequences from old ones. The simplest way to do this is to split and glue exact sequences as follows.

Lemma 4.4 (Splitting and gluing exact sequences)**.**

(a) *(Splitting) If*

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} M_3 \xrightarrow{\varphi_3} M_4 \tag{A}$$

is an exact sequence of R-modules, the two sequences

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} N \longrightarrow 0 \qquad \text{and} \qquad 0 \longrightarrow N \longrightarrow M_3 \xrightarrow{\varphi_3} M_4 \tag{B}$$

are also exact, where $N = \operatorname{im}\varphi_2 = \ker\varphi_3$*, and the middle map in the second sequence is the inclusion of* $\ker\varphi_3$ *in* M_3*.*

(b) *(Gluing) Conversely, if*

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} N \longrightarrow 0 \qquad \text{and} \qquad 0 \longrightarrow N \longrightarrow M_3 \xrightarrow{\varphi_3} M_4 \tag{B}$$

are two exact sequences, with $N \leq M_3$ *a submodule and the middle map in the second sequence the inclusion, the sequence*

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} M_3 \xrightarrow{\varphi_3} M_4 \tag{A}$$

is also exact.

Proof.

(a) The first sequence of (B) is exact at M_2 since $\operatorname{im}\varphi_1 = \ker\varphi_2$, and exact at N as $N = \operatorname{im}\varphi_2$. The second sequence of (B) is exact N as the middle map is an inclusion, and exact at M_3 since $N = \ker\varphi_3$.

(b) The sequence of (A) is exact at M_2 since $\operatorname{im}\varphi_1 = \ker\varphi_2$. Moreover, in (B) exactness of the first sequence at N and of the second sequence at M_3 means that $\operatorname{im}\varphi_2 = N = \ker\varphi_3$, which implies that (A) is exact at M_3. □

Remark 4.5 (Splitting an exact sequence into short ones)**.** With Lemma 4.4 (a) every exact sequence

$$0 \longrightarrow M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} \cdots \xrightarrow{\varphi_{n-1}} M_n \longrightarrow 0 \tag{$*$}$$

(for simplicity with zero modules at the end) can be split up into short exact sequences

$$0 \longrightarrow \ker\varphi_i \longrightarrow M_i \xrightarrow{\varphi_i} \operatorname{im}\varphi_i \longrightarrow 0$$

for $i = 2, \ldots, n-1$ as in Example 4.3 (a), where $M_1 = \ker\varphi_2$ and $M_n = \operatorname{im}\varphi_{n-1}$. Conversely, such short exact sequences with $M_1 = \ker\varphi_2$, $M_n = \operatorname{im}\varphi_{n-1}$, and $\operatorname{im}\varphi_{i-1} = \ker\varphi_i$ for $i = 3, \ldots, n-1$ can be glued back together by Lemma 4.4 (b) to the long exact sequence $(*)$.

Example 4.6 (Exact sequence of a homomorphism)**.** Let $\varphi : M \to N$ be a homomorphism of R-modules. By Example 4.3 (a) and (b) there are then short exact sequences

$$0 \longrightarrow \ker\varphi \longrightarrow M \xrightarrow{\varphi} \operatorname{im}\varphi \longrightarrow 0 \qquad \text{and} \qquad 0 \longrightarrow \operatorname{im}\varphi \longrightarrow N \longrightarrow N/\operatorname{im}\varphi \longrightarrow 0,$$

and hence we get a glued exact sequence

$$0 \longrightarrow \ker\varphi \longrightarrow M \xrightarrow{\varphi} N \longrightarrow N/\operatorname{im}\varphi \longrightarrow 0$$

by Lemma 4.4 (b). So any homomorphism can be completed both to the left and to the right to an exact sequence with zero modules at the ends. Of course, the exactness of this sequence could also be checked directly, without using Lemma 4.4 (b).

06

There is another much more subtle way to construct new exact sequences from old ones. This time, instead of gluing two sequences such that the ending module of the first sequence is the starting module of the second, we consider two short exact sequences such that one of them can be mapped to the other by a sequence of homomorphisms.

Lemma 4.7 (Snake Lemma). *Let*

$$\begin{array}{ccccccc} & & M & \xrightarrow{\varphi} & N & \xrightarrow{\psi} & P & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \longrightarrow & M' & \xrightarrow{\varphi'} & N' & \xrightarrow{\psi'} & P' & & \end{array}$$

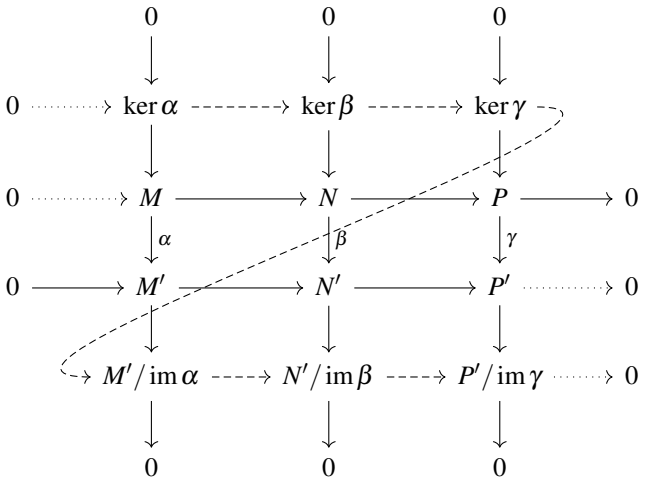
be a commutative diagram of R-modules with exact rows. Then there is a long exact sequence

$$\ker\alpha \longrightarrow \ker\beta \longrightarrow \ker\gamma \longrightarrow M'/\operatorname{im}\alpha \longrightarrow N'/\operatorname{im}\beta \longrightarrow P'/\operatorname{im}\gamma. \qquad (*)$$

Moreover:

- *if φ is injective, then so is the first map in the sequence $(*)$;*
- *if ψ' is surjective, then so is the last map in the sequence $(*)$.*

Remark 4.8. There is a nice graphical way to express the assertion of Lemma 4.7 as follows. First of all, if we complete the vertical homomorphisms α, β, and γ to exact sequences as in Example 4.6 we obtain the solid arrows in the following diagram, in which all columns are exact.



The statement of the lemma is now the existence of an exact sequence indicated by the dashed arrows above. In particular, the probably unexpected homomorphism from $\ker\gamma$ to $M'/\operatorname{im}\alpha$ is the reason for the name "Snake Lemma". The dotted arrows represent the additional statements of Lemma 4.7: one can read the diagram with or without both dotted arrows on the left, and with or without both dotted arrows on the right.

Proof of Lemma 4.7. In the following proof, we will denote elements of M, N, P, M', N', P' by the corresponding lower case letters.

First of all let us construct the homomorphisms in the sequence

$$\ker\alpha \xrightarrow{\tilde\varphi} \ker\beta \xrightarrow{\tilde\psi} \ker\gamma \xrightarrow{\delta} M'/\operatorname{im}\alpha \xrightarrow{\tilde\varphi'} N'/\operatorname{im}\beta \xrightarrow{\tilde\psi'} P'/\operatorname{im}\gamma. \qquad (*)$$

(a) *The first two maps.* Let $\tilde\varphi : \ker\alpha \to \ker\beta$, $m \mapsto \varphi(m)$ be the restriction of φ to $\ker\alpha$. Note that its image lies indeed in $\ker\beta$: for $m \in \ker\alpha$ we have $\beta(\varphi(m)) = \varphi'(\alpha(m)) = 0$ since the given diagram is commutative. In the same way, let $\tilde\psi$ be the restriction of ψ to $\ker\beta$.

(b) *The last two maps.* We have $\varphi'(\operatorname{im}\alpha) \subset \operatorname{im}\beta$, since $\varphi'(\alpha(m)) = \beta(\varphi(m))$. Hence we get a well-defined map $\tilde\varphi' : M'/\operatorname{im}\alpha \to N'/\operatorname{im}\beta$, $\overline{m'} \mapsto \overline{\varphi'(m')}$. Similarly, set $\tilde\psi'(\overline{n'}) := \overline{\psi'(n')}$.

(c) *The middle map δ.* The existence of this map from a submodule of P to a quotient module of M' — it is usually called the *connecting homomorphism* — is the central part of this lemma.

Schematically, its idea is shown by the dashed arrow in the picture below: take an inverse image under ψ, then the image under β, and then again an inverse image under φ'.

$$\begin{array}{ccccccccc} & & M & \xrightarrow{\varphi} & N & \xrightarrow{\psi} & P & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \longrightarrow & M' & \xrightarrow[\varphi']{} & N' & \xrightarrow[\psi']{} & P' & & \end{array}$$

More precisely, let $p \in \ker\gamma \leq P$. As ψ is surjective there is an element $n \in N$ with $\psi(n) = p$. Set $n' = \beta(n)$. Then $\psi'(n') = \psi'(\beta(n)) = \gamma(\psi(n)) = \gamma(p) = 0$, so that $n' \in \ker\psi' = \operatorname{im}\varphi'$. Hence there is an element $m' \in M'$ with $\varphi'(m') = n'$. We want to set $\delta(p) := \overline{m'} \in M'/\operatorname{im}\alpha$.

We have to check that this is well-defined, i. e. does not depend on the two choices of inverse images. Note that the choice of m' was clearly unique since φ' is injective. Now let us suppose that we pick another inverse image $\tilde{n} \in N$ of p in the first step, with image $\tilde{n}' = \beta(\tilde{n})$ under β and inverse image $\tilde{m}'$ under φ'. Then $\psi(\tilde{n} - n) = 0$, so $\tilde{n} - n \in \ker\psi = \operatorname{im}\varphi$, i. e. $\tilde{n} - n = \varphi(m)$ for some $m \in M$. It follows that

$$\varphi'(\tilde{m}' - m') = \tilde{n}' - n' = \beta(\tilde{n} - n) = \beta(\varphi(m)) = \varphi'(\alpha(m)),$$

and hence $\tilde{m}' - m' = \alpha(m)$ since φ' is injective. Consequently, $\tilde{m}'$ and m' define the same class in $M'/\operatorname{im}\alpha$, and thus δ is in fact well-defined.

As all these maps are clearly homomorphisms, it only remains to prove that the sequence $(*)$ is exact. The technique to do this is the same as in the construction of the connecting homomorphism δ above: just following elements through the given commutative diagram, a so-called *diagram chase*. This is purely automatic and does not require any special ideas. The proof is therefore probably not very interesting, but we will give it here nevertheless for the sake of completeness.

- *Exactness at* $\ker\beta$. Clearly, $\tilde{\psi}(\tilde{\varphi}(m)) = \psi(\varphi(m)) = 0$ for $m \in \ker\alpha$, and so $\operatorname{im}\tilde{\varphi} \subset \ker\tilde{\psi}$. Conversely, if $n \in \ker\tilde{\psi} \leq N$ then $n \in \ker\psi = \operatorname{im}\varphi$, so $n = \varphi(m)$ for some $m \in M$. This element satisfies $\varphi'(\alpha(m)) = \beta(\varphi(m)) = \beta(n) = 0$ since $n \in \ker\beta$. By the injectivity of φ' this means that m lies in $\ker\alpha$, which is the source of $\tilde{\varphi}$. Hence $n \in \operatorname{im}\tilde{\varphi}$.
- *Exactness at* $\ker\gamma$. If $p \in \operatorname{im}\tilde{\psi}$ then $p = \psi(n)$ for some $n \in \ker\beta$. In the construction (c) of δ above, we can then choose this element n as inverse image for p, so we get $n' = \beta(n) = 0$ and thus $\delta(p) = 0$, i. e. $p \in \ker\delta$. Conversely, if $p \in \ker\delta$ then $m' = \alpha(m)$ for some $m \in M$ in the construction of δ in (c). Continuing in the notation of (c), we then have $\beta(n) = n' = \varphi'(m') = \varphi'(\alpha(m)) = \beta(\varphi(m))$. Hence, $n - \varphi(m) \in \ker\beta$ with $\tilde{\psi}(n - \varphi(m)) = \psi(n) - \psi(\varphi(m)) = \psi(n) = p$, and thus $p \in \operatorname{im}\tilde{\psi}$.
- *Exactness at* $M'/\operatorname{im}\alpha$. If $\overline{m'} \in \operatorname{im}\delta$, then $\tilde{\varphi}'(\overline{m'}) = \overline{n'} = 0$ in the notation of (c) since $n' = \beta(n) \in \operatorname{im}\beta$. Conversely, if $\overline{m'} \in \ker\tilde{\varphi}'$, then $n' := \varphi'(m') \in \operatorname{im}\beta$, so $n' = \beta(n)$ for some $n \in N$. Then $p := \psi(n)$ satisfies $\gamma(p) = \gamma(\psi(n)) = \psi'(\beta(n)) = \psi'(n') = \psi'(\varphi'(m')) = 0$ and yields the given element $\overline{m'}$ as image under δ, so $\overline{m'} \in \operatorname{im}\delta$.
- *Exactness at* $N'/\operatorname{im}\beta$. If $\overline{n'} \in \operatorname{im}\tilde{\varphi}'$ then $\overline{n'} = \tilde{\varphi}'(\overline{m'})$ for some $m' \in M'$, and thus $\tilde{\psi}'(\overline{n'}) = \overline{\psi'(\varphi'(m'))} = 0$. Conversely, if $\overline{n'} \in \ker\tilde{\psi}'$ then $\psi'(n') \in \operatorname{im}\gamma$, so $\psi'(n') = \gamma(p)$ for some $p \in P$. As ψ is surjective, we can choose $n \in N$ with $\psi(n) = p$. Then $\psi'(n' - \beta(n)) = \gamma(p) - \gamma(\psi(n)) = 0$, and hence $n' - \beta(n) \in \ker\psi' = \operatorname{im}\varphi'$. We therefore find an element $m' \in M'$ with $\varphi'(m') = n' - \beta(n)$, and thus $\tilde{\varphi}'(\overline{m'}) = \overline{n' - \beta(n)} = \overline{n'}$.
- *Injectivity of* $\tilde{\varphi}$. As $\tilde{\varphi}$ is just a restriction of φ, it is clear that $\tilde{\varphi}$ is injective if φ is.
- *Surjectivity of* $\tilde{\psi}'$. If ψ' is surjective then for any $p' \in P'$ there is an element $n' \in N'$ with $\psi'(n') = p'$, and thus $\tilde{\psi}'(\overline{n'}) = \overline{p'}$. □

Exercise 4.9 ($\operatorname{Hom}(\cdot, N)$ is left exact)**.**

(a) Prove that a sequence

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} M_3 \longrightarrow 0$$

of R-modules is exact if and only if the sequence

$$0 \longrightarrow \operatorname{Hom}(M_3,N) \xrightarrow{\varphi_2^*} \operatorname{Hom}(M_2,N) \xrightarrow{\varphi_1^*} \operatorname{Hom}(M_1,N) \qquad (*)$$

is exact for every R-module N, where $\varphi_i^*(\varphi) = \varphi \circ \varphi_i$ for $i \in \{1,2\}$.

(b) Show that the statement of (a) is not true with an additional 0 at the left and right end, respectively — i. e. that φ_1^* need not be surjective if φ_1 is injective. We say that the operation $\operatorname{Hom}(\cdot,N)$ is *left exact* (because the sequence $(*)$ is exact with 0 at the left), but not exact.

After having seen several ways to construct exact sequences, let us now study what sort of information we can get out of them. The general philosophy is that an exact sequence with zero modules at the end is "almost enough" to determine any of its modules in terms of the others in the sequence. As a first statement in this direction, let us show that in order to compute the length of a module it is enough to have the module somewhere in an exact sequence in which the lengths of the other modules are known — at least if all involved lengths are finite.

Corollary 4.10. *Let*

$$0 \longrightarrow M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} \cdots \xrightarrow{\varphi_{n-1}} M_n \longrightarrow 0$$

be an exact sequence of R-modules of finite length. Then $\sum_{i=1}^n (-1)^i l(M_i) = 0$.

Proof. By Corollary 3.23 (b) applied to all morphisms $\varphi_1, \dots, \varphi_{n-1}$ and an index shift we get

$$\begin{aligned}\sum_{i=1}^{n-1} (-1)^i l(M_i) &= \sum_{i=1}^{n-1} (-1)^i \left(l(\ker \varphi_i) + l(\operatorname{im} \varphi_i)\right) \\ &= -l(\ker \varphi_1) + (-1)^{n-1} l(\operatorname{im} \varphi_{n-1}) + \sum_{i=2}^{n-1} (-1)^i \left(l(\ker \varphi_i) - l(\operatorname{im} \varphi_{i-1})\right).\end{aligned}$$

But $l(\ker \varphi_1) = 0$ since φ_1 is injective, $l(\operatorname{im} \varphi_{n-1}) = l(M_n)$ since φ_{n-1} is surjective, and $l(\ker \varphi_i) = l(\operatorname{im} \varphi_{i-1})$ for all $i = 2, \dots, n-1$ by the exactness of the sequence. Plugging this into the above formula, the result follows. □

Of course, knowing the length of a module does not mean that one knows the module up to isomorphism (as we have seen e. g. in Example 3.21). So let us now address the question to what extent a module in an exact sequence can be completely recovered from the other parts of the sequence. For simplicity, let us restrict to short exact sequences.

Example 4.11 (Recovering modules from an exact sequence)**.** Consider an exact sequence

$$0 \longrightarrow M \xrightarrow{\varphi} N \xrightarrow{\psi} P \longrightarrow 0$$

of R-modules, and let us ask whether we can determine one of the three modules if we know the rest of the sequence.

Of course, if we know M and N together with the map φ then P is uniquely determined by this data: as we must have $\operatorname{im} \psi = P$ and $\ker \psi = \operatorname{im} \varphi$, the homomorphism theorem of Proposition 3.10 (a) tells us that $P \cong N/\ker \psi = N/\operatorname{im} \varphi$. In the same way, M is uniquely determined if we know N and P together with the map ψ, since by the injectivity of φ we have $M \cong \operatorname{im} \varphi = \ker \psi$.

The most interesting question is thus if we can recover the middle term N if we know M and P (but none of the maps). The following two examples show that this is in general not possible.

(a) In any case, a possible way to obtain a short exact sequence from given modules M at the left and P at the right is

$$0 \longrightarrow M \xrightarrow{\varphi} M \oplus P \xrightarrow{\psi} P \longrightarrow 0,$$

where φ is the inclusion of M in $M \oplus P$, and ψ the projection of $M \oplus P$ onto P.

(b) There is an exact sequence of $\mathbb{Z}$-modules

$$0 \longrightarrow \mathbb{Z}_2 \xrightarrow{\cdot 2} \mathbb{Z}_4 \longrightarrow \mathbb{Z}_2 \longrightarrow 0,$$

in which the morphism from $\mathbb{Z}_4$ to $\mathbb{Z}_2$ is just the quotient map. Note that in contrast to (a) its middle term $\mathbb{Z}_4$ is not isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ as a $\mathbb{Z}$-module (i. e. as a group).

However, recovering the middle term in an exact sequence does work under a slightly stronger assumption: if we have two short exact sequences, *and one of them maps to the other* in the sense of the following lemma, then the two sequences must agree up to isomorphisms if they agree at any two of their modules.

Corollary 4.12 (5-Lemma). *Let*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M & \xrightarrow{\varphi} & N & \xrightarrow{\psi} & P & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & M' & \xrightarrow{\varphi'} & N' & \xrightarrow{\psi'} & P' & \longrightarrow & 0 \end{array}$$

be a commutative diagram of R-modules with exact rows. If two of the maps α, β, and γ are isomorphisms, then so is the third.

Proof. By the Snake Lemma 4.7 there is a long exact sequence

$$0 \longrightarrow \ker\alpha \longrightarrow \ker\beta \longrightarrow \ker\gamma \longrightarrow M'/\operatorname{im}\alpha \longrightarrow N'/\operatorname{im}\beta \longrightarrow P'/\operatorname{im}\gamma \longrightarrow 0.$$

If now e. g. α and γ are bijective then $\ker\alpha = \ker\gamma = M'/\operatorname{im}\alpha = P'/\operatorname{im}\gamma = 0$, and the sequence becomes

$$0 \longrightarrow 0 \longrightarrow \ker\beta \longrightarrow 0 \longrightarrow 0 \longrightarrow N'/\operatorname{im}\beta \longrightarrow 0 \longrightarrow 0.$$

By Example 4.2 (b) this means that $\ker\beta = N'/\operatorname{im}\beta = 0$, and so β is bijective as well.

Of course, an analogous argument applies if α and β, or β and γ are assumed to be bijective. □

Remark 4.13. There are various versions of the 5-Lemma in the literature. In fact, looking at the proof above we see that our assumptions could be relaxed: in order to show that β is an isomorphism if α and γ are, we do not need the additional zero modules on the left and right of the long exact sequence of the Snake Lemma 4.7. So for this case we do not have to require φ to be injective or ψ' to be surjective — this is only necessary if we want to conclude that α or γ is an isomorphism, respectively.

More generally, if one wants to show that the middle vertical map is an isomorphism, one can show that it is enough to assume a commutative diagram

$$\begin{array}{ccccccccc} M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & M_4 & \longrightarrow & M_5 \\ \downarrow{\scriptstyle \beta_1} & & \downarrow{\scriptstyle \beta_2} & & \downarrow{\scriptstyle \beta_3} & & \downarrow{\scriptstyle \beta_4} & & \downarrow{\scriptstyle \beta_5} \\ M_1' & \longrightarrow & M_2' & \longrightarrow & M_3' & \longrightarrow & M_4' & \longrightarrow & M_5' \end{array}$$

with exact rows in which the zero modules at the left and right ends have been replaced by arbitrary ones, and where now β_1, β_2, β_4, and β_5 are required to be isomorphisms. This version of the lemma is actually the reason for the name "5-Lemma".

Using Corollary 4.12, we can now give easy criteria to check whether a given short exact sequence $0 \longrightarrow M \longrightarrow N \longrightarrow P \longrightarrow 0$ is of the simple form as in Example 4.11 (a), i. e. whether the middle entry N is just the direct sum of M and P:

Corollary 4.14 (Splitting Lemma). *For a short exact sequence*

$$0 \longrightarrow M \xrightarrow{\varphi} N \xrightarrow{\psi} P \longrightarrow 0$$

the following three statements are equivalent:

(a) *$N \cong M \oplus P$, with φ being the inclusion of M in the first summand, and ψ the projection onto the second.*

(b) *φ has a left-sided inverse, i. e. there is a homomorphism $\alpha : N \to M$ such that $\alpha \circ \varphi = \mathrm{id}_M$.*

(c) *ψ has a right-sided inverse, i. e. there is a homomorphism $\beta : P \to N$ such that $\psi \circ \beta = \mathrm{id}_P$.*

*In this case, the sequence is called **split exact**.*

Proof.

(a) $\Rightarrow$ (b) and (c): Under the assumption (a) we can take α to be the projection from $M \oplus P$ onto M, and for β the inclusion of P in $M \oplus P$.

(b) $\Rightarrow$ (a): If α is a left-sided inverse of φ, the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M & \xrightarrow{\varphi} & N & \xrightarrow{\psi} & P & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \mathrm{id}} & & \downarrow{\scriptstyle (\alpha,\psi)} & & \downarrow{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & M & \longrightarrow & M \oplus P & \longrightarrow & P & \longrightarrow & 0 \end{array}$$

is commutative with exact rows. As the left and right vertical map are obviously isomorphisms, so is the middle one by Corollary 4.12.

(c) $\Rightarrow$ (a): If β is a right-sided inverse of ψ, the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M & \longrightarrow & M \oplus P & \longrightarrow & P & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \mathrm{id}} & & \downarrow{\scriptstyle \varphi+\beta} & & \downarrow{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & M & \xrightarrow{\varphi} & N & \xrightarrow{\psi} & P & \longrightarrow & 0 \end{array}$$

is commutative with exact rows, where $(\varphi+\beta)(m,p) := \varphi(m)+\beta(p)$. Again, as the left and right vertical map are isomorphisms, so is the middle one by Corollary 4.12. □

Example 4.15. Every short exact sequence

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

of vector spaces over a field K is split exact: if $(b_i)_{i \in I}$ is a basis of W we can pick inverse images $c_i \in \psi^{-1}(b_i)$ by the surjectivity of ψ. There is then a (unique) linear map $\beta : W \to V$ with $\beta(b_i) = c_i$ [G2, Corollary 16.27]. Hence $\psi \circ \beta = \mathrm{id}_W$, i. e. the above sequence is split exact. So by Corollary 4.14 we conclude that we always have $V \cong U \oplus W$ in the above sequence.

Exercise 4.16. Let I and J be ideals in a ring R.

(a) Prove that there is an exact sequence of R-modules (what are the maps?)

$$0 \longrightarrow I \cap J \longrightarrow I \oplus J \longrightarrow I + J \longrightarrow 0.$$

(b) Use the Snake Lemma 4.7 to deduce from this an exact sequence

$$0 \longrightarrow R/(I \cap J) \longrightarrow R/I \oplus R/J \longrightarrow R/(I+J) \longrightarrow 0.$$

(c) Show by example that the sequences of (a) and (b) are in general not split exact.

Exercise 4.17. Let N be a submodule of a finitely generated R-module M. In Exercise 3.5 (c) you have seen that N need not be finitely generated in this case.

However, prove now that N is finitely generated if it is the kernel of a surjective R-module homomorphism $\varphi : M \to R^n$ for some $n \in \mathbb{N}$.

(Hint: Show and use that the sequence $0 \longrightarrow N \longrightarrow M \xrightarrow{\varphi} R^n \longrightarrow 0$ is split exact.)

5. Tensor Products

In the last two chapters we have developed powerful methods to work with modules and linear maps between them. However, in practice bilinear (or more generally multilinear) maps are often needed as well, so let us have a look at them now. Luckily, it turns out that to study bilinear maps we do not have to start from scratch, but rather can reduce their theory to the linear case. More precisely, for given R-modules M and N we will construct another module named $M \otimes N$ — the so-called *tensor product* of M and N — such that *bilinear* maps from $M \times N$ to any other R-module P are in natural one-to-one correspondence with *linear* maps from $M \otimes N$ to P. So instead of bilinear maps from $M \times N$ we can then always consider linear maps from the tensor product, and thus use all the machinery that we have developed so far for homomorphisms.

As the construction of this tensor product is a bit lengthy, let us first give an easy example that should show the idea behind it.

Example 5.1 (Idea of tensor products)**.** Let M and N be finitely generated free modules over a ring R, and choose bases $B = (b_1, \dots, b_m)$ and $C = (c_1, \dots, c_n)$ of M and N, respectively. Then every bilinear map $\alpha : M \times N \to P$ to a third R-module P satisfies

$$\alpha(\lambda_1 b_1 + \dots + \lambda_m b_m, \mu_1 c_1 + \dots + \mu_n c_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \mu_j \, \alpha(b_i, c_j) \qquad (*)$$

for all $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_n \in R$. Hence α is uniquely determined by specifying the values $\alpha(b_i, c_j) \in P$. Conversely, any choice of these values gives rise to a well-defined bilinear map $\alpha : M \times N \to P$ by the above formula.

Now let F be a free R-module of rank $m \cdot n$. We denote a basis of this space by $b_i \otimes c_j$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ — so at this point this is just a name for a basis of this module, rather than an actual operation between elements of M and N. By the same argument as above, a linear map $\varphi : F \to P$ can then be specified uniquely by giving arbitrary images $\varphi(b_i \otimes c_j) \in P$ of the basis elements [G2, Corollary 16.27]. Putting both results together, we see that there is a one-to-one correspondence between bilinear maps $\alpha : M \times N \to P$ and linear maps $\varphi : F \to P$, given on the bases by $\alpha(b_i, c_j) = \varphi(b_i \otimes c_j)$: both maps can be specified by giving $m \cdot n$ arbitrary elements of P. So in the above sense F is a tensor product of M and N.

If M and N are no longer free, but still finitely generated, we can at least pick generators $(b_1, \dots, b_m)$ and $(c_1, \dots, c_n)$ of M and N, respectively. Then $(*)$ shows that any bilinear map $\alpha : M \times N \to P$ is still determined by the values $\alpha(b_i, c_j)$. But these values can no longer be chosen independently; they have to be compatible with the relations among the generators. For example, if we have the relation $2b_1 + 3b_2 = 0$ in M, we must have $2\alpha(b_1, c_j) + 3\alpha(b_2, c_j) = 0$ for all j in order to get a well-defined bilinear map α. For the tensor product, this means the following: if G is the submodule of F generated by all relations — so in our example we would take $2b_1 \otimes c_j + 3b_2 \otimes c_j$ for all j — then bilinear maps $\alpha : M \times N \to P$ now correspond exactly to those linear maps $\varphi : F \to P$ that are zero on G. As these are the same as linear maps from F/G to P, we can now take F/G to be our tensor product of M and N.

In fact, this idea of the construction of the tensor product should be clearly visible in the proof of Proposition 5.5 below. The main difference will be that, in order to avoid unnatural choices of generators, we will just take *all* elements of M and N as a generating set. This will lead to a huge module F, but also to a huge submodule G of relations among these generators, and so the quotient F/G will again be what we want.

07

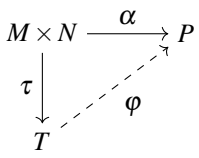
But let us now see how to obtain the tensor product $M \otimes N$ rigorously. There are two options for this: we can either construct it directly and then prove its properties, or define it to be a module having the

desired property — namely that linear maps from $M \otimes N$ are the same as bilinear maps from $M \times N$ — and show that such an object exists and is uniquely determined by this property. As this property is actually much more important than the technical construction of $M \otimes N$, we will take the latter approach.

Definition 5.2 (Tensor products)**.** Let M, N, and P be R-modules.

(a) A map $\alpha : M \times N \to P$ is called R-**bilinear** if $\alpha(\cdot, n) : M \to P$ and $\alpha(m, \cdot) : N \to P$ are R-linear for all $m \in M$ and $n \in N$.

(b) A **tensor product** of M and N over R is an R-module T together with a bilinear map $\tau : M \times N \to T$ such that the following *universal property* holds: for every bilinear map $\alpha : M \times N \to P$ to a third module P there is a unique linear map $\varphi : T \to P$ such that $\alpha = \varphi \circ \tau$, i. e. such that the diagram on the right commutes. The elements of a tensor product are called **tensors**.

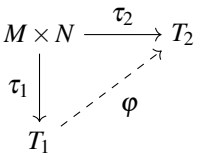


Remark 5.3. In the above notation, Definition 5.2 (b) just means that there is a one-to-one correspondence

$$\begin{array}{ccc} \{\text{bilinear maps } M \times N \to P\} & \overset{1:1}{\longleftrightarrow} & \{\text{homomorphisms } T \to P\} \\ \alpha & \longmapsto & \varphi \\ \varphi \circ \tau & \longmapsto & \varphi \end{array}$$

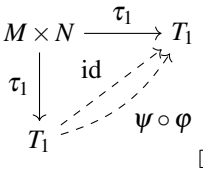
as explained in the motivation above.

Proposition 5.4 (Uniqueness of tensor products)**.** *A tensor product is unique up to unique isomorphism in the following sense: if T_1 and T_2 together with bilinear maps $\tau_1 : M \times N \to T_1$ and $\tau_2 : M \times N \to T_2$ are two tensor products for M and N over R, there is a unique R-module isomorphism $\varphi : T_1 \to T_2$ such that $\tau_2 = \varphi \circ \tau_1$.*



Proof. Consider the universal property of Definition 5.2 (b) for the first tensor product: as $\tau_2 : M \times N \to T_2$ is bilinear, there is a unique morphism $\varphi : T_1 \to T_2$ with $\tau_2 = \varphi \circ \tau_1$. In the same way, reversing the roles of the tensor products we get a unique morphism $\psi : T_2 \to T_1$ with $\tau_1 = \psi \circ \tau_2$.

Now apply the universal property for the first tensor product again, this time for the bilinear map $\tau_1 : M \times N \to T_1$ as shown on the right. Note that we have $\psi \circ \varphi \circ \tau_1 = \psi \circ \tau_2 = \tau_1$ as well as $\mathrm{id}_{T_1} \circ \tau_1 = \tau_1$, so that both $\psi \circ \varphi$ and id_{T_1} make the diagram commute. Hence, by the uniqueness part of the universal property we conclude that $\psi \circ \varphi = \mathrm{id}_{T_1}$. In the same way we see that $\varphi \circ \psi = \mathrm{id}_{T_2}$, and thus φ is an isomorphism. □



Proposition 5.5 (Existence of tensor products)**.** *Any two R-modules have a tensor product.*

Proof. Let M and N be R-modules. We denote by F the R-module of all finite formal linear combinations of elements of $M \times N$, i. e. formal sums of the form

$$a_1 (m_1, n_1) + \cdots + a_k (m_k, n_k)$$

for $k \in \mathbb{N}$, $a_1, \ldots, a_k \in R$, and distinct $(m_i, n_i) \in M \times N$ for $i = 1, \ldots, k$. More precisely, F can be modeled as the set of maps from $M \times N$ to R that have non-zero values at most at finitely many elements, where the values at these elements $(m_1, n_1), \ldots, (m_k, n_k) \in M \times N$ are $a_1, \ldots, a_k$ in the above notation. In this picture, the R-module structure of F is then given by pointwise addition and scalar multiplication. It is more intuitive however to think of the elements of F as linear combinations of elements of $M \times N$ as above (rather than as functions from $M \times N$ to R), and so we will use this notation in the rest of the proof.

Note that the various elements $(m, n) \in M \times N$ are by definition all independent in F — e. g. for given $a \in R$, $m \in M$, and $n \in N$ the linear combinations $a(m, n)$, $1(am, n)$, and $1(m, an)$ are in

general all different elements of F. In order to construct the tensor product we now want to enforce just enough relations so that the formal linear combinations become bilinear: let G be the submodule of F generated by all expressions

$$(m_1+m_2,n)-(m_1,n)-(m_2,n), \qquad (am,n)-a(m,n),$$
$$(m,n_1+n_2)-(m,n_1)-(m,n_2), \qquad \text{and } (m,an)-a(m,n)$$

for all $a \in R$, $m,m_1,m_2 \in M$, and $n,n_1,n_2 \in N$, and set $T := F/G$. Then the map

$$\tau : M \times N \to T, \ \ (m,n) \mapsto \overline{(m,n)}$$

is R-bilinear by the very definition of these relations.

We claim that T together with τ is a tensor product for M and N over R. So to check the universal property let $\alpha : M \times N \to P$ be an R-bilinear map. Then we can define a homomorphism $\varphi : T \to P$ by setting $\varphi\big(\overline{(m,n)}\big) := \alpha(m,n)$ and extending this by linearity, i. e.

$$\varphi\big(a_1\,\overline{(m_1,n_1)}+\cdots+a_k\,\overline{(m_k,n_k)}\big) = a_1\,\alpha(m_1,n_1)+\cdots+a_k\,\alpha(m_k,n_k).$$

Note that φ is well-defined since α is bilinear, and we certainly have $\alpha = \varphi \circ \tau$. Moreover, it is also obvious that setting $\varphi\big(\overline{(m,n)}\big) = \alpha(m,n)$ is the only possible choice such that $\alpha = \varphi \circ \tau$. Hence the universal property is satisfied, and T together with τ is indeed a tensor product. □

Notation 5.6 (Tensor products)**.** Let M and N be R-modules. By Propositions 5.4 and 5.5 there is a unique tensor product of M and N over R up to isomorphism, i. e. an R-module T together with a bilinear map $\tau : M \times N \to T$ satisfying the universal property of Definition 5.2 (b). We write T as $M \otimes_R N$ (or simply $M \otimes N$ if the base ring is understood), and $\tau(m,n)$ as $m \otimes n$. The element $m \otimes n \in M \otimes N$ is often called the tensor product of m and n.

Remark 5.7.

(a) Tensors in $M \otimes N$ that are of the form $m \otimes n$ for $m \in M$ and $n \in N$ are called **pure** or **monomial**. As we can see from the proof of Proposition 5.5, not every tensor in $M \otimes N$ is pure — instead, the pure tensors generate $M \otimes N$ as an R-module, i. e. a general element of $M \otimes N$ can be written as a finite linear combination $\sum_{i=1}^{k} a_i\,(m_i \otimes n_i)$ for $k \in \mathbb{N}$, $a_1,\dots,a_k \in R$, $m_1,\dots,m_k \in M$, and $n_1,\dots,n_k \in N$.

Note that these generators are not independent, so that there are in general many different ways to write a tensor as a linear combination of pure tensors. This makes it often a non-trivial task to decide whether two such linear combinations are the same tensor or not.

(b) The tensor product of two elements of M and N is bilinear by Definition 5.2 (b), i. e. we have

$$(m_1+m_2)\otimes n = m_1 \otimes n + m_2 \otimes n \quad \text{and} \quad a\,(m \otimes n) = (am) \otimes n$$

in $M \otimes N$ for all $a \in R$, $m,m_1,m_2 \in M$, and $n \in N$, and similarly for the second factor. In fact, the tensor product has been defined in such a way that the relations among tensors are *exactly* those bilinear ones.

(c) Of course, using multilinear instead of bilinear maps, we can also define tensor products $M_1 \otimes \cdots \otimes M_k$ of more than two modules in the same way as above. We will see in Exercise 5.9 however that the result is nothing but a repeated application of the tensor product for bilinear maps.

Before we give some examples of tensor product spaces, let us first prove a few simple properties that will also make the study of the examples easier.

Lemma 5.8. *For any R-modules M, N, and P there are natural isomorphisms*

(a) $M \otimes N \cong N \otimes M$;

(b) $M \otimes R \cong M$;

(c) $(M \oplus N) \otimes P \cong (M \otimes P) \oplus (N \otimes P)$.

Proof. The strategy for all three proofs is the same: using the universal property of Definition 5.2 (b) we construct maps between the tensor products from bilinear maps, and then show that they are inverse to each other.

(a) The map $M \times N \to N \otimes M,\ (m,n) \mapsto n \otimes m$ is bilinear by Remark 5.7 (b), and thus induces a (unique) linear map

$$\varphi : M \otimes N \to N \otimes M \quad \text{with} \quad \varphi(m \otimes n) = n \otimes m \quad \text{for all } m \in M \text{ and } n \in N$$

by the universal property of $M \otimes N$. In the same way we get a morphism $\psi : N \otimes M \to M \otimes N$ with $\psi(n \otimes m) = m \otimes n$. Then $(\psi \circ \varphi)(m \otimes n) = m \otimes n$ for all $m \in M$ and $n \in N$, so $\psi \circ \varphi$ is the identity on pure tensors. But the pure tensors generate $M \otimes N$, and so we must have $\psi \circ \varphi = \mathrm{id}_{M \otimes N}$. In the same way we see that $\varphi \circ \psi = \mathrm{id}_{N \otimes M}$. Hence φ is an isomorphism.

(b) From the bilinear map $M \times R \to M,\ (m,a) \mapsto am$ we obtain a linear map

$$\varphi : M \otimes R \to M \quad \text{with} \quad \varphi(m \otimes a) = am \quad \text{for all } m \in M \text{ and } a \in R$$

by the universal property. Furthermore, there is a linear map $\psi : M \to M \otimes R,\ m \mapsto m \otimes 1$. As

$$(\psi \circ \varphi)(m \otimes a) = am \otimes 1 = m \otimes a \quad \text{and} \quad (\varphi \circ \psi)(m) = m$$

for all $m \in M$ and $a \in R$, we conclude as in (a) that φ is an isomorphism.

(c) The bilinear map $(M \oplus N) \times P \to (M \otimes P) \oplus (N \otimes P),\ ((m,n),p) \mapsto (m \otimes p, n \otimes p)$ induces a linear map

$$\varphi : (M \oplus N) \otimes P \to (M \otimes P) \oplus (N \otimes P) \quad \text{with} \quad \varphi((m,n) \otimes p) = (m \otimes p, n \otimes p)$$

as above. Similarly, we get morphisms $M \otimes P \to (M \oplus N) \otimes P$ with $m \otimes p \mapsto (m,0) \otimes p$ and $N \otimes P \to (M \oplus N) \otimes P$ with $n \otimes p \mapsto (0,n) \otimes p$, and thus by addition a linear map

$$\psi : (M \otimes P) \oplus (N \otimes P) \to (M \oplus N) \otimes P \quad \text{with} \quad \psi(m \otimes p, n \otimes q) = (m,0) \otimes p + (0,n) \otimes q.$$

Is is verified immediately that φ and ψ are inverse to each other on pure tensors, and thus also on the whole tensor product space. □

Exercise 5.9 (Associativity of tensor products)**.** Let M, N, and P be three R-modules. Prove that there are natural isomorphisms

$$M \otimes N \otimes P \cong (M \otimes N) \otimes P \cong M \otimes (N \otimes P)$$

where $M \otimes N \otimes P$ is the tensor product constructed from trilinear maps as in Remark 5.7 (c).

Example 5.10.

(a) Let M and N be free R-modules of ranks m and n, respectively. Then $M \cong R^m$ and $N \cong R^n$, and so by Lemma 5.8 we have

$$M \otimes N \cong R^m \otimes (\underbrace{R \oplus \cdots \oplus R}_{n \text{ times}}) \cong (R^m \otimes R) \oplus \cdots \oplus (R^m \otimes R) \cong R^m \oplus \cdots \oplus R^m \cong R^{mn},$$

as expected from Example 5.1. So after a choice of bases the elements of $M \otimes N$ can be described by $m \times n$-matrices over R.

(b) Let I and J be coprime ideals in a ring R. Then there are elements $a \in I$ and $b \in J$ with $a + b = 1$, and so we obtain in the tensor product $R/I \otimes R/J$ for all monomial tensors $\overline{r} \otimes \overline{s}$ with $r, s \in R$

$$\overline{r} \otimes \overline{s} = (a+b)(\overline{r} \otimes \overline{s}) = \overline{ar} \otimes \overline{s} + \overline{r} \otimes \overline{bs} = 0$$

since $\overline{ar} = 0 \in R/I$ and $\overline{bs} = 0 \in R/J$. But these monomial tensors generate the tensor product, and so we conclude that $R/I \otimes R/J = 0$. As a concrete example, we get $\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathbb{Z}_q = 0$ for any two distinct primes p and q.

This shows that (in contrast to (a)) a tensor product space need not be "bigger" than its factors, and might be 0 even if none of its factors are.

(c) In the tensor product $\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2$ we have

$$2 \otimes \overline{1} = 1 \otimes \overline{2} = 0$$

by bilinearity. However, if we now consider $2 \otimes \overline{1}$ as an element of $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2$, the above computation is invalid since $1 \notin 2\mathbb{Z}$. So is it still true that $2 \otimes \overline{1} = 0$ in $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2$?

We can answer this question with Lemma 5.8 (b): we know that $2\mathbb{Z}$ is isomorphic to $\mathbb{Z}$ as a $\mathbb{Z}$-module by sending 2 to 1, and so $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2$ is isomorphic to $\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2 \cong \mathbb{Z}_2$ by the map $\varphi : 2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_2 \to \mathbb{Z}_2$, $2a \otimes \overline{b} \mapsto \overline{ab}$. But now $\varphi(2 \otimes \overline{1}) = \overline{1} \neq 0$, and so indeed we have $2 \otimes \overline{1} \neq 0$ in $2\mathbb{Z} \otimes \mathbb{Z}_2$.

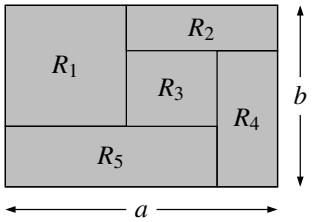
The conclusion is that, when writing down tensor products $m \otimes n$, we have to be very careful to specify which tensor product space we consider: if $n \in N$ and m lies in a submodule M' of a module M, it might happen that $m \otimes n$ is non-zero in $M' \otimes N$, but zero in $M \otimes N$. In other words, for a submodule M' of M it is not true in general that $M' \otimes N$ is a submodule of $M \otimes N$ in a natural way! We will discuss this issue in more detail in Proposition 5.22 (b) and Remark 5.23.

Exercise 5.11. Compute the tensor products $\mathbb{Q}/\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Q}/\mathbb{Z}$, $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Q}$, $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$, and $\mathbb{Q}[x] \otimes_{\mathbb{Q}} \mathbb{C}$.

Exercise 5.12. Assume that we have n rectangles $R_1, \dots, R_n$ in the plane, of size $a_i \times b_i$ for $i = 1, \dots, n$, that fit together to form a rectangle R of size $a \times b$ as in the picture on the right. Prove:

(a) $a \otimes b = \sum_{i=1}^{n} a_i \otimes b_i$ in $\mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R}$.

(b) If each of the rectangles $R_1, \dots, R_n$ has at least one side with a rational length, then R must also have at least one side with a rational length.



Exercise 5.13 (Dual vector spaces)**.** Let V and W be vector spaces over a field K. We call $V^* := \mathrm{Hom}_K(V, K)$ the *dual vector space* to V. Moreover, denote by $\mathrm{BLF}(V)$ the vector space of bilinear forms $V \times V \to K$.

(a) Show that there are (natural, i. e. basis-independent) linear maps

$$\begin{aligned} &\Phi : V^* \otimes W \to \mathrm{Hom}(V, W) && \text{such that } \Phi(\varphi \otimes w)(v) = \varphi(v) \cdot w, \\ &\Psi : V^* \otimes V^* \to \mathrm{BLF}(V) && \text{such that } \Psi(\varphi \otimes \psi)(v, v') = \varphi(v) \cdot \psi(v'), \\ &T : V^* \otimes V \to K && \text{such that } T(\varphi \otimes v) = \varphi(v). \end{aligned}$$

(b) Prove that Φ and Ψ are injective.

(c) Prove that Φ and Ψ are in fact isomorphisms if V and W are finite-dimensional, but not in general for arbitrary vector spaces.

(d) Assume that $\dim_K V = n < \infty$, so that $V^* \otimes V$ is naturally isomorphic to $\mathrm{Hom}_K(V, V)$ by (c), which in turn is isomorphic to $\mathrm{Mat}(n \times n, K)$ in the standard way after choosing a basis of V. Using these isomorphisms, T becomes a linear map $\mathrm{Mat}(n \times n, K) \to K$ that is invariant under a change of basis. Which one?

08

To define homomorphisms between tensor products the following construction will be useful.

Construction 5.14 (Tensor product of homomorphisms)**.** Let $\varphi : M \to N$ and $\varphi' : M' \to N'$ be two homomorphisms of R-modules. Then the map $M \times M' \to N \otimes N'$, $(m, m') \mapsto \varphi(m) \otimes \varphi'(m')$ is bilinear, and thus by the universal property of the tensor product gives rise to a homomorphism

$$\varphi \otimes \varphi' : M \otimes M' \to N \otimes N' \quad \text{such that} \quad (\varphi \otimes \varphi')(m \otimes m') = \varphi(m) \otimes \varphi'(m')$$

for all $m \in M$ and $m' \in M'$. We call $\varphi \otimes \varphi'$ the **tensor product** of φ and φ'.

Remark 5.15. Note that for $\varphi \in \mathrm{Hom}_R(M,N)$ and $\varphi' \in \mathrm{Hom}_R(M',N')$ we have already defined a tensor product $\varphi \otimes \varphi'$ as an element of $\mathrm{Hom}_R(M,N) \otimes_R \mathrm{Hom}_R(M',N')$ in Notation 5.6 — whereas Construction 5.14 gives an element of $\mathrm{Hom}_R(M \otimes_R M', N \otimes_R N')$. In fact, it is easy to see by the universal property of the tensor product that there is a natural homomorphism

$$\mathrm{Hom}_R(M,N) \otimes_R \mathrm{Hom}_R(M',N') \to \mathrm{Hom}_R(M \otimes_R M', N \otimes_R N')$$

that sends $\varphi \otimes \varphi'$ in the sense of Notation 5.6 to the morphism of Construction 5.14. It should therefore not lead to confusion if we denote both constructions by $\varphi \otimes \varphi'$.

One application of tensor products is to extend the ring of scalars for a given module. For vector spaces, this is a process that you know very well: suppose that we have e. g. a real vector space V with $\dim_{\mathbb{R}} V = n < \infty$ and want to study the eigenvalues and eigenvectors of a linear map $\varphi : V \to V$. We then usually set up the matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ corresponding to φ in some chosen basis, and compute its characteristic polynomial. Often it happens that this polynomial does not split into linear factors over $\mathbb{R}$, and that we therefore want to pass from the real to the complex numbers.

But while it is perfectly possible to consider A as a complex matrix in $\mathrm{Mat}(n \times n, \mathbb{C})$ instead and talk about complex eigenvalues and eigenvectors of A, it is not clear what this means in the language of the linear map φ: in the condition $\varphi(x) = \lambda x$ for an eigenvector of φ it certainly does not make sense to take x to be a "complex linear combination" of the basis vectors, since such an element does not exist in V, and so φ is not defined on it. We rather have to extend V first to a complex vector space, and φ to a $\mathbb{C}$-linear map on this extension. It turns out that the tensor product of V with $\mathbb{C}$ over $\mathbb{R}$ is exactly the right construction to achieve this in a basis-independent language.

Construction 5.16 (Extension of scalars)**.** Let M be an R-module, and R' an R-algebra (so that R' is a ring as well as an R-module). Moreover, for any $a \in R'$ we denote by $\mu_a : R' \to R'$, $s \mapsto as$ the multiplication map, which is obviously a homomorphism of R-modules. If we then set $M_{R'} := M \otimes_R R'$, we obtain a scalar multiplication with R' on $M_{R'}$

$$\begin{aligned} R' \times M_{R'} &\to M_{R'} \\ (a, m \otimes s) &\mapsto a \cdot (m \otimes s) := (1 \otimes \mu_a)(m \otimes s) = m \otimes (as) \end{aligned}$$

which turns $M_{R'}$ into an R'-module. We say that $M_{R'}$ is obtained from M by an **extension of scalars** from R to R'. Note that any R-module homomorphism $\varphi : M \to N$ then gives rise to an "extended" R'-module homomorphism $\varphi_{R'} := \varphi \otimes \mathrm{id} : M_{R'} \to N_{R'}$.

Example 5.17 (Complexification of a real vector space)**.** Let V be a real vector space. Extending scalars from $\mathbb{R}$ to $\mathbb{C}$, we call $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$ the *complexification* of V. By Construction 5.16, this is now a complex vector space.

Let us assume for simplicity that V is finitely generated, with basis $(b_1, \dots, b_n)$. Then $V \cong \mathbb{R}^n$ by an isomorphism that maps b_i to the i-th standard basis vector e_i for $i = 1, \dots, n$, and consequently

$$V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{R}^n \otimes_{\mathbb{R}} \mathbb{C} \cong (\mathbb{R} \otimes_{\mathbb{R}} \mathbb{C})^n \cong \mathbb{C}^n$$

as $\mathbb{R}$-modules by Lemma 5.8. But by definition of the scalar multiplication with $\mathbb{C}$ from Construction 5.16 this is also a $\mathbb{C}$-module homomorphism, and thus an isomorphism of complex vector spaces. Moreover, $b_i \otimes 1$ maps to e_i under this chain of isomorphisms for all i, and so as expected the vectors $b_1 \otimes 1, \dots, b_n \otimes 1$ form a basis of the complexified vector space $V_{\mathbb{C}}$.

Finally, let us consider a linear map $\varphi : V \to V$ described by the matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ with respect to the basis $(b_1, \dots, b_n)$, i. e. we have $\varphi(b_i) = \sum_{j=1}^n a_{j,i} b_j$ for all i. Then $\varphi_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$ from Construction 5.16 is an endomorphism of the complexified vector space, and since

$$\varphi_{\mathbb{C}}(b_i \otimes 1) = \varphi(b_i) \otimes 1 = \sum_{j=1}^n a_{j,i} (b_j \otimes 1)$$

we see that the matrix of $\varphi_{\mathbb{C}}$ with respect to the basis $(b_1 \otimes 1, \dots, b_n \otimes 1)$ is precisely A again, just now considered as a complex matrix.

Of course, the same constructions can not only be used to pass from the real to the complex numbers, but also for any field extension.

Exercise 5.18. Let M and N be R-modules, and R' an R-algebra. Show that the extension of scalars commutes with tensor products in the sense that there is an isomorphism of R'-modules

$$(M \otimes_R N)_{R'} \cong M_{R'} \otimes_{R'} N_{R'}.$$

In Construction 5.16 we have considered the case when one of the two factors in a tensor product over a ring R is not only an R-module, but also an R-algebra. If both factors are R-algebras, we can say even more: in this case, the resulting tensor product will also be an R-algebra in a natural way:

Construction 5.19 (Tensor product of algebras)**.** Let R be a ring, and let R_1 and R_2 be R-algebras. Then the map $R_1 \times R_2 \times R_1 \times R_2 \to R_1 \otimes R_2$, $(s,t,s',t') \mapsto (ss') \otimes (tt')$ is multilinear, and so by the universal property of the tensor product (and its associativity as in Exercise 5.9) it induces a homomorphism

$$(R_1 \otimes R_2) \otimes (R_1 \otimes R_2) \to R_1 \otimes R_2 \quad \text{with} \quad (s \otimes t) \otimes (s' \otimes t') \mapsto (ss') \otimes (tt')$$

for all $s,s' \in R_1$ and $t,t' \in R_2$. Again by the universal property of the tensor product this now corresponds to a bilinear map

$$(R_1 \otimes R_2) \times (R_1 \otimes R_2) \to R_1 \otimes R_2 \quad \text{with} \quad (s \otimes t, s' \otimes t') \mapsto (s \otimes t) \cdot (s' \otimes t') := (ss') \otimes (tt').$$

It is obvious that this multiplication makes $R_1 \otimes R_2$ into a ring, and thus into an R-algebra. So the tensor product of two R-algebras has again a natural structure of an R-algebra.

Example 5.20 (Multivariate polynomial rings as tensor products)**.** Let R be a ring. We claim that $R[x,y] \cong R[x] \otimes_R R[y]$ as R-algebras, i. e. that polynomial rings in several variables can be thought of as tensor products of polynomial rings in one variable.

In fact, there are R-module homomorphisms

$$\varphi : R[x] \otimes_R R[y] \to R[x,y], \quad f \otimes g \mapsto fg$$

(by the universal property of the tensor product) and

$$\psi : R[x,y] \mapsto R[x] \otimes R[y], \quad \sum_{i,j} a_{i,j} x^i y^j \mapsto \sum_{i,j} a_{i,j} x^i \otimes y^j.$$

As

$$(\psi \circ \varphi)(x^i \otimes y^j) = \psi(x^i y^j) = x^i \otimes y^j$$
$$\text{and} \quad (\varphi \circ \psi)(x^i y^j) = \varphi(x^i \otimes y^j) = x^i y^j$$

for all $i,j \in \mathbb{N}$ and these elements $x^i \otimes y^j$ and $x^i y^j$ generate $R[x] \otimes_R R[y]$ and $R[x,y]$ as an R-module, respectively, we see that φ and ψ are inverse to each other. Moreover, φ is also a ring homomorphism with the multiplication in $R[x] \otimes_R R[y]$ of Construction 5.19, since

$$\varphi((f \otimes g) \cdot (f' \otimes g')) = \varphi((ff') \otimes (gg')) = ff'gg' = \varphi(f \otimes g) \cdot \varphi(f' \otimes g').$$

Hence $R[x,y] \cong R[x] \otimes_R R[y]$ as R-algebras.

Exercise 5.21.

(a) Let I and J be ideals in a ring R. Prove that $R/I \otimes_R R/J \cong R/(I+J)$ as R-algebras.

(b) Let $X \subset \mathbb{A}^n_K$ and $Y \subset \mathbb{A}^m_K$ be varieties over a field K, so that $X \times Y \subset K^{n+m}$. Show that $X \times Y$ is again a variety, and $A(X \times Y) \cong A(X) \otimes_K A(Y)$ as K-algebras.

Finally, to conclude this chapter we want to study how tensor products behave in exact sequences. The easiest way to see this is to trace it back to Exercise 4.9, in which we applied $\mathrm{Hom}_R(\cdot, N)$ to an exact sequence.

Proposition 5.22 (Tensor products are right exact)**.**

(a) *For any R-modules M, N, and P we have* $\mathrm{Hom}(M, \mathrm{Hom}(N,P)) \cong \mathrm{Hom}(M \otimes N, P)$.

(b) *Let*

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} M_3 \longrightarrow 0$$

be an exact sequence of R-modules. Then for any R-module N the sequence

$$M_1 \otimes N \xrightarrow{\varphi_1 \otimes \mathrm{id}} M_2 \otimes N \xrightarrow{\varphi_2 \otimes \mathrm{id}} M_3 \otimes N \longrightarrow 0$$

is exact as well.

Proof.

(a) A natural isomorphism can be constructed by identifying $\alpha \in \mathrm{Hom}(M, \mathrm{Hom}(N,P))$ with $\beta \in \mathrm{Hom}(M \otimes N, P)$, where

$$\alpha(m)(n) = \beta(m \otimes n) \quad \in P$$

for all $m \in M$ and $n \in N$. Note that this equation can be used to define α in terms of β to get a map $\mathrm{Hom}(M \otimes N, P) \to \mathrm{Hom}(M, \mathrm{Hom}(N,P))$, and also (by the universal property of the tensor product) to define β in terms of α in order to get a map in the opposite direction. Obviously, these two maps are then R-linear and inverse to each other.

(b) Starting from the given sequence, Exercise 4.9 (a) gives us an exact sequence

$$0 \longrightarrow \mathrm{Hom}(M_3, \mathrm{Hom}(N,P)) \longrightarrow \mathrm{Hom}(M_2, \mathrm{Hom}(N,P)) \longrightarrow \mathrm{Hom}(M_1, \mathrm{Hom}(N,P))$$

for all R-modules N and P, where the two non-trivial maps send $\alpha_i \in \mathrm{Hom}(M_i, \mathrm{Hom}(N,P))$ to $\alpha_{i-1} \in \mathrm{Hom}(M_{i-1}, \mathrm{Hom}(N,P))$ with $\alpha_{i-1}(m_{i-1})(p) = \alpha_i(\varphi_{i-1}(m_{i-1}))(p)$ for $i \in \{2,3\}$ and all $m_1 \in M_1$, $m_2 \in M_2$, $p \in P$. Using the isomorphism of (a), this is the same as an exact sequence

$$0 \longrightarrow \mathrm{Hom}(M_3 \otimes N, P) \longrightarrow \mathrm{Hom}(M_2 \otimes N, P) \longrightarrow \mathrm{Hom}(M_1 \otimes N, P),$$

where the maps are now $\beta_i \mapsto \beta_{i-1}$ with

$$\beta_{i-1}(m_{i-1} \otimes p) = \beta_i(\varphi_{i-1}(m_{i-1}) \otimes p) = \beta_i((\varphi_{i-1} \otimes \mathrm{id})(m_{i-1} \otimes p))$$

for $i \in \{2,3\}$. But using Exercise 4.9 (a) again, this means that the sequence

$$M_1 \otimes N \xrightarrow{\varphi_1 \otimes \mathrm{id}} M_2 \otimes N \xrightarrow{\varphi_2 \otimes \mathrm{id}} M_3 \otimes N \longrightarrow 0$$

is also exact. □

Remark 5.23. Similarly to the case of $\mathrm{Hom}(\cdot, N)$ in Exercise 4.9, the statement of Proposition 5.22 (b) is in general not true with an additional zero module at the left, i. e. if φ_1 is injective it does not necessarily follow that $\varphi_1 \otimes \mathrm{id}$ is injective. In fact, we know this already from Example 5.10 (c), where we have seen that for a submodule M_1 of M_2 we cannot conclude that $M_1 \otimes N$ is a submodule of $M_2 \otimes N$ in a natural way.

In analogy to Exercise 4.9 we say that taking tensor products is *right exact*, but not exact. However, we have the following result:

Exercise 5.24. Let $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ be a short exact sequence of R-modules. Show that the sequence $0 \longrightarrow M_1 \otimes N \longrightarrow M_2 \otimes N \longrightarrow M_3 \otimes N \longrightarrow 0$ is also exact if one of the following assumptions hold:

(a) the sequence $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ is split exact;

(b) N is a finitely generated, free R-module.

As an example of how Proposition 5.22 can be used, let us prove the following statement.

Corollary 5.25. *Let I be an ideal in a ring R, and let M be an R-module. Then $M/IM \cong M \otimes_R R/I$.*

Proof. By Example 4.3 (b) we know that the sequence

$$0 \longrightarrow I \longrightarrow R \longrightarrow R/I \longrightarrow 0$$

is exact. So by Proposition 5.22 (b) it follows that

$$M \otimes_R I \xrightarrow{\varphi} M \otimes_R R \xrightarrow{\psi} M \otimes_R R/I \longrightarrow 0$$

is also exact, where ψ is the projection in the second factor, $M \otimes_R R \cong M$ by Lemma 5.8 (b), and $\varphi(m \otimes a) = am$ using this isomorphism. In particular, the image of φ is just IM by Definition 3.12 (a), and so we conclude by the exactness of the sequence and the homomorphism theorem

$$M \otimes_R R/I = \operatorname{im} \psi \cong M/\ker \psi = M/\operatorname{im} \varphi = M/IM. \qquad \square$$

Example 5.26 (Derivatives in terms of tensor products)**.** Let V and W be normed real vector spaces [G2, Definition 23.1], and let $f : U \to W$ be a function on an open subset $U \subset V$. Then for any point $a \in U$ the derivative $f'(a)$ of f in a (if it exists) is an element of $\operatorname{Hom}_{\mathbb{R}}(V,W)$: it is just the homomorphism such that $x \mapsto f(a) + f'(a)(x-a)$ is the affine-linear approximation of f at a [G2, Definition 25.3 and Remark 25.6].

If we now want to define the second derivative f'' of f, the most natural way to do this is to take the derivative of the map $f' : U \to \operatorname{Hom}_{\mathbb{R}}(V,W)$, $a \mapsto f'(a)$, with a suitable norm on $\operatorname{Hom}_{\mathbb{R}}(V,W)$. By the same reasoning as above, this will now lead to an element $f''(a)$ of $\operatorname{Hom}_{\mathbb{R}}(V, \operatorname{Hom}_{\mathbb{R}}(V,W))$ [G2, Remark 26.5 (a)]. Similarly, the third derivative $f'''(a)$ is an element of $\operatorname{Hom}_{\mathbb{R}}(V, \operatorname{Hom}_{\mathbb{R}}(V, \operatorname{Hom}_{\mathbb{R}}(V,W)))$, and so on.

With Proposition 5.22 (a) we can now rephrase this in a simpler way in terms of tensor products: the k-th derivative $f^{(k)}(a)$ of f in a point $a \in U$ is an element of $\operatorname{Hom}_{\mathbb{R}}(V \otimes_{\mathbb{R}} \cdots \otimes_{\mathbb{R}} V, W)$ for $k \in \mathbb{N}_{>0}$, where we take the k-fold tensor product of V with itself. So the higher derivatives of f can again be thought of as linear maps, just with a tensor product source space. Of course, if V and W are finite-dimensional with $n = \dim_{\mathbb{R}} V$ and $m = \dim_{\mathbb{R}} W$, then $\operatorname{Hom}_{\mathbb{R}}(V \otimes_{\mathbb{R}} \cdots \otimes_{\mathbb{R}} V, W)$ is of dimension $n^k m$, and the coordinates of $f^{(k)}(a)$ with respect to bases of V and W are simply the n^k partial derivatives of order k of the m coordinate functions of f.

Exercise 5.27. Show that $l(M \otimes N) \leq l(M) \cdot l(N)$ for any two R-modules M and N (where an expression $0 \cdot \infty$ on the right hand side is to be interpreted as 0). Does equality hold in general?

(Hint: It is useful to consider suitable exact sequences.)

09

6. Localization

Localization is a very powerful technique in commutative algebra that often allows to reduce questions on rings and modules to a union of smaller "local" problems. It can easily be motivated both from an algebraic and a geometric point of view, so let us start by explaining the idea behind it in these two settings.

Remark 6.1 (Motivation for localization)**.**

(a) *Algebraic motivation:* Let R be a ring which is not a field, i. e. in which not all non-zero elements are units. The algebraic idea of localization is then to make more (or even all) non-zero elements invertible by introducing fractions, in the same way as one passes from the integers $\mathbb{Z}$ to the rational numbers $\mathbb{Q}$.

Let us have a more precise look at this particular example: in order to construct the rational numbers from the integers we start with $R = \mathbb{Z}$, and let $S = \mathbb{Z}\backslash\{0\}$ be the subset of the elements of R that we would like to become invertible. On the set $R \times S$ we then consider the equivalence relation

$$(a,s) \sim (a',s') \quad \Leftrightarrow \quad as' - a's = 0$$

and denote the equivalence class of a pair (a,s) by $\frac{a}{s}$. The set of these "fractions" is then obviously $\mathbb{Q}$, and we can define addition and multiplication on it in the expected way by $\frac{a}{s} + \frac{a'}{s'} := \frac{as'+a's}{ss'}$ and $\frac{a}{s} \cdot \frac{a'}{s'} := \frac{aa'}{ss'}$.

(b) *Geometric motivation:* Now let $R = A(X)$ be the ring of polynomial functions on a variety X. In the same way as in (a) we can ask if it makes sense to consider fractions of such polynomials, i. e. rational functions

$$X \to K,\ x \mapsto \frac{f(x)}{g(x)}$$

for $f,g \in R$. Of course, for global functions on all of X this does not work, since g will in general have zeroes somewhere, so that the value of the rational function is ill-defined at these points. But if we consider functions on subsets of X we can allow such fractions. The most important example from the point of view of localization is the following: for a fixed point $a \in X$ let $S = \{f \in A(X) : f(a) \neq 0\}$ be the set of all polynomial functions that do not vanish at a. Then the fractions $\frac{f}{g}$ for $f \in R$ and $g \in S$ can be thought of as rational functions that are well-defined at a.

In fact, the best way to interpret the fractions $\frac{f}{g}$ with $f \in R$ and $g \in S$ is as "functions on an arbitrarily small neighborhood of a", assuming that we have a suitable topology on X: although the only point at which these functions are guaranteed to be well-defined is a, for every such function there is a neighborhood on which it is defined as well, and to a certain extent the polynomials f and g can be recovered from the values of the function on such a neighborhood. In this sense we will refer to functions of the form $\frac{f}{g}$ with $f \in R$ and $g \in S$ as "*local functions*" at a. In fact, this geometric interpretation is the origin of the name "localization" for the process of constructing fractions described in this chapter.

Let us now introduce such fractions in a rigorous way. As above, R will always denote the original ring, and $S \subset R$ the set of elements that we would like to become invertible. Note that this subset S of denominators for our fractions has to be closed under multiplication, for otherwise the formulas $\frac{a}{s} + \frac{a'}{s'} := \frac{as'+a's}{ss'}$ and $\frac{a}{s} \cdot \frac{a'}{s'} := \frac{aa'}{ss'}$ for addition and multiplication of fractions would not make sense. Moreover, the equivalence relation on $R \times S$ to obtain fractions will be slightly different from the one in Remark 6.1 (a) — we will explain the reason for this later in Remark 6.3.

Lemma and Definition 6.2 (Localization of rings)**.** *Let R be a ring.*

(a) *A subset $S \subset R$ is called **multiplicatively closed** if $1 \in S$, and $ab \in S$ for all $a, b \in S$.*

(b) *Let $S \subset R$ be a multiplicatively closed set. Then*

$$(a,s) \sim (a',s') \quad :\Leftrightarrow \quad \text{there is an element } u \in S \text{ such that } u(as' - a's) = 0$$

is an equivalence relation on $R \times S$. We denote the equivalence class of a pair $(a,s) \in R \times S$ by $\frac{a}{s}$. The set of all equivalence classes

$$S^{-1}R := \left\{ \frac{a}{s} : a \in R, s \in S \right\}$$

*is called the **localization** of R **at the multiplicatively closed set** S. It is a ring together with the addition and multiplication*

$$\frac{a}{s} + \frac{a'}{s'} := \frac{as' + a's}{ss'} \quad \text{and} \quad \frac{a}{s} \cdot \frac{a'}{s'} := \frac{aa'}{ss'}.$$

Proof. The relation $\sim$ is clearly reflexive and symmetric. It is also transitive: if $(a,s) \sim (a',s')$ and $(a',s') \sim (a'',s'')$ there are $u, v \in S$ with $u(as' - a's) = v(a's'' - a''s') = 0$, and thus

$$s'uv(as'' - a''s) = vs'' \cdot u(as' - a's) + us \cdot v(a's'' - a''s') = 0. \qquad (*)$$

But this means that $(a,s) \sim (a'',s'')$, since S is multiplicatively closed and therefore $s'uv \in S$.

In a similar way we can check that the addition and multiplication in $S^{-1}R$ are well-defined: if $(a,s) \sim (a'',s'')$, i. e. $u(as'' - a''s) = 0$ for some $u \in S$, then

$$u\big((as' + a's)(s''s') - (a''s' + a's'')(ss')\big) = (s')^2 \cdot u(as'' - a''s) = 0$$
$$\text{and} \quad u\big((aa')(s''s') - (a''a')(ss')\big) = a's' \cdot u(as'' - a''s) = 0,$$

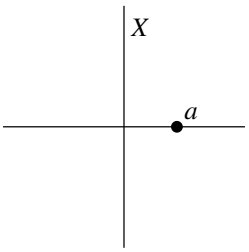
and so $\frac{as'+a's}{ss'} = \frac{a''s'+a's''}{s''s'}$ and $\frac{aa'}{ss'} = \frac{a''a'}{s''s'}$.

It is now verified immediately that these two operations satisfy the ring axioms — in fact, the required computations to show this are exactly the same as for ordinary fractions in $\mathbb{Q}$. □

Remark 6.3 (Explanation for the equivalence relation in Definition 6.2 (b))**.** Compared to Remark 6.1 (a) it might be surprising that the equivalence relation $(a,s) \sim (a',s')$ for $S^{-1}R$ is not just given by $as' - a's = 0$, but by $u(as' - a's) = 0$ for some $u \in S$. Again, there is both an algebraic and a geometric reason for this:

(a) *Algebraic reason:* Without the additional element $u \in S$ the relation $\sim$ would not be transitive. In fact, if we set $u = v = 1$ in the proof of transitivity in Lemma 6.2, we could only show that $s'(as'' - a''s) = 0$ in $(*)$. Of course, this does not imply $as'' - a''s = 0$ if S contains zero-divisors. (On the other hand, if S does not contain any zero-divisors, the condition $u(as' - a's) = 0$ for some $u \in S$ can obviously be simplified to $as' - a's = 0$.)

(b) *Geometric reason, continuing Remark 6.1 (b):* Let $X = V(xy)$ in $\mathbb{A}^2_{\mathbb{R}}$ be the union of the two coordinate axes as in the picture on the right, so that $A(X) = \mathbb{R}[x,y]/(xy)$. Then the functions y and 0 on X agree in a neighborhood of the point $a = (1,0)$, and so $\frac{y}{1}$ and $\frac{0}{1}$ should be the same local function at a, i. e. the same element in $S^{-1}R$ with $S = \{f \in A(X) : f(a) \neq 0\}$. But without the additional element $u \in S$ in Definition 6.2 (b) this would be false, since $y \cdot 1 - 0 \cdot 1 \neq 0 \in A(X)$. However, if we can set $u = x \in S$, then $x(y \cdot 1 - 0 \cdot 1) = xy = 0 \in A(X)$, and hence $\frac{y}{1} = \frac{0}{1} \in S^{-1}R$ as desired.



Remark 6.4. Let S be a multiplicatively closed subset of a ring R.

(a) There is an obvious ring homomorphism $\varphi : R \to S^{-1}R$, $a \mapsto \frac{a}{1}$ that makes $S^{-1}R$ into an R-algebra. However, φ is only injective if S does not contain zero-divisors, as $\frac{a}{1} = \frac{0}{1}$ by definition only implies the existence of an element $u \in S$ with $u(a \cdot 1 - 0 \cdot 1) = ua = 0$. We have already seen a concrete example of this in Remark 6.3 (b): in that case we had $y \neq 0 \in R$, but $\frac{y}{1} = \frac{0}{1} \in S^{-1}R$.

(b) In Definition 6.2 (a) of a multiplicatively closed subset we have not excluded the case $0 \in S$, i. e. that we "want to allow division by 0". However, in this case we trivially get $S^{-1}R = 0$, since we can then always take $u = 0$ in Definition 6.2 (b).

Example 6.5 (Standard examples of localization)**.** Let us now give some examples of multiplicatively closed sets in a ring R, and thus of localizations. In fact, the following are certainly the most important examples — probably any localization that you will meet is of one of these forms.

(a) Of course, $S = \{1\}$ is a multiplicatively closed subset, and leads to the localization $S^{-1}R \cong R$.

(b) Let S be the set of non-zero-divisors in R. Then S is multiplicatively closed: if $a, b \in S$ then ab is also not a zero-divisor, since $abc = 0$ for some $c \in R$ first implies $bc = 0$ since $a \in S$, and then $c = 0$ since $b \in S$.

Of particular importance is the case when R is an integral domain. Then $S = R \backslash \{0\}$, and every non-zero element $\frac{a}{s}$ is a unit in $S^{-1}R$, with inverse $\frac{s}{a}$. Hence $S^{-1}R$ is a field, the so-called **quotient field** $\operatorname{Quot} R$ of R.

Moreover, in this case every localization $T^{-1}R$ at a multiplicatively closed subset T with $0 \notin T$ is naturally a subring of $\operatorname{Quot} R$, since

$$T^{-1}R \to \operatorname{Quot} R, \quad \frac{a}{s} \mapsto \frac{a}{s}$$

is an injective ring homomorphism. In particular, by (a) the ring R itself can be thought of as a subring of its quotient field $\operatorname{Quot} R$ as well.

(c) For a fixed element $a \in R$ let $S = \{a^n : n \in \mathbb{N}\}$. Then S is obviously multiplicatively closed. The corresponding localization $S^{-1}R$ is often written as R_a; we call it the **localization** of R **at the element** a.

(d) Let $P \trianglelefteq R$ be a prime ideal. Then $S = R \backslash P$ is multiplicatively closed, since $a \notin P$ and $b \notin P$ implies $ab \notin P$ by Definition 2.1 (a). The resulting localization $S^{-1}R$ is usually denoted by R_P and called the **localization** of R **at the prime ideal** P.

In fact, this construction of localizing a ring at a prime ideal is probably the most important case of localization. In particular, if $R = A(X)$ is the ring of functions on a variety X and $P = I(a) = \{f \in A(X) : f(a) = 0\}$ the ideal of a point $a \in X$ (which is maximal and hence prime, see Remark 2.7), the localization R_P is exactly the *ring of local functions* on X at a as in Remark 6.1 (b). So localizing a coordinate ring at the maximal ideal corresponding to a point can be thought of as studying the variety locally around this point.

Example 6.6 (Localizations of $\mathbb{Z}$)**.** As concrete examples, let us apply the constructions of Example 6.5 to the ring $R = \mathbb{Z}$ of integers. Of course, localizing at $\mathbb{Z} \backslash \{0\}$ as in (b) gives the quotient field $\operatorname{Quot} \mathbb{Z} = \mathbb{Q}$ as in Remark 6.1 (a). If $p \in \mathbb{Z}$ is a prime number then localization at the element p as in (c) gives the ring

$$\mathbb{Z}_p = \left\{ \frac{a}{p^n} : a \in \mathbb{Z}, n \in \mathbb{N} \right\} \subset \mathbb{Q},$$

whereas localizing at the prime ideal (p) as in (d) yields

$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} : a, b \in \mathbb{Z}, p \nmid b \right\} \subset \mathbb{Q}$$

(these are both subrings of $\mathbb{Q}$ by Example 6.5 (b)). So note that we have to be careful about the notations, in particular since $\mathbb{Z}_p$ is usually also used to denote the quotient $\mathbb{Z}/p\mathbb{Z}$.

After having seen many examples of localizations, let us now turn to the study of their properties. We will start by relating ideals in a localization $S^{-1}R$ to ideals in R.

Proposition 6.7 (Ideals in localizations)**.** *Let S be a multiplicatively closed subset of a ring R. In the following, we will consider contractions and extensions by the ring homomorphism $\varphi : R \to S^{-1}R$.*

(a) *For any ideal $I \trianglelefteq R$ we have $I^e = \{\frac{a}{s} : a \in I, s \in S\}$.*

(b) *For any ideal $I \trianglelefteq S^{-1}R$ we have $(I^c)^e = I$.*

(c) *Contraction and extension by φ give a one-to-one correspondence*

$$\begin{array}{ccc} \{\text{prime ideals in } S^{-1}R\} & \overset{1:1}{\longleftrightarrow} & \{\text{prime ideals } I \text{ in } R \text{ with } I \cap S = \emptyset\} \\ I & \longmapsto & I^c \\ I^e & \longleftarrow\!\shortmid & I. \end{array}$$

Proof.

(a) As $\frac{1}{s} \in S^{-1}R$ for $s \in S$, the ideal I^e generated by $\varphi(I)$ must contain the elements $\frac{1}{s} \cdot \varphi(a) = \frac{a}{s}$ for $a \in I$. But it is easy to check that $\{\frac{a}{s} : a \in I, s \in S\}$ is already an ideal in $S^{-1}R$, and so it has to be equal to I^e.

(b) By Exercise 1.19 (b) it suffices to show the inclusion "$\supset$". So let $\frac{a}{s} \in I$. Then $a \in \varphi^{-1}(I) = I^c$ since $\varphi(a) = \frac{a}{1} = s \cdot \frac{a}{s} \in I$. By (a) applied to I^c it follows that $\frac{a}{s} \in (I^c)^e$.

(c) We have to check a couple of assertions.

The map "$\to$" is well-defined: If I is a prime ideal in $S^{-1}R$ then I^c is a prime ideal in R by Exercise 2.9 (b). Moreover, we must have $I^c \cap S = \emptyset$ as all elements of S are mapped to units by φ, and I cannot contain any units.

The map "$\leftarrow$" is well-defined: Let $I \trianglelefteq R$ be a prime ideal with $I \cap S = \emptyset$. We have to check that I^e is a prime ideal. So let $\frac{a}{s}, \frac{b}{t} \in S^{-1}R$ with $\frac{a}{s} \cdot \frac{b}{t} \in I^e$. By (a) this means that $\frac{ab}{st} = \frac{c}{u}$ for some $c \in I$ and $u \in S$, i. e. there is an element $v \in S$ with $v(abu - stc) = 0$. This implies that $uvab \in I$. Since I is prime, one of these four factors must lie in I. But u and v are in S and thus not in I. Hence we conclude that $a \in I$ or $b \in I$, and therefore $\frac{a}{s} \in I^e$ or $\frac{b}{t} \in I^e$ by (a).

$(I^c)^e = I$ for any prime ideal I in $S^{-1}R$: this follows from (b).

$(I^e)^c = I$ for any prime ideal I in R with $I \cap S = \emptyset$: By Exercise 1.19 (a) we only have to check the inclusion "$\subset$". So let $a \in (I^e)^c$, i. e. $\varphi(a) = \frac{a}{1} \in I^e$. By (a) this means $\frac{a}{1} = \frac{b}{s}$ for some $b \in I$ and $s \in S$, and thus there is an element $u \in S$ with $u(as - b) = 0$. Therefore $sua \in I$, which means that one of these factors is in I since I is prime. But s and u lie in S and hence not in I. So we conclude that $a \in I$. □

Example 6.8 (Prime ideals in R_P)**.** Applying Proposition 6.7 (c) to the case of a localization of a ring R at a prime ideal P, i. e. at the multiplicatively closed subset $R \setminus P$ as in Example 6.5 (d), gives a one-to-one correspondence by contraction and extension

$$\{\text{prime ideals in } R_P\} \overset{1:1}{\longleftrightarrow} \{\text{prime ideals } I \text{ in } R \text{ with } I \subset P\}. \qquad (*)$$

One should compare this to the case of prime ideals in the quotient ring R/P: by Lemma 1.21 and Corollary 2.4 we have a one-to-one correspondence

$$\{\text{prime ideals in } R/P\} \overset{1:1}{\longleftrightarrow} \{\text{prime ideals } I \text{ in } R \text{ with } I \supset P\},$$

this time by contraction and extension by the quotient map. Unfortunately, general ideals do not behave as nicely: whereas ideals in the quotient R/P still correspond to ideals in R containing P, the analogous statement for general ideals in the localization R_P would be false.

As we will see now, another consequence of the one-to-one correspondence $(*)$ is that the localization R_P has *exactly one* maximal ideal (of course it always has at least one by Corollary 2.17). Since this is a very important property of rings, it has a special name.

Definition 6.9 (Local rings)**.** A ring is called **local** if it has exactly one maximal ideal.

Corollary 6.10 (R_P is local)**.** *Let P be a prime ideal in a ring R. Then R_P is local with maximal ideal $P^e = \{\frac{a}{s} : a \in P, s \notin P\}$.*

Proof. Obviously, the set of all prime ideals of R contained in P has the upper bound P. So as the correspondence $(*)$ of Example 6.8 preserves inclusions, we see that every prime ideal in R_P is contained in P^e. In particular, every maximal ideal in R_P must be contained in P^e. But this means that P^e is the only maximal ideal. The formula for P^e is just Proposition 6.7 (a). □

Note however that arbitrary localizations as in Definition 6.2 are in general not local rings.

Remark 6.11 (Rings of local functions on a variety are local)**.** If $R = A(X)$ is the coordinate ring of a variety X and $P = I(a)$ the maximal ideal of a point $a \in X$, the localization R_P is the ring of local functions on X at a as in Example 6.5 (d). By Corollary 6.10, this is a local ring with unique maximal ideal P^e, i. e. with only the maximal ideal corresponding to the point a. This fits nicely with the interpretation of Remark 6.1 (b) that R_P describes an "arbitrarily small neighborhood" of a: although every function in R_P is defined on X in a neighborhood of a, there is no point except a at which every function in R_P is well-defined.

Exercise 6.12 (Universal property of localization)**.** Let S be a multiplicatively closed subset of a ring R. Prove that the localization $S^{-1}R$ satisfies the following *universal property*: for any homomorphism $\alpha : R \to R'$ to another ring R' that maps S to units in R' there is a unique ring homomorphism $\varphi : S^{-1}R \to R'$ such that $\varphi(\frac{r}{1}) = \alpha(r)$ for all $r \in R$.

Exercise 6.13.

(a) Let $R = \mathbb{R}[x,y]/(xy)$ and $P = (x-1) \trianglelefteq R$. Show that P is a maximal ideal of R, and that $R_P \cong \mathbb{R}[x]_{(x-1)}$. What does this mean geometrically?

(b) Let R be a ring and $a \in R$. Show that $R_a \cong R[x]/(ax-1)$, where R_a denotes the localization of R at a as in Example 6.5 (c). Can you give a geometric interpretation of this statement?

Exercise 6.14. Let S be a multiplicatively closed subset in a ring R, and let $I \trianglelefteq R$ be an ideal with $I \cap S = \emptyset$. Prove the following "localized version" of Corollary 2.17:

(a) The ideal I is contained in a prime ideal P of R such that $P \cap S = \emptyset$ and $S^{-1}P$ is a maximal ideal in $S^{-1}R$.

(b) The ideal I is not necessarily contained in a maximal ideal P of R with $P \cap S = \emptyset$.

Exercise 6.15 (Saturations)**.** For a multiplicatively closed subset S of a ring R we call

$$\overline{S} := \{s \in R : as \in S \text{ for some } a \in R\}$$

the *saturation* of S. Show that $\overline{S}$ is a multiplicatively closed subset as well, and that $\overline{S}^{-1}R \cong S^{-1}R$.

Exercise 6.16 (Alternative forms of Nakayama's Lemma for local rings)**.** Let M be a finitely generated module over a local ring R. Prove the following two versions of *Nakayama's Lemma* (see Corollary 3.27):

(a) If there is a proper ideal $I \trianglelefteq R$ with $IM = M$, then $M = 0$.

(b) If $I \trianglelefteq R$ is the unique maximal ideal and $m_1, \dots, m_k \in M$ are such that $\overline{m_1}, \dots, \overline{m_k}$ generate M/IM as an R/I-vector space, then $m_1, \dots, m_k$ generate M as an R-module.

So far we have only applied the technique of localization to rings. But in fact this works equally well for modules, so let us now give the corresponding definitions. However, as the computations to check that these definitions make sense are literally the same as in Lemma 6.2, we will not repeat them here.

Definition 6.17 (Localization of modules)**.** Let S be a multiplicatively closed subset of a ring R, and let M be an R-module. Then

$$(m,s) \sim (m',s') \;:\Leftrightarrow\; \text{there is an element } u \in S \text{ such that } u(s'm - sm') = 0$$

is an equivalence relation on $M \times S$. We denote the equivalence class of $(m,s) \in M \times S$ by $\frac{m}{s}$. The set of all equivalence classes

$$S^{-1}M := \left\{ \frac{m}{s} : m \in M, s \in S \right\}$$

is then called the **localization** of M at S. It is an $S^{-1}R$-module together with the addition and scalar multiplication

$$\frac{m}{s} + \frac{m'}{s'} := \frac{s'm + sm'}{ss'} \quad \text{and} \quad \frac{a}{s} \cdot \frac{m'}{s'} := \frac{am'}{ss'}$$

for all $a \in R$, $m, m' \in M$, and $s, s' \in S$.

In the cases of example 6.5 (c) and (d), when S is equal to $\{a^n : n \in \mathbb{N}\}$ for an element $a \in R$ or $R \backslash P$ for a prime ideal $P \trianglelefteq R$, we will write $S^{-1}M$ also as M_a or M_P, respectively.

Example 6.18. Let S be a multiplicatively closed subset of a ring R, and let I be an ideal in R. Then by Proposition 6.7 (a) the localization $S^{-1}I$ in the sense of Definition 6.17 is just the extension I^e of I with respect to the canonical map $R \to S^{-1}R$, $a \mapsto \frac{a}{1}$.

10

In fact, it turns out that the localization $S^{-1}M$ of a module M is just a special case of a tensor product. As one might already expect from Definition 6.17, it can be obtained from M by an extension of scalars from R to the localization $S^{-1}R$ as in Construction 5.16. Let us prove this first, so that we can then carry over some of the results of Chapter 5 to our present situation.

Lemma 6.19 (Localization as a tensor product)**.** *Let S be a multiplicatively closed subset in a ring R, and let M be an R-module. Then $S^{-1}M \cong M \otimes_R S^{-1}R$.*

Proof. There is a well-defined R-module homomorphism

$$\varphi : S^{-1}M \to M \otimes_R S^{-1}R, \quad \frac{m}{s} \mapsto m \otimes \frac{1}{s},$$

since for $m' \in M$ and $s' \in S$ with $\frac{m}{s} = \frac{m'}{s'}$, i. e. such that $u(s'm - sm') = 0$ for some $u \in S$, we have

$$m \otimes \frac{1}{s} - m' \otimes \frac{1}{s'} = u(s'm - sm') \otimes \frac{1}{uss'} = 0.$$

Similarly, by the universal property of the tensor product we get a well-defined homomorphism

$$\psi : M \otimes_R S^{-1}R \to S^{-1}M \quad \text{with} \quad \psi\left(m \otimes \frac{a}{s}\right) = \frac{am}{s},$$

as for $a' \in R$ and $s' \in S$ with $u(s'a - sa') = 0$ for some $u \in S$ we also get $u(s'am - sa'm) = 0$, and thus $\frac{am}{s} = \frac{a'm}{s'}$. By construction, φ and ψ are inverse to each other, and thus φ is an isomorphism. □

Remark 6.20 (Localization of homomorphisms)**.** Let $\varphi : M \to N$ be a homomorphism of R-modules, and let $S \subset R$ be a multiplicatively closed subset. Then by Construction 5.16 there is a well-defined homomorphism of $S^{-1}R$-modules

$$\varphi \otimes \mathrm{id} : M \otimes_R S^{-1}R \to N \otimes_R S^{-1}R \quad \text{such that} \quad (\varphi \otimes \mathrm{id})\left(m \otimes \frac{a}{s}\right) = \varphi(m) \otimes \frac{a}{s}.$$

By Lemma 6.19 this can now be considered as an $S^{-1}R$-module homomorphism

$$S^{-1}\varphi : S^{-1}M \to S^{-1}N \quad \text{with} \quad S^{-1}\varphi\left(\frac{m}{s}\right) = \frac{\varphi(m)}{s}.$$

We will call this the *localization $S^{-1}\varphi$ of the homomorphism φ*. As before, we will denote it by φ_a or φ_P if $S = \{a^n : n \in \mathbb{N}\}$ for an element $a \in R$ or $S = R \backslash P$ for a prime ideal $P \trianglelefteq R$, respectively.

The localization of rings and modules is a very well-behaved concept in the sense that it is compatible with almost any other construction that we have seen so far. One of the most important properties is that it preserves exact sequences, which then means that it is also compatible with everything that can be expressed in terms of such sequences.

Proposition 6.21 (Localization is exact)**.** *For every short exact sequence*

$$0 \longrightarrow M_1 \xrightarrow{\varphi} M_2 \xrightarrow{\psi} M_3 \longrightarrow 0$$

of R-modules, and any multiplicatively closed subset $S \subset R$, the localized sequence

$$0 \longrightarrow S^{-1}M_1 \xrightarrow{S^{-1}\varphi} S^{-1}M_2 \xrightarrow{S^{-1}\psi} S^{-1}M_3 \longrightarrow 0$$

is also exact.

Proof. By Lemma 6.19, localization at S is just the same as tensoring with $S^{-1}R$. But tensor products are right exact by Proposition 5.22 (b), hence it only remains to prove the injectivity of $S^{-1}\varphi$.

So let $\frac{m}{s} \in S^{-1}M_1$ with $S^{-1}\varphi(\frac{m}{s}) = \frac{\varphi(m)}{s} = 0$. This means that there is an element $u \in S$ with $u\varphi(m) = \varphi(um) = 0$. But φ is injective, and therefore $um = 0$. Hence $\frac{m}{s} = \frac{1}{us} \cdot um = 0$, i. e. $S^{-1}\varphi$ is injective. □

Corollary 6.22. *Let S be a multiplicatively closed subset of a ring R.*

(a) *For any homomorphism $\varphi : M \to N$ of R-modules we have* $\ker(S^{-1}\varphi) = S^{-1}\ker\varphi$ *and* $\operatorname{im}(S^{-1}\varphi) = S^{-1}\operatorname{im}\varphi$.

In particular, if φ is injective / surjective then $S^{-1}\varphi$ is injective / surjective, and if M is a submodule of N then $S^{-1}M$ is a submodule of $S^{-1}N$ in a natural way.

(b) *If M is a submodule of an R-module N then $S^{-1}(N/M) \cong S^{-1}N/S^{-1}M$.*

(c) *For any two R-modules M and N we have $S^{-1}(M \oplus N) \cong S^{-1}M \oplus S^{-1}N$.*

Proof.

(a) Localizing the exact sequence $0 \longrightarrow \ker\varphi \longrightarrow M \xrightarrow{\varphi} \operatorname{im}\varphi \longrightarrow 0$ of Example 4.3 (a) at S, we get by Proposition 6.21 an exact sequence

$$0 \longrightarrow S^{-1}\ker\varphi \longrightarrow S^{-1}M \xrightarrow{S^{-1}\varphi} S^{-1}\operatorname{im}\varphi \longrightarrow 0.$$

But again by Example 4.3 this means that the modules $S^{-1}\ker\varphi$ and $S^{-1}\operatorname{im}\varphi$ in this sequence are the kernel and image of the map $S^{-1}\varphi$, respectively.

(b) This time we localize the exact sequence $0 \longrightarrow M \longrightarrow N \longrightarrow N/M \longrightarrow 0$ of Example 4.3 (b) to obtain by Proposition 6.21

$$0 \longrightarrow S^{-1}M \longrightarrow S^{-1}N \longrightarrow S^{-1}(N/M) \to 0,$$

which by Example 4.3 means that the last module $S^{-1}(N/M)$ in this sequence is isomorphic to the quotient $S^{-1}N/S^{-1}M$ of the first two.

(c) This follows directly from the definition, or alternatively from Lemma 5.8 (c) as localization is the same as taking the tensor product with $S^{-1}R$. □

Remark 6.23. If an operation such as localization preserves short exact sequences as in Proposition 6.21, the same then also holds for longer exact sequences: in fact, we can split a long exact sequence of R-modules into short ones as in Remark 4.5, localize it to obtain corresponding short exact sequences of $S^{-1}R$-modules, and then glue them back to a localized long exact sequence by Lemma 4.4 (b).

Exercise 6.24. Let S be a multiplicatively closed subset of a ring R, and let M be an R-module. Prove:

(a) For $M_1, M_2 \leq M$ we have $S^{-1}(M_1 + M_2) = S^{-1}M_1 + S^{-1}M_2$.

(b) For $M_1, M_2 \leq M$ we have $S^{-1}(M_1 \cap M_2) = S^{-1}M_1 \cap S^{-1}M_2$. However, if $M_i \leq M$ for all i in an arbitrary index set J, it is in general not true that $S^{-1}\left(\bigcap_{i\in J} M_i\right) = \bigcap_{i\in J} S^{-1}M_i$. Does at least one inclusion hold?

(c) For an ideal $I \trianglelefteq R$ we have $S^{-1}\sqrt{I} = \sqrt{S^{-1}I}$.

Example 6.25. Let P and Q be maximal ideals in a ring R; we want to compute the ring R_Q/P_Q. Note that this is also an R_Q-module, and as such isomorphic to $(R/P)_Q$ by Corollary 6.22 (b). So when viewed as such a module, it does not matter whether we first localize at Q and then take the quotient by P or vice versa.

(a) If $P \neq Q$ we must also have $P \not\subset Q$ since P is maximal and $Q \neq R$. So there is an element $a \in P \backslash Q$, which leads to $\frac{1}{1} = \frac{a}{a} \in P_Q$. Hence in this case the ideal P_Q in R_Q is the whole ring, and we obtain $R_Q/P_Q = (R/P)_Q = 0$.

(b) On the other hand, if $P = Q$ there is a well-defined ring homomorphism

$$\varphi : R/P \to R_P/P_P, \ \overline{a} \mapsto \overline{\left(\frac{a}{1}\right)}.$$

We can easily construct an inverse of this map as follows: the morphism $R \to R/P$, $a \mapsto \overline{a}$ sends $R\backslash P$ to units since R/P is a field by Lemma 2.3 (b), hence it passes to a morphism

$$R_P \to R/P, \ \frac{a}{s} \mapsto \overline{s}^{-1}\overline{a}$$

by Exercise 6.12. But P_P is in the kernel of this map, and thus we get a well-defined ring homomorphism

$$R_P/P_P \to R/P, \ \overline{\left(\frac{a}{s}\right)} \mapsto \overline{s}^{-1}\overline{a},$$

which clearly is an inverse of φ. So the ring R_P/P_P is isomorphic to R/P.

Note that this result is also easy to understand from a geometric point of view: let R be the ring of functions on a variety X over an algebraically closed field, and let $p, q \in X$ be the points corresponding to the maximal ideals P and Q, respectively (see Remark 2.7 (a)). Then taking the quotient by $P = I(p)$ means restricting the functions to the point p by Lemma 0.9 (d), and localizing at Q corresponds to restricting them to an arbitrarily small neighborhood of q by Example 6.5 (d).

So first of all we see that the order of these two operations should not matter, since restrictions of functions commute. Moreover, if $p \neq q$ then our small neighborhood of q does not meet p, and we get the zero ring as the space of functions on the empty set. In contrast, if $p = q$ then after restricting the functions to the point p another restriction to a neighborhood of p does not change anything, and thus we obtain the isomorphism of (b) above.

As an application of this result, we get the following refinement of the notion of length of a module: not only is the length of all composition series the same, but also the maximal ideals occurring in the quotients of successive submodules in the composition series as in Definition 3.18 (a).

Corollary 6.26 (Length of localized modules)**.** *Let M be an R-module of finite length, and let*

$$0 = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_n = M$$

be a composition series for M, so that $M_i/M_{i-1} \cong R/I_i$ by Definition 3.18 (a) with some maximal ideals $I_i \trianglelefteq R$ for $i = 1, \ldots, n$.

Then for any maximal ideal $P \trianglelefteq R$ the localization M_P is of finite length as an R_P-module, and its length $l_{R_P}(M_P)$ is equal to the number of indices i such that $I_i = P$. In particular, up to permutations the ideals I_i (and their multiplicities) as above are the same for all choices of composition series.

Proof. Let $P \trianglelefteq R$ be a maximal ideal. From the given composition series we obtain a chain of localized submodules

$$0 = (M_0)_P \subset (M_1)_P \subset \cdots \subset (M_n)_P = M_P. \qquad (*)$$

For its successive quotients $(M_i)_P/(M_{i-1})_P$ for $i = 1,\dots,n$ we get by Corollary 6.22 (b)

$$(M_i)_P/(M_{i-1})_P \cong (M_i/M_{i-1})_P \cong (R/I_i)_P \cong R_P/(I_i)_P.$$

But by Example 6.25 this is 0 if $I_i \neq P$, whereas for $I_i = P$ it is equal to R_P modulo its unique maximal ideal P_P by Corollary 6.10. So by deleting all equal submodules from the chain $(*)$ we obtain a composition series for M_P over R_P whose length equals the number of indices $i \in \{1,\dots,n\}$ with $I_i = P$. □

We can rephrase Corollary 6.26 by saying that the length of a module is a "local concept": if an R-module M is of finite length and we know the lengths of the localized modules M_P for all maximal ideals $P \trianglelefteq R$, we also know the length of M: it is just the sum of all the lengths of the localized modules.

In a similar way, there are many other properties of an R-module M that carry over from the localized modules M_P, i. e. in geometric terms from considering M locally around small neighborhoods of every point. This should be compared to properties in geometric analysis such as continuity and differentiability: if we have e. g. a function $f : U \to \mathbb{R}$ on an open subset U of $\mathbb{R}$ we can define what it means that it is continuous at a point $p \in U$ — and to check this we only need to know f in an arbitrarily small neighborhood of p. In total, the function f is then continuous if and only if it is continuous at all points $p \in U$: we can say that "continuity is a local property". Here are some algebraic analogues of such local properties:

Proposition 6.27 (Local properties)**.** *Let R be a ring.*

(a) *If M is an R-module with $M_P = 0$ for all maximal ideals $P \trianglelefteq R$, then $M = 0$.*

(b) *Consider a sequence of R-module homomorphisms*

$$0 \longrightarrow M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} M_3 \longrightarrow 0. \qquad (*)$$

If the localized sequences of R_P-modules

$$0 \longrightarrow (M_1)_P \xrightarrow{(\varphi_1)_P} (M_2)_P \xrightarrow{(\varphi_2)_P} (M_3)_P \longrightarrow 0$$

are exact for all maximal ideals $P \trianglelefteq R$, then the original sequence $(*)$ is exact as well.

We say that being zero and being exact are "local properties" of a module and a sequence, respectively. Note that the converse of both statements holds as well by Proposition 6.21.

Proof.

(a) Assume that $M \neq 0$. If we then choose an element $m \neq 0$ in M, its annihilator $\operatorname{ann}(m)$ does not contain $1 \in R$, so it is a proper ideal of R and thus by Corollary 2.17 contained in a maximal ideal P. Hence for all $u \in R \setminus P$ we have $u \notin \operatorname{ann}(m)$, i. e. $um \neq 0$ in M. This means that $\frac{m}{1}$ is non-zero in M_P, and thus $M_P \neq 0$.

(b) It suffices to show that a three-term sequence $L \xrightarrow{\varphi} M \xrightarrow{\psi} N$ is exact if all localizations $L_P \xrightarrow{\varphi_P} M_P \xrightarrow{\psi_P} N_P$ are exact for maximal $P \trianglelefteq R$, since we can then apply this to all three possible positions in the five-term sequence of the statement.

To prove this, note first that

$$\begin{aligned}(\psi(\varphi(L)))_P &= \left\{ \frac{\psi(\varphi(m))}{s} : m \in L, s \in S \right\} = \left\{ \psi_P\left(\varphi_P\left(\frac{m}{s}\right)\right) : m \in L, s \in S \right\} \\ &= \psi_P(\varphi_P(L_P)) = 0,\end{aligned}$$

since the localized sequence is exact. This means by (a) that $\psi(\varphi(L)) = 0$, i. e. $\operatorname{im}\varphi \subset \ker\psi$. We can therefore form the quotient $\ker\psi/\operatorname{im}\varphi$, and get by the exactness of the localized sequence

$$(\ker\psi/\operatorname{im}\varphi)_P \cong (\ker\psi)_P/(\operatorname{im}\varphi)_P = \ker\psi_P/\operatorname{im}\varphi_P = 0$$

using Corollary 6.22. But this means $\ker\psi/\operatorname{im}\varphi = 0$ by (a), and so $\ker\psi = \operatorname{im}\varphi$, i. e. the sequence $L \xrightarrow{\varphi} M \xrightarrow{\psi} N$ is exact. □

Remark 6.28. As usual, Proposition 6.27 (b) means that all properties that can be expressed in terms of exact sequences are local as well: e. g. an R-module homomorphism $\varphi : M \to N$ is injective (resp. surjective) if and only if its localizations $\varphi_P : M_P \to N_P$ for all maximal ideals $P \trianglelefteq R$ are injective (resp. surjective). Moreover, as in Remark 6.23 it follows that Proposition 6.27 (b) holds for longer exact sequences as well.

Exercise 6.29. Let R be a ring. Show:

(a) Ideal containment is a local property: for two ideals $I, J \trianglelefteq R$ we have $I \subset J$ if and only if $I_P \subset J_P$ in R_P for all maximal ideals $P \trianglelefteq R$.

(b) Being reduced is a local property, i. e. R is reduced if and only if R_P is reduced for all maximal ideals $P \trianglelefteq R$.

(c) Being an integral domain is not a local property, i. e. R might not be an integral domain although the localizations R_P are integral domains for all maximal ideals $P \trianglelefteq R$. Can you give a geometric interpretation of this statement?

7. Chain Conditions

In the previous chapters we have met several finiteness conditions: an algebra can be finitely generated as in Definition 1.26 (b), a module can also be finitely generated as in Definition 3.3 (b) or have finite length as in Definition 3.18, and in our study of Zorn's Lemma in Remark 2.13 we discussed whether an increasing chain of ideals in a ring has to stop after finitely many steps. Of course, this can often make a difference: for example, a reduced algebra over a field is the coordinate ring of a variety if and only if it is finitely generated (see Remark 1.31), and the Cayley-Hamilton theorem in Proposition 3.25 together with its corollaries such as Nakayama's Lemma in Corollary 3.27 only hold for finitely generated modules. So we will now take some time to study such finiteness questions in more detail and see how they are related.

The key idea is to consider chains of submodules in a given module and check whether they have to stop after finitely many terms.

Definition 7.1 (Noetherian and Artinian modules)**.** Let M be an R-module.

(a) M is called **Noetherian** if every ascending chain

$$M_0 \subset M_1 \subset M_2 \subset \cdots$$

of submodules of M becomes stationary, i. e. if for every such chain there is an index $n \in \mathbb{N}$ such that $M_k = M_n$ for all $k \geq n$. Obviously, this is the same as saying that there is no infinite *strictly* ascending chain $M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$.

(b) Similarly, M is called **Artinian** if every descending chain

$$M_0 \supset M_1 \supset M_2 \supset \cdots$$

of submodules becomes stationary. Again, this is equivalent to requiring that there is no infinite strictly descending chain $M_0 \supsetneq M_1 \supsetneq M_2 \supsetneq \cdots$.

The conditions of (a) and (b) are often referred to as the *ascending* and *descending chain condition*, respectively. The ring R itself is called Noetherian or Artinian if it has this property as an R-module; the submodules above are then just ideals of R by Example 3.4 (a).

Example 7.2.

(a) Any field K is trivially Noetherian and Artinian as it has only the trivial ideals (0) and K. More generally, a K-vector space V is Noetherian if and only if it is Artinian if and only if it is finite-dimensional (i. e. finitely generated):

- If V is finite-dimensional, there can only be finite strictly ascending or descending chains of vector subspaces of V since the dimension has to be strictly increasing or decreasing in such a chain, respectively.
- If V is infinite-dimensional, we can obviously form a chain $M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$ with $\dim_K M_n = n$ for all $n \in \mathbb{N}$: set $M_0 = 0$, and $M_{n+1} = M_n + \langle v_{n+1} \rangle$ with $v_{n+1} \notin M_n$ for all $n \in \mathbb{N}$. Similarly, we can also find an infinite descending chain $M_0 \supsetneq M_1 \supsetneq M_2 \supsetneq \cdots$ of infinite-dimensional subspaces of V with $\dim_K(M/M_n) = n$ for all n: let $M_0 = M$, and $M_{n+1} = M_n \cap \ker \varphi_{n+1}$ for some linear map $\varphi_{n+1} : V \to K$ that is not identically zero on M_n. Then $M/M_n \cong (M/M_{n+1})/(M_n/M_{n+1})$ by Proposition 3.10 (b), and so $\dim M_n/M_{n+1} = 1$ implies $\dim M/M_n = n$ for all n by induction.

(b) The ring $\mathbb{Z}$ is Noetherian: if we had a strictly increasing chain of ideals $I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq \cdots$ in $\mathbb{Z}$, then certainly $I_1 \neq 0$, and thus $I_1 = (a)$ for some non-zero $a \in \mathbb{Z}$. But there are only finitely many ideals in $\mathbb{Z}$ that contain I_1 since they correspond to ideals of the finite ring $\mathbb{Z}/(a)$ by Lemma 1.21. Hence the chain cannot be infinitely long, and thus $\mathbb{Z}$ is Noetherian.

On the other hand, $\mathbb{Z}$ is not Artinian, since there is an infinite decreasing chain of ideals

$$\mathbb{Z} \supsetneq 2\mathbb{Z} \supsetneq 4\mathbb{Z} \supsetneq 8\mathbb{Z} \supsetneq \cdots.$$

(c) Let $R = \bigcup_{n\in\mathbb{N}} \mathbb{R}[x_0, x_1, \ldots, x_n]$ be the polynomial ring over $\mathbb{R}$ in infinitely many variables. Then R is neither Noetherian nor Artinian, since there are infinite chains of ideals

$$(x_0) \subsetneq (x_0, x_1) \subsetneq (x_0, x_1, x_2) \subsetneq \cdots \qquad \text{and} \qquad (x_0) \supsetneq (x_0^2) \supsetneq (x_0^3) \supsetneq \cdots.$$

Exercise 7.3. For a prime number $p \in \mathbb{N}$, consider $M = \mathbb{Z}_p/\mathbb{Z}$ as a $\mathbb{Z}$-module (i. e. as an Abelian group), where $\mathbb{Z}_p \subset \mathbb{Q}$ denotes the localization of $\mathbb{Z}$ at the element p as in Example 6.5 (c). Show that:

(a) Every proper submodule of M is of the form $\left\langle \overline{\frac{1}{p^n}} \right\rangle$.

(b) M is Artinian, but not Noetherian.

As you might expect, we will see in the following that Noetherian and Artinian modules have many similar properties that can be obtained from one another by just reversing all inclusions — as e. g. in the first two parts of the following lemma. However, there are also aspects in which the two concepts of Noetherian and Artinian modules are fundamentally different (in particular when we specialize to rings in the second half of this chapter). A first example of this is part (c) of the following equivalent reformulation of our chain conditions, which asserts that a module is Noetherian if and only if each of its submodules is finitely generated. There is no corresponding condition for Artinian modules, and in fact this is one of the main reasons why in practice Noetherian modules are much more important than Artinian ones.

Lemma 7.4 (Equivalent conditions for Noetherian and Artinian modules)**.** *Let M be an R-module.*

(a) *M is Noetherian if and only if every non-empty family of submodules of M has a maximal element.*

(b) *M is Artinian if and only if every non-empty family of submodules of M has a minimal element.*

(c) *M is Noetherian if and only if every submodule of M is finitely generated.*

Proof.

(a) "⇒" Let A be a non-empty family of submodules of M. If there was no maximal element of A, we could choose recursively a chain $M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$ from A, contradicting the assumption that M is Noetherian.

"⇐" Consider a chain $M_0 \subset M_1 \subset M_2 \subset \cdots$ of submodules of M. By assumption, the set $A = \{M_n : n \in \mathbb{N}\}$ has a maximal element M_n. But then $M_k = M_n$ for all $k \geq n$, hence M is Noetherian.

(b) is proven in the same way as (a), just reversing all inclusions. 11

(c) "⇒" Assume that we have a submodule $N \leq M$ that is not finitely generated. Then we can recursively pick $m_0 \in N$ and $m_{n+1} \in N \setminus \langle m_0, \ldots, m_n \rangle$ for $n \in \mathbb{N}$, and obtain a chain $M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$ in M. This is a contradiction since M is Noetherian.

"⇐" Let $M_0 \subset M_1 \subset M_2 \subset \cdots$ be a chain of submodules of M. Then $N = \bigcup_{n\in\mathbb{N}} M_n$ is also a submodule of M (which can be shown in the same way as in the proof of Corollary 2.17). So by assumption N can be generated by finitely many elements $m_1, \ldots, m_r \in N$. Now by definition of N we must have $m_i \in M_{n_i}$ for all $i = 1, \ldots, r$ and some $n_1, \ldots, n_r \in \mathbb{N}$. With $n = \max\{n_1, \ldots, n_r\}$ we then have $m_1, \ldots, m_r \in M_n$. Hence $N = \langle m_1, \ldots, m_r \rangle \leq M_n \leq N$, which implies $M_k = M_n = N$ for all $k \geq n$. □

Example 7.5.

(a) Any principal ideal domain R (as e. g. $R = \mathbb{Z}$ in Example 7.2 (b)) is Noetherian by Lemma 7.4 (c), since every ideal in R is even generated by one element.

(b) Note that the $\mathbb{R}$-algebra $\mathbb{R}[x]$ is a Noetherian ring by (a), but not a Noetherian $\mathbb{R}$-module by Example 7.2 (a) (as it is an infinite-dimensional $\mathbb{R}$-vector space). So when applying the chain conditions to algebras we have to be very careful whether we talk about Noetherian resp. Artinian *rings* or *modules*, as this might make a difference! There is one important case however in which there is no such difference:

(c) Let I be an ideal in a ring R, and let M be an R-module. Then M/IM is both an R/I-module and an R-module, and by definition a subset of M/IM is an R/I-submodule if and only if it is an R-submodule. So we conclude by Definition 7.1 that M/IM is Noetherian resp. Artinian as an R/I-module if and only if it has this property as an R-module.

In particular, applying this result to $M = R$ we see that the R-algebra R/I is Noetherian resp. Artinian as a ring if and only if it has this property as an R-module.

Remark 7.6 (Maximal ideals in Noetherian rings)**.** Let I be an ideal in a Noetherian ring R with $I \neq R$. Then every ascending chain of ideals in R becomes stationary, so by Remark 2.13 this means that the existence of a maximal ideal in R that contains I is trivial and does not require Zorn's Lemma (for which we had to work quite a bit). In fact, this is just the statement of Lemma 7.4 (a) which tells us that the family of all proper ideals of R containing I must have a maximal element.

Let us now prove some basic properties of Noetherian and Artinian modules.

Lemma 7.7. *Let N be a submodule of an R-module M.*

(a) *M is Noetherian if and only if N and M/N are Noetherian.*

(b) *M is Artinian if and only if N and M/N are Artinian.*

Proof. We just prove (a), since (b) can be proven in the same way, reversing all inclusions.

"$\Rightarrow$" Let M be Noetherian. First of all, any chain $N_0 \subset N_1 \subset N_2 \subset \cdots$ of submodules of N is also a chain of submodules of M, and thus becomes stationary. Hence N is Noetherian.

Similarly, let $P_0 \subset P_1 \subset P_2 \subset \cdots$ be a chain of submodules of M/N. If we set $M_k = q^{-1}(P_k)$ for all $k \in \mathbb{N}$, where $q : M \to M/N$ is the quotient map, then $M_0 \subset M_1 \subset M_2 \subset \cdots$ is a chain of submodules of M. As M is Noetherian, we have $M_k = M_n$ for all $k \geq n$ with some $n \in \mathbb{N}$. But since q is surjective we then also have $P_k = q(M_k) = q(M_n) = P_n$ for all $k \geq n$. Hence M/N is Noetherian.

"$\Leftarrow$" Let $M_0 \subset M_1 \subset M_2 \subset \cdots$ be an ascending chain of submodules in M. If we set $N_k := M_k \cap N$ and $P_k := (M_k + N)/N$ for all $k \in \mathbb{N}$, then

$$N_0 \subset N_1 \subset N_2 \subset \cdots \qquad \text{and} \qquad P_0 \subset P_1 \subset P_2 \subset \cdots$$

are chains of submodules of N and M/N, respectively. By assumption, both of them become stationary, and hence there is an element $n \in \mathbb{N}$ such that $N_k = N_n$ and $P_k = P_n$ for all $k \geq n$. But then we obtain a commutative diagram for all $k \geq n$

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N_n & \longrightarrow & M_n & \longrightarrow & P_n & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & N_k & \longrightarrow & M_k & \longrightarrow & P_k & \longrightarrow & 0 \end{array}$$

whose rows are exact by Proposition 3.10 (c) and Example 4.3 (b), and whose columns are induced by the inclusions $M_n \to M_k$. As the left and right vertical map are isomorphisms, so is the middle one by Corollary 4.12, and thus we have $M_k = M_n$ for $k \geq n$ as well. Hence M is Noetherian. □

Remark 7.8.

(a) Of course, by Example 4.3 we can rephrase Lemma 7.7 by saying that for any short exact sequence $0 \longrightarrow N \longrightarrow M \longrightarrow P \longrightarrow 0$ of R-modules the middle entry M is Noetherian (or Artinian) if and only if N and P are Noetherian (or Artinian).

(b) Let I be an ideal in a Noetherian (resp. Artinian) ring R. Combining Lemma 7.7 with Example 7.5 (c) we see that then the quotient R/I is a Noetherian (resp. Artinian) ring as well.

Corollary 7.9. *Let M and N be R-modules.*

(a) *The direct sum $M \oplus N$ is Noetherian if and only if M and N are Noetherian.*

(b) *If R is Noetherian and M is finitely generated, then M is also Noetherian.*

The same statements also hold with "Noetherian" replaced by "Artinian".

Proof. Again, we only show the statement for Noetherian modules, since the Artinian counterpart follows in exactly the same way.

(a) By Remark 7.8 (a), this follows from the exact sequence $0 \longrightarrow M \longrightarrow M \oplus N \longrightarrow N \longrightarrow 0$.

(b) Let $M = \langle m_1, \dots, m_k \rangle$ for some $m_1, \dots, m_k \in M$. Then the ring homomorphism

$$\varphi : R^k \to M, \ \ (a_1, \dots, a_k) \mapsto a_1 m_1 + \dots + a_k m_k$$

is surjective, so that we have an exact sequence $0 \longrightarrow \ker \varphi \longrightarrow R^k \xrightarrow{\varphi} M \longrightarrow 0$. Now as R is Noetherian, so is R^k by (a), and hence also M by Remark 7.8 (a). □

Remark 7.10 (Structure Theorem for finitely generated Abelian groups)**.** Let M be a finitely generated Abelian group, viewed as a finitely generated $\mathbb{Z}$-module as in Example 3.2 (d). Of course, M is then a Noetherian $\mathbb{Z}$-module by Corollary 7.9 (b). But we can follow the idea of the proof of this statement one step further: the $\mathbb{Z}$-module $\ker \varphi$ occurring in the proof is (again by Remark 7.8 (a)) finitely generated as well, and so we also find a surjective ring homomorphism $\psi : \mathbb{Z}^l \to \ker \varphi$ for some $l \in \mathbb{N}$. This results in another short exact sequence $0 \longrightarrow \ker \psi \longrightarrow \mathbb{Z}^l \xrightarrow{\psi} \ker \varphi \longrightarrow 0$, and thus by gluing as in Lemma 4.4 (b) in the exact sequence

$$0 \longrightarrow \ker \psi \longrightarrow \mathbb{Z}^l \xrightarrow{\psi} \mathbb{Z}^k \xrightarrow{\varphi} M \longrightarrow 0,$$

which means that $M \cong \mathbb{Z}^k / \ker \varphi = \mathbb{Z}^k / \operatorname{im} \psi$.

Now $\psi \in \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}^l, \mathbb{Z}^k)$ is just given by an integer $k \times l$ matrix by Remark 3.17 (b). Similarly to the case of matrices over a field [G2, Proposition 16.43] one can show that it is possible to change bases in $\mathbb{Z}^l$ and $\mathbb{Z}^k$ such that the matrix of ψ has non-zero entries only on the diagonal. But this means that $\operatorname{im} \psi$ is generated by $a_1 e_1, \dots, a_k e_k$ for some $a_1, \dots, a_k \in \mathbb{Z}$, where $e_1, \dots, e_k$ denotes the standard basis of $\mathbb{Z}^k$. Thus

$$M \ \cong \ \mathbb{Z}^k / \operatorname{im} \psi \ \cong \ \mathbb{Z}/a_1\mathbb{Z} \times \dots \times \mathbb{Z}/a_k\mathbb{Z},$$

and so we conclude that every finitely generated Abelian group is a product of cyclic groups. Of course, by the Chinese Remainder Theorem [G1, Proposition 11.22] this can also be rewritten as

$$M \ \cong \ \mathbb{Z}^r \times \mathbb{Z}/q_1\mathbb{Z} \times \dots \times \mathbb{Z}/q_n\mathbb{Z}$$

for $r, n \in \mathbb{N}$ and (not necessarily distinct) prime powers $q_1, \dots, q_n$.

Let us now see how the finite length condition on modules is related to the concepts introduced in this chapter.

Lemma 7.11. *An R-module M is of finite length if and only if it is both Noetherian and Artinian.*

Proof. If M is of finite length, then all strict chains of submodules of M are finite by Exercise 3.19 (b) and (c). So in this case M is clearly both Noetherian and Artinian.

Conversely, assume that M is both Noetherian and Artinian. Starting from $M_0 = 0$, we try to construct a chain $M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots$ of submodules of M as follows: for $n \in \mathbb{N}$ let M_{n+1} be a minimal submodule of M that strictly contains M_n — as long as $M_n \neq M$ this works by Lemma 7.4 (b) since M is Artinian. But as M is Noetherian as well, we cannot get such an infinite ascending chain of submodules, and thus we conclude that we must have $M_n = M$ for some $n \in \mathbb{N}$. The resulting chain $0 = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_n = M$ is then a composition series for M, since by construction there are no submodules between M_{i-1} and M_i for all $i = 1, \dots, n$. □

Exercise 7.12. Let M be an R-module, and let $\varphi : M \to M$ be an R-module homomorphism. If M is Noetherian (hence finitely generated) and φ is surjective, you already know by Corollary 3.28 that φ has to be an isomorphism.

Now show that if M is Artinian and φ is injective, then φ is again an isomorphism.

(Hint: Consider the images of φ^n for $n \in \mathbb{N}$.)

So far we have mostly considered chain conditions for general modules. For the rest of this chapter we now want to specialize to the case of rings. In this case we can obtain stronger results, however we will also see that this is where the Noetherian and Artinian conditions begin to diverge drastically. So let us consider these two conditions in turn, starting with the more important case of Noetherian rings.

The one central result on Noetherian rings is Hilbert's Basis Theorem, which implies that "most rings that you will meet in practice are Noetherian".

Proposition 7.13 (Hilbert's Basis Theorem). *If R is a Noetherian ring, then so is the polynomial ring $R[x]$.*

Proof. Assume that $R[x]$ is not Noetherian. Then by Lemma 7.4 (c) there is an ideal $I \trianglelefteq R[x]$ that is not finitely generated. We can therefore pick elements $f_0, f_1, f_2, \ldots \in I$ as follows: let $f_0 \in I$ be a non-zero polynomial of minimal degree, and for $k \in \mathbb{N}$ let f_{k+1} be a polynomial of minimal degree in $I \setminus \langle f_0, \ldots, f_k \rangle$.

Now for all $k \in \mathbb{N}$ let $d_k \in \mathbb{N}$ be the degree and $a_k \in R$ the leading coefficient of f_k, so that we can write $f_k = a_k x^{d_k} + (\text{lower order terms})$. Note that $d_k \leq d_{k+1}$ for all k by construction of the polynomials. Moreover, since R is Noetherian the chain of ideals $(a_0) \subset (a_0, a_1) \subset (a_0, a_1, a_2) \subset \cdots$ becomes stationary, and thus we must have $a_{n+1} = c_0 a_0 + \cdots + c_n a_n$ for some $n \in \mathbb{N}$ and $c_0, \ldots, c_n \in R$. We can therefore cancel the leading term in f_{n+1} by subtracting a suitable linear combination of $f_0, \ldots, f_n$: in the polynomial

$$f'_{n+1} := f_{n+1} - \sum_{k=0}^{n} c_k x^{d_{n+1}-d_k} f_k$$

the $x^{d_{n+1}}$-coefficient is $a_{n+1} - c_0 a_0 - \cdots - c_n a_n = 0$. But this means that $\deg f'_{n+1} < \deg f_{n+1}$, and as $f_{n+1} \notin \langle f_0, \ldots, f_n \rangle$ we must have $f'_{n+1} \notin \langle f_0, \ldots, f_n \rangle$ as well. This contradicts our choice of f_{n+1}, proving that an ideal I as above cannot exist, and thus that $R[x]$ is Noetherian. □

Corollary 7.14. *Any finitely generated algebra over a Noetherian ring is itself a Noetherian ring.*

Proof. Let R be a Noetherian ring. By Lemma 1.30, any finitely generated R-algebra is of the form $R[x_1, \ldots, x_n]/I$ for an ideal I in the polynomial ring $R[x_1, \ldots, x_n]$. Hilbert's Basis Theorem now implies by induction that the polynomial ring $R[x_1, \ldots, x_n] = R[x_1][x_2] \cdots [x_n]$ is Noetherian. So by Remark 7.8 (b) the quotient $R[x_1, \ldots, x_n]/I$ is a Noetherian ring as well. □

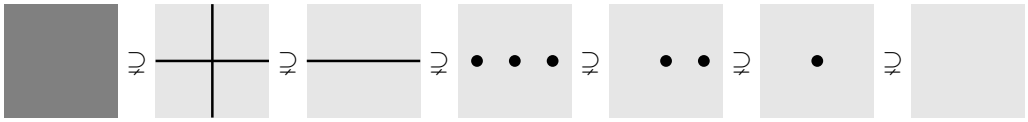
Remark 7.15 (Geometric interpretation of Noetherian and Artinian rings)**.** We have seen in Remark 1.31 that any coordinate ring $A(X)$ of a variety X over a field K is a finitely generated K-algebra. So by Corollary 7.14 we see that $A(X)$ is always a Noetherian ring. In particular, by Lemma 7.4 (c) this means that every ideal in $A(X)$ is finitely generated, and hence that any subvariety of X can be defined by *finitely many* polynomial equations.

It is also instructive to study the original chain conditions of Definition 7.1 in a geometric setting. As the correspondence of Remark 1.10 between (radical) ideals in $A(X)$ and subvarieties of X reverses inclusions, the *ascending* chain condition on ideals for the Noetherian ring $A(X)$ translates into a *descending* chain condition on subvarieties of X, i. e. every chain

$$X_0 \supset X_1 \supset X_2 \supset \cdots$$

of subvarieties of X must become stationary. The geometric picture behind this is the following: to make a subvariety smaller one has to drop an irreducible component or to reduce the dimension of the subvariety (a concept that we will introduce in Chapter 11), and this process must always

terminate since the number of components and the dimension are natural numbers. For example, in $X = \mathbb{A}^2_{\mathbb{R}}$ we could have the following descending chain of subvarieties:



Of course, we can easily construct finite descending chains of subvarieties of $\mathbb{A}^2_{\mathbb{R}}$ of any length in the same way, but infinite chains are impossible.

In contrast, as soon as X contains infinitely many points $a_1, a_2, a_3, \ldots$, it is easy to construct an infinite strictly ascending chain of subvarieties $X_0 \subsetneq X_1 \subsetneq X_2 \subsetneq$ of X by setting $X_n = \{a_k : k \leq n\}$ for all $n \in \mathbb{N}$. As this corresponds to a strictly decreasing chain of ideals in $A(X)$, we expect that a coordinate ring $A(X)$ is Artinian if and only if X is a finite collection of points — so that the Artinian condition is a very strong one, in contrast to the Noetherian one.

12

To turn these expectations into rigorous statements, let us now study Artinian rings in detail and prove some algebraic results that all correspond to the geometric idea that an Artinian ring R should describe a finite union of points X. More precisely, consider the correspondence of subvarieties of X and ideals of R as in Remark 2.7: as the only irreducible subvarieties of X are single points, we would expect that any prime ideal of R is already maximal. Let us prove this now, together with the fact that in an Artinian ring the zero ideal is always a product of maximal ideals — which can also be translated into geometry by Remark 1.12 by saying that X is a union of finitely many points.

Proposition 7.16. *Let R be an Artinian ring.*

(a) *There are (not necessarily distinct) maximal ideals $P_1, \ldots, P_n \trianglelefteq R$ such that $P_1 \cdot \cdots \cdot P_n = 0$.*

(b) *R has only finitely many prime ideals, all of them are maximal, and occur among the $P_1, \ldots, P_n$ in (a).*

Proof.

(a) Let $I = P_1 \cdot \cdots \cdot P_n$ be a product of maximal ideals $P_1, \ldots, P_n$ such that I is minimal among all ideals that can be written in this form — such a minimal element exists by Lemma 7.4 (b) since R is Artinian. We need to show that $I = 0$. First we note:

(i) I^2 is obviously also a product of maximal ideals, and we have $I^2 \subset I$. Hence $I^2 = I$ by the minimality of I.

(ii) If $P \trianglelefteq R$ is any maximal ideal, then PI is also a product of maximal ideals with $PI \subset I$. So again by the minimality of I we see that $I = PI$, which is clearly contained in P. Hence I is contained in every maximal ideal of R.

Now let us assume for a contradiction that $I \neq 0$. Then, as R is Artinian, Lemma 7.4 (b) implies that there is a minimal ideal $J \trianglelefteq R$ with $IJ \neq 0$. About this ideal we note:

- J is a principal ideal (so in particular finitely generated): there must be an element $b \in J$ with $I \cdot (b) \neq 0$, and we clearly have $(b) \subset J$, so $(b) = J$ by the minimality of J.
- $IJ = J$ again by the minimality of J, since $IJ \subset J$ and $I \cdot IJ = I^2 J = IJ \neq 0$ by (i).

Because of these two properties of J Nakayama's Lemma in Corollary 3.27 gives us an element $a \in I$ with $(1-a)J = 0$. As $J \neq 0$, this means that $1-a$ is not a unit in R. But then $(1-a) \neq R$, hence $1-a$ is contained in a maximal ideal $P \trianglelefteq R$. But so is $a \in I \subset P$ by (ii), and thus we obtain the contradiction $1 = (1-a) + a \in P$. Hence we conclude that $I = 0$.

(b) Let $P \trianglelefteq R$ be any prime ideal. Then $P_i \subset P$ for some $i = 1, \ldots, n$, since otherwise there would be elements $a_i \in P_i \setminus P$ for all i, which implies $a_1 \cdot \cdots \cdot a_n \in P_1 \cdot \cdots \cdot P_n = 0 \subset P$ although no a_i lies in P. But P_i is maximal, and so we must have $P = P_i$. □

A somewhat surprising consequence of this proposition is that every Artinian ring is Noetherian — a statement that is false for general modules as we have seen in Exercise 7.3:

Proposition 7.17 (Hopkins). *For any ring R we have:*

$$R \text{ is Artinian} \quad \Leftrightarrow \quad R \text{ is Noetherian and every prime ideal of } R \text{ is maximal.}$$

Proof.

"$\Rightarrow$" Let R be Artinian. Then every prime ideal is maximal by Proposition 7.16 (b). Moreover, by Proposition 7.16 (a) there are maximal ideals $P_1, \dots, P_n$ of R giving a chain

$$0 = Q_0 \subset Q_1 \subset \cdots \subset Q_n = R$$

of ideals in R, where $Q_i = P_{i+1} \cdot P_{i+2} \cdot \cdots \cdot P_n$ for $i = 0, \dots, n$. Now for all $i = 1, \dots, n$ the quotient $Q_i/Q_{i-1} = Q_i/P_iQ_i$ is an Artinian R-module by Lemma 7.7 (b), hence an Artinian R/P_i-vector space by Example 7.5 (c), therefore also a Noetherian R/P_i-vector space by Example 7.2 (a), and thus a Noetherian R-module again by Example 7.5 (c). So by induction on i it follows from Lemma 7.7 (a) that Q_i is a Noetherian R-module for all $i = 0, \dots, n$. In particular, $R = Q_n$ is Noetherian.

"$\Leftarrow$" Assume that R is Noetherian, but not Artinian. We have to find a prime ideal P of R that is not maximal.

Consider the family of all ideals I of R such that R/I is not Artinian (as a ring or as an R-module, see Example 7.5 (c)). This family is non-empty since it contains the zero ideal, so as R is Noetherian it must have a maximal element P by Lemma 7.4 (a). Note that P is certainly *not* a maximal ideal: otherwise R/P would be a field by Lemma 2.3 (b), hence Artinian by Example 7.2 (a) — in contradiction to the choice of P.

It therefore suffices to prove that P is prime, i. e. by Lemma 2.3 (a) that $S := R/P$ is an integral domain. For any $a \in R$ consider the exact sequence

$$0 \longrightarrow S/\operatorname{ann}(\overline{a}) \xrightarrow{\cdot \overline{a}} S \longrightarrow S/(\overline{a}) \longrightarrow 0$$

of S-modules. As S is not Artinian, we know by Remark 7.8 (a) that at least one of the rings $S/\operatorname{ann}(\overline{a})$ and $S/(\overline{a})$ cannot be Artinian either. But since P was chosen to be maximal such that $S = R/P$ is not Artinian, taking a further quotient of S by a non-zero ideal must yield an Artinian ring, and thus we conclude that $\operatorname{ann}(\overline{a}) = 0$ or $\overline{a} = 0$. In other words, every non-zero element of R/P is a non-zero-divisor, i. e. R/P is an integral domain. □

Example 7.18. Let I be a non-zero ideal in a principal ideal domain R. Then as in Example 1.4 we have $I = (a)$ with $a = p_1^{a_1} \cdot \cdots \cdot p_n^{a_n}$, where $a_1, \dots, a_n \in \mathbb{N}_{>0}$ and $p_1, \dots, p_n$ are distinct prime elements of R. We want to check the conditions of Proposition 7.17 for the ring $S := R/I$.

First of all, by Lemma 1.21 the ideals of S correspond to the ideals of R that contain I. These are the ideals (b) with $b \mid a$, i. e. such that $b = p_1^{b_1} \cdot \cdots \cdot p_n^{b_n}$ with $b_i \leq a_i$ for all i. In particular, S has only finitely many ideals, which means by Definition 7.1 that S is trivially Noetherian as well as Artinian.

Moreover, since R is a principal ideal domain we have already seen in Example 2.6 (b) that every non-zero prime ideal in R is maximal, and hence by Corollary 2.4 that every prime ideal in S is maximal as well. Hence all three conditions of Proposition 7.17 are satisfied for S. In fact, the maximal ideals of S are just the ideals $(\overline{p_i})$ for $i = 1, \dots, n$. So the equation $(\overline{p_1})^{a_1} \cdot \cdots \cdot (\overline{p_n})^{a_n} = 0$ also verifies the statement of Proposition 7.16 (a).

Example 7.19. Specializing Example 7.18 to the geometric case $R = \mathbb{C}[x]$, we see that

$$S = \mathbb{C}[x]/(f) \quad \text{with} \quad f = (x - x_1)^{a_1} \cdot \cdots \cdot (x - x_n)^{a_n}$$

is an Artinian ring, where $x_1, \dots, x_n \in \mathbb{C}$ are distinct and $a_1, \dots, a_n \in \mathbb{N}_{>0}$. In fact, $X = V(f) = \{x_1, \dots, x_n\} \subset \mathbb{C}$ is a finite collection of points, in accordance with Remark 7.15. But S is not reduced unless $a_1 = \cdots = a_n = 1$ since the class of $(x - x_1) \cdots (x - x_n)$ is nilpotent in S, and consequently S is not the coordinate ring of X. Instead, the ring S remembers the local multiplicity information at each point, i. e. the exponents $a_1, \dots, a_n$ in f.

So even if S corresponds to a finite collection of points, the structure of S is not completely determined by giving these points: it also contains some local information at each of these points. The corresponding precise algebraic statement is that an Artinian ring is completely determined by its localizations at all maximal ideals:

Proposition 7.20 (Structure Theorem for Artinian rings)**.** *Every Artinian ring R is a finite product of local Artinian rings.*

More precisely, if $P_1,\dots,P_n$ are the distinct maximal ideals of R (see Proposition 7.16 (b)), then the localizations R_{P_i} are also Artinian for all $i=1,\dots,n$, and the ring R is isomorphic to $R_{P_1}\times\cdots\times R_{P_n}$.

Proof. By Proposition 7.16 we can find $k\in\mathbb{N}$ such that $P_1^k\cdot\ \cdots\ \cdot P_n^k=0$. Note that $P_1^k,\dots,P_n^k$ are pairwise coprime by Exercise 2.24. Hence $P_1^k\cap\cdots\cap P_n^k=P_1^k\cdot\ \cdots\ \cdot P_n^k=0$ by Exercise 1.8, and so the Chinese Remainder Theorem of Proposition 1.14 implies that

$$R\ \cong\ R/P_1^k\times\cdots\times R/P_n^k.$$

As the factors R/P_i^k are clearly Artinian by Lemma 7.7 (b), it therefore only remains to be shown that the ring R/P_i^k is isomorphic to the localization R_{P_i} for all i. In fact, it is straightforward to construct mutually inverse ring homomorphisms, without loss of generality for $i=1$:

- The ring homomorphism $R\to R_{P_1}$, $a\mapsto\frac{a}{1}$ contains P_1^k in its kernel: if $a\in P_1^k$ we can choose $a_j\in P_j\backslash P_1$ for all $j=2,\dots,n$. Then $u:=a_2^k\cdot\ \cdots\ \cdot a_n^k\notin P_1$ since P_1 is prime, and $ua=a\cdot a_2^k\cdot\ \cdots\ \cdot a_n^k\in P_1^k\cdot\ \cdots\ \cdot P_n^k=0$. This means that $\frac{a}{1}=0$ in R_{P_1}. Hence the above map gives rise to a ring homomorphism $R/P_1^k\to R_{P_1}$, $\overline{a}\mapsto\frac{a}{1}$.
- Now consider the ring homomorphism $R\to R/P_1^k$, $a\mapsto\overline{a}$. It maps any $a\in R\backslash P_1$ to a unit: otherwise $(\overline{a})$ would be contained in a maximal ideal of R/P_1^k, which must be of the form P/P_1^k for a maximal ideal $P\supset P_1^k$ of R by Lemma 1.21 and Corollary 2.4. As P is prime this means that $P\supset P_1$, and hence $P=P_1$ since P_1 is maximal. But $\overline{a}\in P_1/P_1^k$ implies $a\in P_1$, a contradiction. By Exercise 6.12 we conclude that the above map gives us a ring homomorphism $R_{P_1}\to R/P_1^k$ with $\frac{a}{1}\mapsto\overline{a}$. □

Example 7.21. Let us continue Example 7.18, i. e. consider the ring $S=R/I$ for a principal ideal domain R and $I=(a)$ with $a=p_1^{a_1}\cdot\ \cdots\ \cdot p_n^{a_n}$. In the proof of Proposition 7.20 we can then take any $k\in\mathbb{N}$ with $k\geq a_i$ for all $i=1,\dots,n$, and obtain

$$S/(\overline{p_i})^k\cong R/(a,p_i^k)=R/(p_i^{a_i})$$

by Example 1.4 (a). So the decomposition of S into local Artinian rings given by Proposition 7.20 is just

$$S\ \cong\ R/(p_1^{a_1})\times\cdots\times R/(p_n^{a_n}),$$

which by the proposition is also isomorphic to $S_{(p_1)}\times\cdots\times S_{(p_n)}$.

Exercise 7.22. Let R be a Noetherian ring. Show:

(a) If R is an integral domain, every non-zero non-unit $a\in R$ can be written as a product of irreducible elements of R.

(b) For any ideal $I\trianglelefteq R$ there is an $n\in\mathbb{N}$ such that $(\sqrt{I})^n\subset I$.

Exercise 7.23. Let S be a multiplicatively closed subset of a ring R. If R is Noetherian (resp. Artinian), show that the localization $S^{-1}R$ is also Noetherian (resp. Artinian).

Exercise 7.24. Prove for any R-module M:

(a) If M is Noetherian then $R/\operatorname{ann}M$ is Noetherian as well.

(b) If M is finitely generated and Artinian, then M is also Noetherian.

(Hint: You can reduce this to the statement of Proposition 7.17 that an Artinian ring is Noetherian.)

8. Prime Factorization and Primary Decompositions

When it comes to actual computations, Euclidean domains (or more generally principal ideal domains) are probably the "nicest" rings that are not fields. One of the main reasons for this is that their elements admit a unique prime factorization [G1, Proposition 11.9]. This allows for an easy computation of many concepts in commutative algebra, such as e. g. the operations on ideals in Example 1.4.

Unfortunately however, such rings are rather rare in practice, with the integers $\mathbb{Z}$ and the polynomial ring $K[x]$ over a field K being the most prominent examples. So we now want to study in this chapter if there are more general rings that allow a prime factorization of their elements, and what we can use as a substitute in rings that do not admit such a factorization.

More precisely, given an element $a \neq 0$ in an integral domain R which is not a unit we ask if we can write $a = p_1 \cdot \dots \cdot p_n$ for some $n \in \mathbb{N}_{>0}$ and $p_1, \dots, p_n \in R$ such that:

- The p_i for $i = 1, \dots, n$ are *irreducible* or *prime* — recall that in a principal ideal domain these two notions are equivalent, but in a general integral domain we only know that every prime element is irreducible [G1, Lemma 11.3 and Proposition 11.5].
- The decomposition is unique up to permutation and multiplication with units, i. e. if we also have $a = q_1 \cdot \dots \cdot q_m$ with $q_1, \dots, q_m$ irreducible resp. prime, then $m = n$ and there are units $c_1, \dots, c_n \in R$ and a permutation $\sigma \in S_n$ such that $q_i = c_i p_{\sigma(i)}$ for all $i = 1, \dots, n$.

Let us first discuss the precise relation between the different variants of these conditions.

Proposition and Definition 8.1 (Unique factorization domains)**.** *For an integral domain R the following statements are equivalent:*

(a) *Every non-zero non-unit of R is a product of prime elements.*

(b) *Every non-zero non-unit of R is a product of irreducible elements, and this decomposition is unique up to permutation and multiplication with units.*

(c) *Every non-zero non-unit of R is a product of irreducible elements, and every irreducible element is prime.*

If these conditions hold, R is called ***factorial*** *or a* ***unique factorization domain*** *(short:* ***UFD****).*

13

Proof.

(a) $\Rightarrow$ (b): Let $a \in R$ be a non-zero non-unit. By assumption we know that $a = p_1 \cdot \dots \cdot p_n$ for some prime elements $p_1, \dots, p_n$. As prime elements are irreducible [G1, Lemma 11.3], we therefore also have a decomposition into irreducible elements.

Moreover, let $a = q_1 \cdot \dots \cdot q_m$ be another decomposition into irreducible elements. Then p_1 divides $a = q_1 \cdot \dots \cdot q_m$, and as p_1 is prime this means that p_1 divides one of these factors, without loss of generality $p_1 \mid q_1$. Hence $q_1 = c\, p_1$ for some $c \in R$. But q_1 is irreducible and p_1 is not a unit, so c must be a unit. This means that p_1 and q_1 agree up to multiplication with a unit. Canceling p_1 in the equation $p_1 \cdot \dots \cdot p_n = q_1 \cdot \dots \cdot q_m$ by p_1 now yields $p_2 \cdot \dots \cdot p_n = c \cdot q_2 \cdot \dots \cdot q_m$, and continuing with this equation in the same way for $p_2, \dots, p_n$ gives the desired uniqueness statement.

(b) $\Rightarrow$ (c): Let $p \in R$ be irreducible, we have to show that p is prime. So let $p \mid ab$, i. e. $ab = pc$ for some $c \in R$. By assumption we can write all these four elements as products of irreducible elements and thus obtain an equation

$$a_1 \cdot \dots \cdot a_n \cdot b_1 \cdot \dots \cdot b_m = p \cdot c_1 \cdot \dots \cdot c_r.$$

But by the uniqueness assumption, the factor p on the right must up to a unit be one of the $a_1,\dots,a_n$ or $b_1,\dots,b_m$, which implies that $p \mid a$ or $p \mid b$.

(c) $\Rightarrow$ (a) is trivial. □

Remark 8.2. In Proposition 8.1, the assumption in (b) and (c) that every non-zero non-unit can be written as a product of irreducible elements is a very weak one: it is satisfied e. g. in every Noetherian domain by Exercise 7.22 (a). The other conditions are much stronger, as we will see in Examples 8.3 (b) and 8.7.

Example 8.3. The following two examples are already known from the "Algebraic Structures" class:

(a) As mentioned above, principal ideal domains (so in particular $\mathbb{Z}$ and univariate polynomial rings over a field) are unique factorization domains [G1, Proposition 11.9].

(b) In the ring $R = \mathbb{Z}[\sqrt{5}i] \subset \mathbb{C}$ the element 2 obviously divides $6 = (1+\sqrt{5}i)(1-\sqrt{5}i)$, but neither $1+\sqrt{5}i$ nor $1-\sqrt{5}i$. Hence 2 is not prime. But one can show that 2 is irreducible in R, and thus R is not a unique factorization domain [G1, Example 11.4]. In fact, $2\cdot 3 = (1+\sqrt{5}i)(1-\sqrt{5}i)$ are two decompositions of the same number 6 that do not agree up to permutation and units.

It follows by (a) that R cannot be a principal ideal domain. One can also check this directly: the ideal $(2, 1+\sqrt{5}i)$ is not principal [G1, Exercise 10.37]. In fact, we will see in Example 13.28 that up to multiplication with a constant this is the only non-principal ideal in R.

We will see in Example 13.8 however that R admits a "unique prime factorization" for ideals (instead of for elements) — a property that holds more generally in so-called Dedekind domains that we will study in Chapter 13.

Remark 8.4. The most important feature of the unique factorization property is that it is preserved when passing from a domain R to the polynomial ring $R[x]$. Often this is already shown in the "Introduction to Algebra" class, and so we will only sketch the proof of this statement here. It relies on the well-known *Lemma of Gauß* stating that an irreducible polynomial over $\mathbb{Z}$ (or more generally over a unique factorization domain R) remains irreducible when considered as a polynomial over $\mathbb{Q}$ (resp. the quotient field $K = \operatorname{Quot} R$). More precisely, if $f \in R[x]$ is reducible in $K[x]$ and factors as $f = gh$ with non-constant $g,h \in K[x]$, then there is an element $c \in K\setminus\{0\}$ such that cg and $\frac{h}{c}$ lie in $R[x]$, and so $f = (cg)\cdot\frac{h}{c}$ is already reducible in $R[x]$ [G3, Proposition 3.2 and Remark 3.3].

Proposition 8.5. *If R is a unique factorization domain, then so is $R[x]$.*

Proof sketch. We will check condition (c) of Definition 8.1 for $R[x]$.

Let $f \in R[x]$, and let c be a greatest common divisor of all coefficients of f. Then $f = cf'$ for some polynomial $f' \in R[x]$ with coefficients whose greatest common divisor is 1, so that no constant polynomial which is not a unit can divide f'. Now split f' into factors until all of them are irreducible — this process has to stop for degree reasons as we have just seen that the degree of each factor must be at least 1. But c can also be written as a product of irreducible elements since R is a unique factorization domain, and so $f = cf'$ is a product of irreducible elements as well.

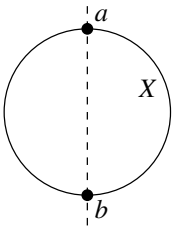
Next assume that f is irreducible, in particular we may assume that $c = 1$. We have to show that f is also prime. So let f divide gh for some $g,h \in R[x]$. If we denote by K the quotient field of R, then f is also irreducible in $K[x]$ by Remark 8.4, hence prime in $K[x]$ by Example 8.3 (a), and so without loss of generality $f \mid g$ in $K[x]$. This means that $g = fk$ for some $k \in K[x]$. But now by Remark 8.4 we can find $\frac{a}{b} \in K$ (with a and b coprime) such that $\frac{a}{b}f$ and $\frac{b}{a}k$ are in $R[x]$. Since the greatest common divisor of the coefficients of f is 1 this is only possible if b is a unit. But then $k = ab^{-1}(\frac{b}{a}k) \in R[x]$, and so $f \mid g$ in $R[x]$. □

Remark 8.6. Of course, Proposition 8.5 implies by induction that $R[x_1,\dots,x_n] = R[x_1][x_2]\cdots[x_n]$ is a unique factorization domain if R is. In particular, the polynomial ring $K[x_1,\dots,x_n]$ over a field K is a unique factorization domain. This also shows that there are more unique factorization domains

than principal ideal domains: as the ideal $(x_1,\dots,x_n)$ in $K[x_1,\dots,x_n]$ cannot be generated by fewer than n elements by Exercise 1.9, this polynomial ring is a principal ideal domain only for $n = 1$.

Example 8.7 (Geometric interpretation of prime and irreducible elements)**.** Consider the coordinate ring $R = \mathbb{R}[x,y]/(x^2+y^2-1)$ of the unit circle $X = V(x^2+y^2-1) \subset \mathbb{A}^2_{\mathbb{R}}$. Note that x^2+y^2-1 is obviously irreducible (it cannot be written as a product of two linear polynomials since otherwise X would have to be a union of two lines), and hence prime by Proposition 8.1 since $\mathbb{R}[x,y]$ is a unique factorization domain by Remark 8.6. So (x^2+y^2-1) is a prime ideal by Example 2.6 (a), and consequently R is an integral domain by Lemma 2.3 (a).

We are going to show that R is not a unique factorization domain. In fact, we will prove — and interpret geometrically — that $\bar{x} \in R$ is irreducible, but not prime. Note that the zero locus $V(\bar{x})$ of $\bar{x}$ in X consists of the two points $a = (0,1)$ and $b = (0,-1)$ shown in the picture on the right. In particular, $\bar{x}$ is neither 0 in R (otherwise $V(\bar{x})$ would be X) nor a unit (otherwise $V(\bar{x})$ would be empty).



(a) $\bar{x}$ is not prime: Geometrically, by Remark 2.7 (b) this is just the statement that $V(\bar{x})$ is not an irreducible variety since it consists of two points.

Algebraically, $\bar{x}$ divides $\bar{x}^2 = (1+\bar{y})(1-\bar{y})$ in R, but it does not divide $1 \pm \bar{y}$: if e. g. we had $\bar{x} \mid 1+\bar{y}$ this would mean $1+y = gx+h(x^2+y^2-1)$ for some $g,h \in \mathbb{R}[x,y]$, but plugging in the point a would then give the contradiction $2 = 0$.

(b) $\bar{x}$ is irreducible: otherwise we would have $\bar{x} = \bar{f}\,\bar{g}$ for two non-units $\bar{f}$ and $\bar{g}$ in R.

Intuitively, as the function $\bar{x}$ vanishes on X exactly at the two points a and b with multiplicity 1, this would mean that one of the two factors, say $\bar{f}$, would have to vanish exactly at a with multiplicity 1, and the other $\bar{g}$ exactly at b. But this would mean that the curve $V(f)$ in $\mathbb{A}^2_{\mathbb{R}}$ meets the circle X exactly at one point a with multiplicity 1. This seems geometrically impossible since the circle X has an outside and an inside, so if $V(f)$ crosses the circle and goes from the outside to the inside, it has to cross it again somewhere (as the dashed line in the picture above) since it cannot end in the interior of the circle.

To give an exact argument for this requires a bit more work. Note that every element $\bar{h} \in R$ has a unique representative of the form $h_0 + xh_1 \in \mathbb{R}[x,y]$ with $h_0, h_1 \in \mathbb{R}[y]$. We define a "norm" map

$$N : R \to \mathbb{R}[y],\ \ \bar{h} \mapsto h_0^2 + (y^2-1)h_1^2$$

which can also be thought of as taking the unique representative of $h(x,y) \cdot h(-x,y)$ not containing x. In particular, N is multiplicative (which can of course also be checked directly). Hence we have

$$(y+1)(y-1) = N(\bar{x}) = N(\bar{f})N(\bar{g}).$$

As $\mathbb{R}[y]$ is a unique factorization domain, there are now two possibilities (up to symmetry in $\bar{f}$ and $\bar{g}$):

- $N(\bar{f})$ is constant: Then $f_0^2 + (y^2-1)f_1^2$ is constant. But the leading coefficients of both f_0^2 and $(y^2-1)f_1^2$ are non-negative and thus cannot cancel in the sum, and hence we must have that f_0 is constant and $f_1 = 0$. But then $\bar{f}$ is a unit in R, which we excluded.
- $N(\bar{f}) = a(y-1)$ for some $a \in \mathbb{R}\backslash\{0\}$: Then $f_0^2 + (y^2-1)f_1^2 = a(y-1)$, and so we have $y-1 \mid f_0$. So we can write $f_0 = (y-1)f_0'$ for some polynomial $f_0' \in \mathbb{R}[y]$ and obtain $(y-1)f_0'^2 + (y+1)f_1^2 = a$. This is again a contradiction, since the left hand side must have a positive non-constant leading term.

Altogether, this contradiction shows that $\bar{x}$ is in fact irreducible.

Exercise 8.8. In contrast to Example 8.7, show that the following rings are unique factorization domains:

(a) the coordinate rings $\mathbb{R}[x,y]/(y-x^2)$ and $\mathbb{R}[x,y]/(xy-1)$ of the standard parabola and hyperbola in $\mathbb{A}^2_{\mathbb{R}}$, respectively;

(b) $\mathbb{C}[x,y]/(x^2+y^2-1)$.

We will also see in Proposition 12.14 "(e) $\Rightarrow$ (c)" that the localization of $\mathbb{R}[x,y]/(x^2+y^2-1)$ at the maximal ideal corresponding to one of the points a and b in Example 8.7 is a unique factorization domain. For the moment however we just note that the unique factorization property easily breaks down, and in addition is only defined for integral domains — so let us now study how the concept of prime factorization can be generalized to a bigger class of rings.

To see the idea how this can be done, let us consider Example 8.7 again. The function $\bar{x} \in R$ was not prime since its zero locus $V(\bar{x})$ was a union of two points. We could not decompose $\bar{x}$ as a product of two functions vanishing at only one of the points each, but we can certainly decompose the *ideal* $(\bar{x})$ into two maximal (and hence prime) ideals as

$$(\bar{x}) = I(a) \cap I(b) = (\bar{x}, \bar{y}-1) \cap (\bar{x}, \bar{y}+1),$$

which by Remark 1.12 is just the algebraic version of saying that $V(\bar{x})$ is the union of the two points a and b. So instead of elements we should rather decompose ideals of R, in the sense that we write them as intersections of "easier" ideals. (Note that in the above example we could also have taken the product of the two ideals instead of the intersection, but for general rings intersections turn out to be better-behaved, in particular under the presence of zero-divisors. Decompositions into products of maximal or prime ideals will be studied in Chapter 13, see e. g. Proposition 13.7 (b) and Example 13.12.)

To see what these "easier" ideals should be, consider the simple case of a principal ideal domain R: any non-zero ideal $I \trianglelefteq R$ can be written as $I = (p_1^{k_1} \cdot \cdots \cdot p_n^{k_n})$ for some distinct prime elements $p_1, \ldots, p_n \in R$ and $k_1, \ldots, k_n \in \mathbb{N}_{>0}$ by Example 8.3 (a), and the "best possible" decomposition of this ideal as an intersection of easier ideals is

$$I = (p_1)^{k_1} \cap \cdots \cap (p_n)^{k_n}.$$

So it seems that we are looking for decompositions of ideals as intersections of powers of prime ideals. Actually, whereas this is the correct notion for principal ideal domains, we need a slight variant of prime powers for the case of general rings:

Definition 8.9 (Primary ideals)**.** Let R be a ring. An ideal $Q \trianglelefteq R$ with $Q \neq R$ is called **primary** if for all $a, b \in R$ with $ab \in Q$ we have $a \in Q$ or $b^n \in Q$ for some $n \in \mathbb{N}$ (which is obviously equivalent to $a \in Q$ or $b \in \sqrt{Q}$).

Example 8.10 (Primary ideals = powers of prime ideals in principal ideal domains)**.** In a principal ideal domain R, the primary ideals are in fact exactly the ideals of the form (p^n) for a prime element $p \in R$ and $n \in \mathbb{N}_{>0}$:

- The ideal (p^n) is primary: if $ab \in (p^n)$ then $ab = cp^n$ for some $c \in R$. But now the n factors of p are either all contained in a (in which case $a \in (p^n)$), or at least one of them is contained in b (in which case $b^n \in (p^n)$).
- Conversely, let $I = (p_1^{k_1} \cdot \cdots \cdot p_n^{k_n})$ be any primary ideal, where $p_1, \ldots, p_n$ are distinct primes and $k_1, \ldots, k_n \in \mathbb{N}_{>0}$. Then we must have $n = 1$, since otherwise $p_1^{k_1} \cdot (p_2^{k_2} \cdot \cdots \cdot p_n^{k_n}) \in I$, but neither $p_1^{k_1}$ nor any power of $p_2^{k_2} \cdot \cdots \cdot p_n^{k_n}$ are in I.

Note that if R is only a unique factorization domain the same argument still works for principal ideals — but not for arbitrary ideals as we will see in Example 8.13 (a).

Remark 8.11. Let R be a ring.

(a) Obviously, every prime ideal in R is primary, but the converse does not hold by Example 8.10.

(b) However, if $Q \trianglelefteq R$ is primary then $P = \sqrt{Q}$ is prime: if $ab \in P$ then $(ab)^n \in Q$ for some $n \in \mathbb{N}$. Hence $a^n \in Q$ or $b^n \in \sqrt{Q}$, which means that a or b lie in $\sqrt{Q} = P$. In fact, P is then the smallest prime ideal containing Q by Lemma 2.21.

If we want to specify the underlying prime ideal $P = \sqrt{Q}$ of a primary ideal Q we often say that Q is **P-primary**.

(c) The condition of an ideal $Q \trianglelefteq R$ with $Q \neq R$ being primary can also be expressed in terms of the quotient ring R/Q: obviously, Definition 8.9 translates into the requirement that $\overline{a}\,\overline{b} = 0$ implies $\overline{a} = 0$ or $\overline{b}^n = 0$ for some $n \in \mathbb{N}$. This means for every element $\overline{b}$ that $\overline{b}^n = 0$ for some $n \in \mathbb{N}$ if there is an exists $\overline{a} \neq 0$ with $\overline{a}\,\overline{b} = 0$. So Q is primary if and only if every zero-divisor of R/Q is nilpotent.

As in Corollary 2.4 this means that primary ideals are preserved under taking quotients, i. e. for an ideal $I \subset Q$ we have that Q is primary in R if and only if Q/I is primary in R/I.

To obtain more examples of primary ideals, we need the following lemma. It gives a particularly easy criterion to detect a primary ideal Q if its underlying prime ideal $\sqrt{Q}$ is maximal.

Lemma 8.12 (Primary ideals over maximal ideals)**.** *Let P be a maximal ideal in a ring R. If an ideal $Q \trianglelefteq R$ satisfies one of the following conditions:*

(a) $\sqrt{Q} = P$;

(b) $P^n \subset Q \subset P$ *for some* $n \in \mathbb{N}_{>0}$;

then Q is P-primary.

Proof.

(a) The given condition means that in R/Q the nilradical $\sqrt{(0)}$ is equal to P/Q, hence maximal by Corollary 2.4. Exercise 2.25 then implies that every element of R/Q is either a unit or nilpotent. Therefore every zero-divisor of R/Q (which is never a unit) is nilpotent, and so Q is primary by Remark 8.11 (c).

(b) Taking radicals in the given inclusions yields $\sqrt{P} = \sqrt{P^n} \subset \sqrt{Q} \subset \sqrt{P}$, and hence we get $\sqrt{Q} = \sqrt{P} = P$. So the result then follows from (a). □

Example 8.13 (Primary ideals $\neq$ powers of prime ideals)**.** In general, primary ideals and powers of prime ideals are different objects:

(a) Primary ideals need not be powers of prime ideals: let $Q = (x^2, y)$ and $P = (x, y)$ in $\mathbb{R}[x, y]$. Then $\sqrt{Q} = P$ is maximal by Example 2.6 (c), and so Q is P-primary by Lemma 8.12 (a). However, $(x^2, xy, y^2) = P^2 \subsetneq Q \subsetneq P = (x, y)$, hence Q is not a power of P.

(b) Powers of prime ideals need not be primary: Let $R = \mathbb{R}[x, y, z]/(xy - z^2)$ and $P = (\overline{x}, \overline{z}) \trianglelefteq R$. Then P is prime by Lemma 2.3 (a) since $R/P \cong \mathbb{R}[y]$ is an integral domain. But the power $P^2 = (\overline{x}^2, \overline{x}\,\overline{z}, \overline{z}^2)$ is not primary as $\overline{x}\,\overline{y} = \overline{z}^2 \in P^2$, but neither is $\overline{x}$ in P^2 nor any power of $\overline{y}$.

However, if R is Noetherian and Q primary with $\sqrt{Q} = P$, then Q always *contains* a power of the prime ideal P by Exercise 7.22 (b).

Remark 8.14. Note that the condition of Definition 8.9 for a primary ideal Q is not symmetric in the two factors a and b, i. e. it does not say that $ab \in Q$ implies that one of the two factors a and b lie in $\sqrt{Q}$. In fact, Example 8.13 (b) shows that this latter condition is not equivalent to Q being primary as it is always satisfied by powers of prime ideals: if P is prime and $ab \in P^n$ for some $n \in \mathbb{N}_{>0}$ then we also have $ab \in P$, hence $a \in P$ or $b \in P$, and so a or b lie in $P = \sqrt{P} = \sqrt{P^n}$.

Let us now prove that every ideal in a Noetherian ring can be decomposed as an intersection of primary ideals.

Definition 8.15 (Primary decompositions)**.** Let I be an ideal in a ring R. A **primary decomposition** of I is a finite collection $\{Q_1, \dots, Q_n\}$ of primary ideals such that $I = Q_1 \cap \dots \cap Q_n$.

Proposition 8.16 (Existence of primary decompositions)**.** *In a Noetherian ring every ideal has a primary decomposition.*

Proof. Assume for a contradiction that R is a Noetherian ring that has an ideal without primary decomposition. By Lemma 7.4 (a) there is then an ideal $I \trianglelefteq R$ which is maximal among all ideals in R without a primary decomposition. In the quotient ring $S := R/I$ the zero ideal I/I is then the only one without a primary decomposition, since by Remark 8.11 (c) contraction and extension by the quotient map give a one-to-one correspondence between primary decompositions of an ideal $J \supset I$ in R and primary decompositions of J/I in R/I.

In particular, the zero ideal $(0) \trianglelefteq S$ is not primary itself, and so there are $a, b \in S$ with $ab = 0$, but $a \neq 0$ and $b^n \neq 0$ for all $n \in \mathbb{N}$. Now as R is Noetherian, so is S by Remark 7.8 (b), and hence the chain of ideals

$$\operatorname{ann}(b) \subset \operatorname{ann}(b^2) \subset \operatorname{ann}(b^3) \subset \cdots$$

becomes stationary, i. e. there is an $n \in \mathbb{N}$ such that $\operatorname{ann}(b^n) = \operatorname{ann}(b^{n+1})$.

Note that $(a) \neq 0$ and $(b^n) \neq 0$ by our choice of a and b. In particular, these two ideals have a primary decomposition. Taking the primary ideals of these two decompositions together, we then obtain a primary decomposition of $(a) \cap (b^n)$ as well. But $(a) \cap (b^n) = 0$: if $x \in (a) \cap (b^n)$ then $x = ca$ and $x = db^n$ for some $c, d \in S$. As $ab = 0$ we then have $0 = cab = xb = db^{n+1}$, hence $d \in \operatorname{ann}(b^{n+1}) = \operatorname{ann}(b^n)$, which means that $x = db^n = 0$. This is a contradiction, since the zero ideal in S does not have a primary decomposition by assumption. □

14

Example 8.17.

(a) In a unique factorization domain R every principal ideal $I = (p_1^{k_1} \cdot \cdots \cdot p_n^{k_n})$ has a primary decomposition

$$I = (p_1)^{k_1} \cap \cdots \cap (p_n)^{k_n}$$

by Example 8.10 (where $p_1, \ldots, p_n$ are distinct prime elements and $k_1, \ldots, k_n \in \mathbb{N}_{>0}$).

(b) Geometrically, if $I = Q_1 \cap \cdots \cap Q_n$ is a primary decomposition of an ideal I in the coordinate ring of a variety X, we have

$$V(I) = V(Q_1) \cup \cdots \cup V(Q_n) = V(P_1) \cup \cdots \cup V(P_n)$$

by Remark 1.12, where $P_i = \sqrt{Q_i}$ for $i = 1, \ldots, n$. So by Remark 2.7 we have decomposed the subvariety $Y = V(I)$ as a union of irreducible subvarieties $V(P_i)$. As coordinate rings of varieties are always Noetherian by Remark 7.15, Proposition 8.16 asserts that such a decomposition of a (sub-)variety into finitely many irreducible subvarieties is always possible.

However, the primary decomposition of an ideal $I \trianglelefteq A(X)$ contains more information that is not captured in its zero locus $V(I)$ alone: we do not only get subvarieties whose union is $V(I)$, but also (primary) ideals whose zero loci are these subvarieties. These primary ideals can be thought of as containing additional "multiplicity information": for example, the zero locus of the ideal $I = ((x-a_1)^{k_1} \cdot \cdots \cdot (x-a_n)^{k_n})$ in $\mathbb{R}[x]$ is the subset $\{a_1, \ldots, a_n\}$ of $\mathbb{R}$ — but the ideal also associates to each point a_i a multiplicity k_i, and the primary decomposition

$$I = ((x-a_1)^{k_1}) \cap \cdots \cap ((x-a_n)^{k_n})$$

as in (a) remembers these multiplicities.

In fact, in higher dimensions the additional information at each subvariety is more complicated than just a multiplicity. We will not study this here in detail, however we will see an instance of this in Example 8.23.

Having proven that primary decompositions always exist in Noetherian rings, we now want to see in the rest of this chapter to what extent these decompositions are unique. However, with our current definitions it is quite obvious that they are far from being unique, since they can be changed in two simple ways:

Example 8.18 (Non-uniqueness of primary decompositions)**.**

(a) We can always add "superfluous ideals" to a primary decomposition, i. e. primary ideals that are already contained in the intersection of the others. For example, (x^2) and $(x) \cap (x^2)$ in $\mathbb{R}[x]$ are two primary decompositions of the same ideal (x^2) by Example 8.10.

(b) In a given primary decomposition we might have several primary ideals with the same underlying radical, such as in

$$(x^2, xy, y^2) = (x^2, y) \cap (x, y^2) \qquad (*)$$

in $\mathbb{R}[x,y]$. Note that this equation holds since the ideals (x^2,y), (x,y^2), and (x^2,xy,y^2) contain exactly the polynomials without the monomials 1 and x, 1 and y, and without any constant or linear terms, respectively. Moreover, all three ideals have the radical $P = (x,y)$, and hence they are all P-primary by Lemma 8.12 (a). So $(*)$ are two different primary decompositions of the same ideal, in which none of the ideals is superfluous as in (a).

By a slight refinement of the definitions it is actually easy to remove these two ambiguities from primary decompositions. To do this, we need a lemma first.

Lemma 8.19 (Intersections of primary ideals)**.** *Let P be a prime ideal in a ring R. If Q_1 and Q_2 are two P-primary ideals in R, then $Q_1 \cap Q_2$ is P-primary as well.*

Proof. First of all, by Lemma 1.7 (b) we have $\sqrt{Q_1 \cap Q_2} = \sqrt{Q_1} \cap \sqrt{Q_2} = P \cap P = P$. Now let $ab \in Q_1 \cap Q_2$, i. e. $ab \in Q_1$ and $ab \in Q_2$. As Q_1 and Q_2 are P-primary we know that $a \in Q_1$ or $b \in P$, as well as $a \in Q_2$ or $b \in P$. This is the same as $a \in Q_1 \cap Q_2$ or $b \in P = \sqrt{Q_1 \cap Q_2}$. Hence $Q_1 \cap Q_2$ is P-primary. □

Definition 8.20 (Minimal primary decompositions)**.** Let $\{Q_1, \dots, Q_n\}$ be a primary decomposition of an ideal I in a ring R, and let $P_i = \sqrt{Q_i}$ for $i = 1, \dots, n$. Then the decomposition is called **minimal** if

(a) none of the ideals is superfluous in the intersection, i. e. $\bigcap_{j \neq i} Q_j \not\subset Q_i$ for all i;

(b) $P_i \neq P_j$ for all i, j with $i \neq j$.

Corollary 8.21 (Existence of minimal primary decompositions)**.** *If an ideal in a ring has a primary decomposition, it also has a minimal one.*

In particular, in a Noetherian ring every ideal has a minimal primary decomposition.

Proof. Starting from any primary decomposition, leave out superfluous ideals, and replace ideals with the same radical by their intersection, which is again primary with the same radical by Lemma 8.19.

The additional statement follows in combination with Proposition 8.16. □

Exercise 8.22. Find a minimal primary decomposition of ...

(a) the ideal $I = (\bar{x}^2)$ in the ring $R = \mathbb{R}[x,y]/(x^2+y^2-1)$ (see Example 8.7);

(b) the ideal $I = (6)$ in the ring $R = \mathbb{Z}[\sqrt{5}\,i]$ (see Example 8.3 (b));

(c) the ideal $I = (x,y) \cdot (y,z)$ in the ring $\mathbb{R}[x,y,z]$.

As a consequence of Corollary 8.21, one is usually only interested in minimal primary decompositions. However, even then the decompositions will in general not be unique, as the following example shows.

Example 8.23 (Non-uniqueness of minimal primary decompositions)**.** Let us consider the ideal $I = (y) \cdot (x,y) = (xy, y^2)$ in $\mathbb{R}[x,y]$. Geometrically, the zero locus of this ideal is just the horizontal axis $V(I) = V(y)$, which is already irreducible. However, I is not primary since $yx \in I$, but $y \notin I$ and

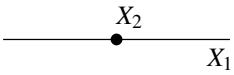
$x^n \notin I$ for all $n \in \mathbb{N}$. Hence, I is not its own minimal primary decomposition. However, we claim that

$$I = Q_1 \cap Q_2 = (y) \cap (x^2, xy, y^2) \quad \text{and} \quad I = Q_1 \cap Q_2' = (y) \cap (x, y^2)$$

are two different minimal primary decompositions of I. In fact, both equations can be checked in the same way as in Example 8.18 (b) (the ideal I contains exactly the polynomials with no monomial 1, y, or x^n for $n \in \mathbb{N}$). Moreover, Q_1 is primary since it is prime, and Q_2 and Q_2' are primary by Lemma 8.12 (a) since both of them have the same maximal radical $P_2 = (x, y)$. Finally, it is clear that in both decompositions none of the ideals is superfluous, and that the radicals of the two ideals are different — namely $P_1 = (y)$ with zero locus $X_1 = V(P_1) = \mathbb{R} \times \{0\}$ and $P_2 = (x, y)$ with zero locus $X_2 = V(P_2) = \{(0,0)\}$, respectively.

So geometrically, even our minimal primary decompositions contain a so-called *embedded* component, i. e. a subvariety X_2 contained in another subvariety X_1 of the decomposition, so that it is not visible in the zero locus of I. The other component X_1 is usually called an *isolated* component. The corresponding algebraic statement is that P_2 contains another prime ideal P_1 occurring as a radical in the decomposition; we will also say that P_2 is an embedded and P_1 an isolated prime ideal (see Definition 8.25 and Example 8.28).

The intuitive reason why this embedded component occurs is that X_2 has a higher "multiplicity" in I than X_1 (in a sense that we do not want to make precise here). We can indicate this as in the picture on the right by a "fat point" X_2 on a "thin line" X_1.



In any case, we conclude that even minimal primary decompositions of an ideal are not unique. However, this non-uniqueness is very subtle: we will now show that it can only occur in the primary ideals of embedded components. More precisely, we will prove that in a minimal primary decomposition:

(a) the underlying prime ideals of all primary ideals are uniquely determined (see Proposition 8.27); and

(b) the primary ideals corresponding to all isolated components are uniquely determined (see Proposition 8.34).

So in our example above, P_1, P_2, and Q_1 are uniquely determined, and only the primary ideal corresponding to P_2 can depend on the decomposition.

To prove the first statement (a), we give an alternative way to reconstruct all underlying prime ideals of a minimal primary decomposition (the so-called *associated prime ideals*) without knowing the decomposition at all.

Lemma 8.24. *Let Q be a P-primary ideal in a ring R. Then for any $a \in R$ we have*

$$\sqrt{Q : a} = \begin{cases} R & \text{if } a \in Q, \\ P & \text{if } a \notin Q. \end{cases}$$

Proof. If $a \in Q$ then clearly $Q : a = R$, and thus $\sqrt{Q : a} = R$ as well.

Now let $a \notin Q$. Then for any $b \in Q : a$ we have $ab \in Q$, and so $b \in P$ since Q is P-primary. Hence $Q \subset Q : a \subset P$, and by taking radicals we obtain $P \subset \sqrt{Q : a} \subset P$ as desired. □

Definition 8.25 (Associated, isolated, and embedded prime ideals)**.** Let I be an ideal in a ring R.

(a) An **associated prime ideal** of I is a prime ideal that can be written as $\sqrt{I : a}$ for some $a \in R$. We denote the set of these associated primes by $\mathrm{Ass}(I)$.

(b) The minimal elements of $\mathrm{Ass}(I)$ are called **isolated prime ideals** of I, the other ones **embedded prime ideals** of I.

Remark 8.26. For an ideal I of a ring R, note that not every ideal that can be written as $\sqrt{I : a}$ for some $a \in R$ is prime. By definition, $\mathrm{Ass}(I)$ contains only the prime ideals of this form.

Proposition 8.27 (**First Uniqueness Theorem** for primary decompositions)**.** *Let $Q_1, \dots, Q_n$ form a minimal primary decomposition for an ideal I in a ring R, and let $P_i = \sqrt{Q_i}$ for $i = 1, \dots, n$.*

Then $\{P_1, \dots, P_n\} = \operatorname{Ass}(I)$. In particular, the number of components in a minimal primary decomposition and their radicals do not depend on the chosen decomposition.

Proof.

"$\subset$": Let $i \in \{1, \dots, n\}$, we will show that $P_i \in \operatorname{Ass}(I)$. As the given decomposition is minimal, we can find $a \in \bigcap_{j \neq i} Q_j$ with $a \notin Q_i$. Then

$$\begin{aligned} \sqrt{I : a} &= \sqrt{Q_1 : a \cap \dots \cap Q_n : a} \\ &= \sqrt{Q_1 : a} \cap \dots \cap \sqrt{Q_n : a} && \text{(Lemma 1.7 (b))} \\ &= P_i && \text{(Lemma 8.24),} \end{aligned}$$

and so $P_i \in \operatorname{Ass}(I)$.

"$\supset$": Let $P \in \operatorname{Ass}(I)$, so $P = \sqrt{I : a}$ for some $a \in R$. Then as above we have

$$P = \sqrt{I : a} = \sqrt{Q_1 : a} \cap \dots \cap \sqrt{Q_n : a},$$

and thus $P \supset \sqrt{Q_i : a}$ for some i by Exercise 2.10 (a) since P is prime. But of course the above equation also implies $P \subset \sqrt{Q_i : a}$, and so $P = \sqrt{Q_i : a}$. Now by Lemma 8.24 this radical can only be P_i or R, and since P is prime we conclude that we must have $P = P_i$. □

Example 8.28. In the situation of Example 8.23, Proposition 8.27 states that the associated prime ideals of I are P_1 and P_2 since we have found a minimal primary decomposition of I with these underlying prime ideals. It is then obvious by Definition 8.25 that P_1 is an isolated and P_2 an embedded prime of I.

Exercise 8.29. Let I be an ideal in a ring R. In Definition 8.25 we have introduced $\operatorname{Ass}(I)$ as the set of prime ideals that are of the form $\sqrt{I : a}$ for some $a \in R$. If R is Noetherian, prove that we do not have to take radicals, i. e. that $\operatorname{Ass}(I)$ is also equal to the set of all prime ideals that are of the form $I : a$ for some $a \in R$.

Corollary 8.30 (Isolated prime ideals = minimal prime ideals)**.** *Let I be an ideal in a Noetherian ring R. Then the isolated prime ideals of I are exactly the minimal prime ideals over I as in Exercise 2.23, i. e. the prime ideals $P \supset I$ such that there is no prime ideal Q with $I \subset Q \subsetneq P$.*

In particular, in a Noetherian ring there are only finitely many minimal prime ideals over any given ideal.

Proof. As R is Noetherian, there is a minimal primary decomposition $I = Q_1 \cap \dots \cap Q_n$ of I by Corollary 8.21. As usual we set $P_i = \sqrt{Q_i}$ for all i, so that $\operatorname{Ass}(I) = \{P_1, \dots, P_n\}$ by Proposition 8.27.

Note that if $P \supset I$ is any prime ideal, then $P \supset Q_1 \cap \dots \cap Q_n$, hence $P \supset Q_i$ for some i by Exercise 2.10 (a), and so by taking radicals $P \supset \sqrt{Q_i} = P_i$. With this we now show both implications stated in the corollary:

- Let $P_i \in \operatorname{Ass}(I)$ be an isolated prime ideal of I. If P is any prime ideal with $I \subset P \subset P_i$ then by what we have just said $P_j \subset P \subset P_i$ for some j. But as P_i is isolated we must have equality, and so $P = P_i$. Hence P_i is minimal over I.
- Now let P be a minimal prime over I. By the above $I \subset Q_i \subset P_i \subset P$ for some i. As P is minimal over I this means that $P = P_i$ is an associated prime, and hence also an isolated prime of I. □

Remark 8.31. In particular, Corollary 8.30 states that the isolated prime ideals of an ideal I in a coordinate ring of a variety $A(X)$ correspond exactly to the maximal subvarieties, i. e. to the irreducible components of $V(I)$ — as already motivated in Example 8.23.

Exercise 8.32. Let R be a Noetherian integral domain. Show:

(a) R is a unique factorization domain if and only if every minimal prime ideal over a principal ideal is itself principal.

(b) If R is a unique factorization domain then every minimal non-zero prime ideal of R is principal.

Finally, to prove the second uniqueness statement (b) of Example 8.23 the idea is to use localization at an isolated prime ideal to remove from I all components that do not belong to this prime ideal.

Lemma 8.33. *Let S be a multiplicatively closed subset in a ring R, and let Q be a P-primary ideal in R. Then with respect to the ring homomorphism $\varphi : R \to S^{-1}R$, $a \mapsto \frac{a}{1}$ we have*

$$(Q^e)^c = \begin{cases} R & \text{if } S \cap P \neq \emptyset, \\ Q & \text{if } S \cap P = \emptyset. \end{cases}$$

Proof. If $S \cap P \neq \emptyset$ there is an element $s \in S$ with $s \in P = \sqrt{Q}$, and thus $s^n \in Q$ for some $n \in \mathbb{N}$. So $\frac{1}{1} = \frac{s^n}{s^n} \in S^{-1}Q = Q^e$ by Example 6.18. Hence $Q^e = S^{-1}R$, and therefore $(Q^e)^c = R$.

On the other hand, assume now that $S \cap P = \emptyset$. By Exercise 1.19 (a) it suffices so prove $(Q^e)^c \subset Q$. If $a \in (Q^e)^c$ we have $\frac{a}{1} \in Q^e$, and so $\frac{a}{1} = \frac{q}{s}$ for some $q \in Q$ and $s \in S$ by Proposition 6.7 (a). Hence $u(q - as) = 0$ for some $u \in S$, which implies that $a \cdot us = uq \in Q$. As Q is P-primary it follows that $a \in Q$ or $us \in P$. But $us \in S$, so $us \notin P$ since $S \cap P = \emptyset$, and we conclude that $a \in Q$. □

Proposition 8.34 (**Second Uniqueness Theorem** for primary decompositions)**.** *Let $Q_1, \ldots, Q_n$ form a minimal primary decomposition for an ideal I in a ring R, and let $P_i = \sqrt{Q_i}$ for $i = 1, \ldots, n$.*

If $i \in \{1, \ldots, n\}$ such that P_i is minimal over I, then $(I^e)^c = Q_i$, where contraction and extension are taken with respect to the canonical localization map $R \to R_{P_i}$. In particular, in a minimal primary decomposition the primary components corresponding to minimal prime ideals do not depend on the chosen decomposition.

Proof. Localizing the equation $I = Q_1 \cap \cdots \cap Q_n$ at S gives $S^{-1}I = S^{-1}Q_1 \cap \cdots \cap S^{-1}Q_n$ by Exercise 6.24 (b), hence $I^e = Q_1^e \cap \cdots \cap Q_n^e$ by Example 6.18, and so $(I^e)^c = (Q_1^e)^c \cap \cdots \cap (Q_n^e)^c$ by Exercise 1.19 (d).

Now let P_i be minimal over I, and set $S = R \backslash P_i$. Then $S \cap P_i = \emptyset$, whereas $S \cap P_j \neq \emptyset$ for all $j \neq i$ since $P_j \not\subset P_i$. So applying Lemma 8.33 gives $(I^e)^c = (Q_i^e)^c = Q_i$ as desired. □

15

9. Integral Ring Extensions

In this chapter we want to discuss a concept in commutative algebra that has its original motivation in algebra, but turns out to have surprisingly many applications and consequences in geometry as well. To explain its background, recall from the "Introduction to Algebra" class that the most important objects in field theory are algebraic and finite field extensions. More precisely, if $K \subset K'$ is an inclusion of fields an element $a \in K'$ is called *algebraic* over K if there is a non-zero polynomial $f \in K[x]$ with coefficients in K such that $f(a) = 0$. The field extension $K \subset K'$ is then called algebraic if every element of K' is algebraic over K [G3, Definition 2.1].

Of course, for an algebraic element $a \in K'$ over K there is then also a *monic* polynomial relation $a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$ for some $n \in \mathbb{N}_{>0}$ and $c_0, \ldots, c_{n-1} \in K$, since we can just divide f by its leading coefficient. This trivial statement actually has far-reaching consequences: it means that we can use the equation $a^n = -c_{n-1}a^{n-1} - \cdots - c_0$ to reduce any monomial a^k for $k \in \mathbb{N}$ (and in fact also the multiplicative inverses of non-zero polynomial expressions in a) to a K-linear combination of the first n powers $1, a, \ldots, a^{n-1}$, and that consequently the extension field $K(a)$ of K generated by a is a *finite-dimensional* vector space over K — we say that $K \subset K(a)$ is a *finite field extension* [G3, Definition 2.12 and Proposition 2.14 (b)]. This means that the field extension $K \subset K(a)$ is quite easy to deal with, since we can use the whole machinery of (finite-dimensional) linear algebra for its study.

What happens now if instead of an extension $K \subset K'$ of *fields* we consider an extension $R \subset R'$ of *rings*? We can certainly still have a look at elements $a \in R'$ satisfying a polynomial relation $c_n a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$ with $c_0, \ldots, c_n \in R$ (and not all of them being 0). But now it will in general not be possible to divide this equation by its leading coefficient c_n to obtain a monic relation. Consequently, we can in general not use this relation to express higher powers of a in terms of lower ones, and hence the R-algebra $R[a]$ generated by a over R need not be a finite R-module. A simple example for this can already be found in the ring extension $\mathbb{Z} \subset \mathbb{Q}$: for example, the rational number $a = \frac{1}{2}$ satisfies a (non-monic) polynomial relation $2a - 1 = 0$ with coefficients in $\mathbb{Z}$, but certainly $\mathbb{Z}[a]$, which is the ring of all rational numbers with a finite binary expansion, is not a finitely generated $\mathbb{Z}$-module.

We learn from this that in the case of a ring extension $R \subset R'$ we should often not consider elements of R' satisfying polynomial relations with coefficients in R, but rather require *monic* relations in the first place. So let us start by giving the corresponding definitions.

Definition 9.1 (Integral and finite ring extensions)**.**

(a) If $R \subset R'$ are rings, we call R' an **extension ring** of R. We will also say in this case that $R \subset R'$ is a **ring extension**.

Note: sometimes in the literature a ring extension is meant to be any ring homomorphism $R \to R'$, even if it is not injective (so that R' is an arbitrary R-algebra as in Definition 1.23 (a)).

(b) Let R be a ring. An element a of an extension ring R' of R is called **integral** over R if there is a monic polynomial $f \in R[x]$ with $f(a) = 0$, i. e. if there are $n \in \mathbb{N}_{>0}$ and $c_0, \ldots, c_{n-1} \in R$ with $a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$. We say that R' is **integral** over R if every element of R' is integral over R.

(c) A ring extension $R \subset R'$ is called **finite** if R' is finitely generated as an R-module.

Remark 9.2.

(a) Note that the usage of the word "integral" for the concept of Definition 9.1 (b) is completely unrelated to the appearance of the same word in the term "integral domain".

(b) If R and R' are fields, the notions of integral and finite ring extensions actually coincide with those of algebraic and finite field extensions [G3, Definitions 2.1 and 2.12]. In fact, for this case you might already know some of the next few results in this chapter from the "Introduction to Algebra" class, in particular Proposition 9.5 and Lemma 9.6.

Example 9.3. Let R be a unique factorization domain, and let $R' = \operatorname{Quot} R$ be its quotient field. We claim that $a \in R'$ is integral over R if and only if $a \in R$.

In fact, it is obvious that any element of R is integral over R, so let us prove the converse. Assume that $a = \frac{p}{q}$ is integral over R with p and q coprime, i. e. there is a polynomial equation

$$\left(\frac{p}{q}\right)^n + c_{n-1}\left(\frac{p}{q}\right)^{n-1} + \cdots + c_0 = 0$$

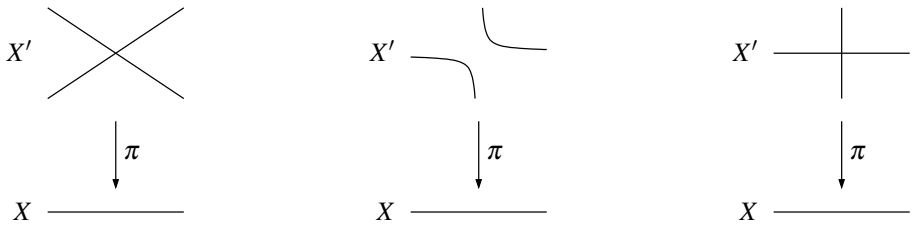
with $c_0, \ldots, c_{n-1} \in R$. We multiply with q^n to get

$$p^n + c_{n-1}p^{n-1}q + \cdots + c_0 q^n = 0, \quad \text{i. e.} \quad p^n = -q(c_{n-1}p^{n-1} + \cdots + c_0 q^{n-1}).$$

in R. So $q \mid p^n$, which is only possible if q is a unit since p and q are coprime. Hence $a = \frac{p}{q} \in R$.

So in particular, if R is not a field then $R' \neq R$, and hence the ring extension $R \subset R'$ is not integral.

Example 9.4 (Geometric examples of integral extensions)**.** Let $R = \mathbb{C}[x]$ and $R' = R[y]/(f) = \mathbb{C}[x,y]/(f)$, where $f \in R[y]$ is a (non-constant) polynomial relation for the additional variable y. Geometrically, we then have $R = A(X)$ and $R' = A(X')$ for $X = \mathbb{A}^1_{\mathbb{C}}$ and the curve X' in $\mathbb{A}^2_{\mathbb{C}}$ given by the equation $f(x,y) = 0$. The ring extension map $R \to R'$ corresponds to the morphism of varieties $\pi : X' \to X$, $(x,y) \mapsto x$ in the sense of Construction 0.11. In the following figure we show three examples of this setting, where we only draw the real points to keep the pictures inside $\mathbb{A}^2_{\mathbb{R}}$.



$f = y^2 - x^2$ (a) $\quad f = xy - 1$ (b) $\quad f = xy$ (c)

The subtle but important difference between these examples is that in case (a) the given relation $f \in R[y]$ is monic in y, whereas in (b) and (c) it is not (with the leading term being xy). This has geometric consequences for the inverse images $\pi^{-1}(x)$ of points $x \in X$, the so-called *fibers* of π:

(a) In this case, the generator $\bar{y}$ of R' over R is integral since it satisfies the monic relation $\bar{y}^2 - \bar{x}^2 = 0$. In fact, Proposition 9.5 will show that this implies that the whole ring extension $R \subset R'$ is integral. Geometrically, the monic relation means that plugging in an arbitrary value for x will always give a quadratic equation $y^2 - x^2 = 0$ for y, leading to two points in any fiber $\pi^{-1}(x)$ (counted with multiplicities).

(b) In this example, $\bar{y} \in R'$ does not satisfy a monic relation over R: considering the leading term in y it is obvious that there are no polynomials $g, h \in R[y]$ with g monic in y such that $g = h(xy - 1)$. Hence the extension $R \subset R'$ is not integral. Geometrically, the consequence is now that after plugging in a value for x the relation $xy - 1 = 0$ for y is linear for $x \neq 0$ but constant for $x = 0$, leading to an empty fiber $\pi^{-1}(0) = \emptyset$ whereas all other fibers contain a single point.

(c) This case is similar to (c): again, the ring extension $R \subset R'$ is not integral, and the relation $xy = 0$ does not remain linear in y when setting $x = 0$. This time however, this leads to an infinite fiber $\pi^{-1}(0)$ instead of an empty one.

Summarizing, we would expect that a morphism of varieties corresponding to an integral ring extension is surjective with finite fibers. In fact, we will see this in Example 9.19, and thinking of integral extensions geometrically as morphisms with this property is a good first approximation — even if the precise geometric correspondence is somewhat more complicated (see e. g. Example 9.25).

But let us now start our rigorous study of integral and finite extensions by proving their main algebraic properties.

Proposition 9.5 (Integral and finite ring extensions)**.** *An extension ring R' is finite over R if and only if $R' = R[a_1, \dots, a_n]$ for integral elements $a_1, \dots, a_n \in R'$ over R.*

Moreover, in this case the whole ring extension $R \subset R'$ is integral.

Proof.

"$\Rightarrow$": Let $R' = \langle a_1, \dots, a_n \rangle$ be finitely generated as an R-module. Of course, we then also have $R' = R[a_1, \dots, a_n]$, i. e. R' is generated by the same elements as an R-algebra. We will prove that every element of R' is integral over R, which then also shows the "moreover" statement.

So let $a \in R'$. As R' is finite over R, we can apply the Cayley-Hamilton theorem of Proposition 3.25 to the R-module homomorphism $\varphi : R' \to R'$, $x \mapsto ax$ to obtain a monic polynomial equation

$$\varphi^k + c_{k-1}\varphi^{k-1} + \dots + c_0 = 0$$

in $\mathrm{Hom}_R(R', R')$ for some $c_0, \dots, c_{k-1} \in R$, and hence $a^k + c_{k-1}a^{k-1} + \dots + c_0 = 0$ by plugging in the value 1. Thus a is integral over R.

"$\Leftarrow$": Let $R' = R[a_1, \dots, a_n]$ with $a_1, \dots, a_n$ integral over R, i. e. every a_i satisfies a monic polynomial relation of some degree r_i over R. Now by Lemma 1.28 every element of R' is a polynomial expression of the form $\sum_{k_1,\dots,k_n} c_{k_1,\dots,k_n} a_1^{k_1} \cdot \dots \cdot a_n^{k_n}$ for some $c_{k_1,\dots,k_n} \in R$, and we can use the above polynomial relations to reduce the exponents to $k_i < r_i$ for all $i = 1, \dots, n$. Hence R' is finitely generated over R by all monomial expressions $a_1^{k_1} \cdot \dots \cdot a_n^{k_n}$ with $k_i < r_i$ for all i. □

Lemma 9.6 (Transitivity of integral and finite extensions)**.** *Let $R \subset R' \subset R''$ be rings.*

(a) *If $R \subset R'$ and $R' \subset R''$ are finite, then so is $R \subset R''$.*

(b) *If $R \subset R'$ and $R' \subset R''$ are integral, then so is $R \subset R''$.*

Proof.

(a) Let $a_1, \dots, a_n$ generate R' as an R-module, and $b_1, \dots, b_m$ generate R'' as an R'-module. Then every element of R'' is of the form $\sum_{i=1}^m c_i b_i$ for some $c_i \in R'$, i. e. $\sum_{i=1}^m \left(\sum_{j=1}^n c_{i,j} a_j\right) \cdot b_i$ for some $c_{i,j} \in R$. Hence the finitely many products $a_j b_i$ generate R'' as an R-module.

(b) Let $a \in R''$. As a is integral over R', there are $n \in \mathbb{N}_{>0}$ and elements $c_0, \dots, c_{n-1}$ of R' such that $a^n + c_{n-1}a^{n-1} + \dots + c_0 = 0$. Then a is also integral over $R[c_0, \dots, c_{n-1}]$. In addition, we know that $c_0, \dots, c_{n-1}$ are integral over R. Hence Proposition 9.5 tells us that $R[c_0, \dots, c_{n-1}, a]$ is finite over $R[c_0, \dots, c_{n-1}]$ and $R[c_0, \dots, c_{n-1}]$ is finite over R. Therefore $R[c_0, \dots, c_{n-1}, a]$ is finite over R by (a), and thus a is integral over R by Proposition 9.5 again. □

A nice property of integral extensions is that they are compatible with quotients, localizations, and polynomial rings in the following sense.

Lemma 9.7. *Let R' be an integral extension ring of R.*

(a) *If I is an ideal of R' then R'/I is an integral extension ring of $R/(I \cap R)$.*

(b) *If S is a multiplicatively closed subset of R then $S^{-1}R'$ is an integral extension ring of $S^{-1}R$.*

(c) *$R'[x]$ is an integral extension ring of $R[x]$.*

Proof.

(a) Note that the map $R/(I\cap R) \to R'/I$, $\overline{a} \mapsto \overline{a}$ is well-defined and injective, hence we can regard R'/I as an extension ring of $R/(I\cap R)$. Moreover, for all $a \in R'$ there is a monic relation $a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$ with $c_0,\ldots,c_{n-1} \in R$, and hence by passing to the quotient also $\overline{a}^n + \overline{c_{n-1}}\,\overline{a}^{n-1} + \cdots + \overline{c_0} = 0$. So $\overline{a}$ is integral over $R/(I\cap R)$.

(b) Again, the ring homomorphism $S^{-1}R \to S^{-1}R'$, $\frac{a}{s} \mapsto \frac{a}{s}$ is obviously well-defined and injective. Moreover, for $\frac{a}{s} \in S^{-1}R'$ we have a monic relation $a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$ with $c_0,\ldots,c_{n-1} \in R$, and thus also

$$\left(\frac{a}{s}\right)^n + \frac{c_{n-1}}{s}\left(\frac{a}{s}\right)^{n-1} + \cdots + \frac{c_0}{s^n} = 0.$$

Hence $\frac{a}{s}$ is integral over $S^{-1}R$.

(c) Let $f = a_n x^n + \cdots + a_0 \in R'[x]$, i. e. $a_0,\ldots,a_n \in R'$. Then $a_0,\ldots,a_n$ are integral over R, so also over $R[x]$, and thus $R[x][a_0,\ldots,a_n] = R[a_0,\ldots,a_n][x]$ is integral over $R[x]$ by Proposition 9.5. In particular, this means that f is integral over $R[x]$. □

Exercise 9.8.

(a) Is $\sqrt{2+\sqrt{2}} + \frac{1}{2}\sqrt[3]{3} \in \mathbb{R}$ integral over $\mathbb{Z}$?

(b) Let $R' = \mathbb{R}[x]$, $R = \mathbb{R}[x^2-1] \subset R'$, $P' = (x-1) \trianglelefteq R'$, and $P = P' \cap R$. Show that R' is an integral extension of R, but the localization $R'_{P'}$ is not an integral extension of R_P. Is this a contradiction to Lemma 9.7 (b)?

(Hint: consider the element $\frac{1}{x+1}$.)

An important consequence of our results obtained so far is that the integral elements of a ring extension always form a ring themselves. This leads to the notion of integral closure.

Corollary and Definition 9.9 (Integral closures)**.**

(a) *Let $R \subset R'$ be a ring extension. The set $\overline{R}$ of all integral elements in R' over R is a ring with $R \subset \overline{R} \subset R'$. It is called the* ***integral closure*** *of R in R'. We say that R is* ***integrally closed in*** *R' if its integral closure in R' is R.*

(b) *An integral domain R is called* ***integrally closed*** *or* ***normal*** *if it is integrally closed in its quotient field* $\operatorname{Quot} R$.

Proof. It is clear that $R \subset \overline{R} \subset R'$, so we only have to show that $\overline{R}$ is a subring of R'. But this follows from Proposition 9.5: if $a,b \in R'$ are integral over R then so is $R[a,b]$, and hence in particular $a+b$ and $a\cdot b$. □

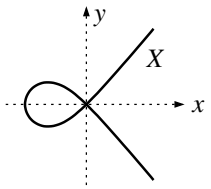
Example 9.10. Every unique factorization domain R is normal, since by Example 9.3 the only elements of $\operatorname{Quot} R$ that are integral over R are the ones in R.

In contrast to Remark 9.2 (b), note that for a field R Definition 9.9 (b) of an integrally closed domain does not specialize to that of an algebraically closed field: we do not require that R admits no integral extensions at all, but only that it has no integral extensions within its quotient field $\operatorname{Quot} R$. The approximate geometric meaning of this concept can be seen in the following example.

Example 9.11 (Geometric interpretation of normal domains)**.** Let $R = A(X)$ be the coordinate ring of an irreducible variety X. The elements $\varphi = \frac{f}{g} \in \operatorname{Quot} R$ of the quotient field can then be interpreted as rational functions on X, i. e. as quotients of polynomial functions that are well-defined except at some isolated points of X (where the denominator g vanishes). Hence the condition of R being normal means that every rational function φ satisfying a monic relation $\varphi^n + c_{n-1}\varphi^{n-1} + \cdots + c_0 = 0$ with $c_0,\ldots,c_{n-1} \in R$ is already an element of R, so that its value is well-defined at every point of X. Now let us consider the following two concrete examples:

(a) Let $R = \mathbb{C}[x]$, corresponding to the variety $X = \mathbb{A}^1_{\mathbb{C}}$. By Example 9.10 we know that R is normal. In fact, this can also be understood geometrically: the only way a rational function φ on $\mathbb{A}^1_{\mathbb{C}}$ can be ill-defined at a point $a \in \mathbb{A}^1_{\mathbb{C}}$ is that it has a pole, i. e. that it is of the form $x \mapsto \frac{f}{(x-a)^k}$ for some $k \in \mathbb{N}_{>0}$ and $f \in \operatorname{Quot} R$ that is well-defined and non-zero at a. But then φ cannot satisfy a monic relation of the form $\varphi^n + c_{n-1}\varphi^{n-1} + \cdots + c_0 = 0$ with $c_0, \ldots, c_{n-1} \in \mathbb{C}[x]$ since φ^n has a pole of order kn at a which cannot be canceled by the lower order poles of the other terms $c_{n-1}\varphi^{n-1} + \cdots + c_0 = 0$. Hence any rational function satisfying such a monic relation is already a polynomial function, which means that R is normal.

(b) Let $X = V(y^2 - x^2 - x^3) \subset \mathbb{A}^2_{\mathbb{R}}$ and $R = A(X) = \mathbb{R}[x,y]/(y^2 - x^2 - x^3)$. The curve X is shown in the picture on the right: locally around the origin (i. e. for small x and y) the term x^3 is small compared to x^2 and y^2, and thus X is approximatively given by $(y-x)(y+x) = y^2 - x^2 \approx 0$, which means that it consists of two branches crossing the origin with slopes ± 1.



In this case the ring R is not normal: the rational function $\varphi = \frac{y}{x} \in \operatorname{Quot}(R) \setminus R$ satisfies the monic equation $\varphi^2 - x - 1 = \frac{y^2}{x^2} - x - 1 = \frac{x^3+x^2}{x^2} - x - 1 = 0$. Geometrically, the reason why φ is ill-defined at 0 is not that it has a pole (i. e. tends to ∞ there), but that it approaches two different values 1 and -1 on the two local branches of the curve. This means that φ^2 approaches a unique value 1 at the origin, and thus has a well-defined value at this point — leading to the monic quadratic equation for φ. So the reason for R not being normal is the "singular point" of X at the origin. In fact, one can think of the normality condition geometrically as some sort of "non-singularity" statement (see also Example 11.37 and Remark 12.15 (b)).

Exercise 9.12 (Integral closures can remove singularities)**.** As in Example 9.11 (b) let us again consider the ring $R = \mathbb{R}[x,y]/(y^2 - x^2 - x^3)$, and let $K = \operatorname{Quot} R$ be its quotient field. For the element $t := \frac{y}{x} \in K$, show that the integral closure $\overline{R}$ of R in K is $R[t]$, and that this is equal to $\mathbb{R}[t]$. What is the geometric morphism of varieties corresponding to the ring extension $R \subset \overline{R}$?

Exercise 9.13. Let R be an integral domain, and let $S \subset R$ be a multiplicatively closed subset. Prove:

(a) Let $R' \supset R$ be an extension ring of R. If $\overline{R}$ is the integral closure of R in R', then $S^{-1}\overline{R}$ is the integral closure of $S^{-1}R$ in $S^{-1}R'$.

(b) If R is normal, then $S^{-1}R$ is normal.

(c) (Normality is a local property) If R_P is normal for all maximal ideals $P \trianglelefteq R$, then R is normal.

Exercise 9.14. Let $R \subset R'$ be an extension of integral domains, and let $\overline{R}$ be the integral closure of R in R'.

Show that for any two monic polynomials $f, g \in R'[t]$ with $fg \in \overline{R}[t]$ we have $f, g \in \overline{R}[t]$.

(Hint: From the "Introduction to Algebra" class you may use the fact that any polynomial over a field K has a splitting field, so in particular an extension field $L \supset K$ over which it splits as a product of linear factors [G3, Proposition 4.15].)

16

Checking whether an element $a \in R'$ of a ring extension $R \subset R'$ is integral over R can be difficult since we have to show that there is no monic polynomial over R at all that vanishes at a. The following lemma simplifies this task if R is a normal domain: it states that in this case it suffices to consider only the *minimal polynomial* of a over the quotient field $K = \operatorname{Quot} R$ (i. e. the uniquely determined monic polynomial $f \in K[x]$ of smallest degree having a as a zero [G3, Definition 2.4]), since this polynomial must already have coefficients in R if a is integral.

Lemma 9.15. *Let $R \subset R'$ be an integral extension of integral domains, and assume that R is normal.*

(a) *For any $a \in R'$ its minimal polynomial f over $K = \operatorname{Quot} R$ actually has coefficients in R.*

(b) *If moreover $a \in PR'$ for some prime ideal $P \trianglelefteq R$, then the non-leading coefficients of f are even contained in P.*

Proof.

(a) As a is integral over R there is a monic polynomial $g \in R[x]$ with $g(a) = 0$. Then $f \mid g$ over K [G3, Remark 2.5], i. e. we have $g = fh$ for some $h \in K[x]$. Applying Exercise 9.14 to the extension $R \subset K$ (and $\overline{R} = R$ since R is normal) it follows that $f \in R[x]$ as required.

(b) Let $a = p_1 a_1 + \cdots + p_k a_k$ for some $p_1, \ldots, p_k \in P$ and $a_1, \ldots, a_k \in R'$. Replacing R' by $R[a_1, \ldots, a_k]$ we may assume by Proposition 9.5 that R' is finite over R. Then we can apply the Cayley-Hamilton theorem of Proposition 3.25 to $\varphi : R' \to R'$, $x \mapsto ax$: since the image of this R-module homomorphism lies in PR', we obtain a polynomial relation

$$\varphi^n + c_{n-1}\varphi^{n-1} + \cdots + c_0 = 0 \quad \in \mathrm{Hom}_R(R', R')$$

with $c_0, \ldots, c_{n-1} \in P$. Plugging in the value 1 this means that we have a monic polynomial $g \in R[x]$ with non-leading coefficients in P such that $g(a) = 0$.

By the proof of (a) we can now write $g = fh$, where $f \in R[x]$ is the minimal polynomial of a and $h \in R[x]$. Reducing this equation modulo P gives $\overline{x}^n = \overline{f}\,\overline{h}$ in $(R/P)[x]$. But as R/P is an integral domain by Lemma 2.3 (a) this is only possible if $\overline{f}$ and $\overline{h}$ are powers of $\overline{x}$ themselves (otherwise the highest and lowest monomial in their product would have to differ). Hence the non-leading coefficients of f lie in P. □

Example 9.16 (Integral elements in quadratic number fields)**.** Let $d \in \mathbb{Z}\setminus\{0,1\}$ be a square-free integer. We want to compute the elements in $\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\} \subset \mathbb{C}$ that are integral over $\mathbb{Z}$, i. e. the integral closure $\overline{R}$ of $R = \mathbb{Z}$ in $R' = \mathbb{Q}(\sqrt{d})$. These subrings of $\mathbb{C}$ play an important role in number theory; you have probably seen them already in the "Elementary Number Theory" class [M, Chapter 8].

It is obvious that the minimal polynomial of $a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})\setminus\mathbb{Q}$ over $\mathbb{Q}$ is

$$(x - a - b\sqrt{d})(x - a + b\sqrt{d}) = x^2 - 2ax + a^2 - db^2.$$

So as $\mathbb{Z}$ is normal by Example 9.10 we know by Lemma 9.15 (a) (applied to the integral extension $R \subset \overline{R}$) that $a + b\sqrt{d}$ is integral over $\mathbb{Z}$ if and only if this polynomial has integer coefficients. Hence

$$\overline{R} = \{a + b\sqrt{d} : a, b \in \mathbb{Q}, -2a \in \mathbb{Z}, a^2 - db^2 \in \mathbb{Z}\},$$

which is the usual way how this ring is defined in the "Elementary Number Theory" class [M, Definition 8.12]. Note that this is in general not the same as the ring $\mathbb{Z} + \mathbb{Z}\sqrt{d}$ — in fact, one can show that $\overline{R} = \mathbb{Z} + \mathbb{Z}\sqrt{d}$ only if $d \neq 1 \mod 4$, whereas for $d = 1 \mod 4$ we have $\overline{R} = \mathbb{Z} + \mathbb{Z}\frac{1+\sqrt{d}}{2}$ [M, Proposition 8.16].

Remark 9.17 (Geometric properties of integral ring extensions)**.** After having discussed the algebraic properties of integral extensions, let us now turn to the geometric ones (some of which were already motivated in Example 9.4). So although we will continue to allow general (integral) ring extensions $R \subset R'$, our main examples will now be coordinate rings $R = A(X)$ and $R' = A(X')$ of varieties X and X', respectively. The inclusion map $i : R \to R'$ then corresponds to a morphism of varieties $\pi : X' \to X$ as in Construction 0.11 and Example 9.4.

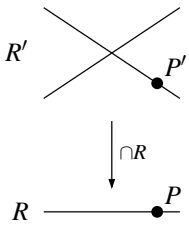
We are going to study the contraction and extension of prime ideals by the ring homomorphism i — by Remarks 1.18 and 2.7 this means that we consider images and inverse images of irreducible subvarieties under π. As i is injective, i. e. R is a subring of R', note that

- the contraction of a prime ideal $P' \trianglelefteq R'$ is exactly $(P')^c = P' \cap R$; and
- the extension of a prime ideal $P \trianglelefteq R$ is exactly $P^e = PR'$.

Moreover, by Example 2.9 (b) we know that the contraction $P' \cap R$ of a prime ideal $P' \trianglelefteq R'$ is always prime again, which means geometrically that the image of an irreducible subvariety X' under π is irreducible. The main geometric questions that we can ask are whether conversely any prime ideal

$P \trianglelefteq R$ is of the form $P' \cap R$ for some prime ideal $P' \trianglelefteq R'$ (maybe with some additional requirements about P'), corresponding to studying whether there are irreducible subvarieties of X' with given image under π (and what we can say about them).

For the rest of this chapter, our pictures to illustrate such questions will always be of the form as shown on the right: we will draw the geometric varieties and subvarieties, but label them with their algebraic counterparts. The arrow from the top to the bottom represents the map π, or algebraically the contraction map $P' \mapsto P' \cap R$ on prime ideals (with the actual ring homomorphism i going in the opposite direction). For example, in the picture on the right the point $V(P')$ in X' for a maximal ideal $P' \trianglelefteq R'$ is mapped to the point $V(P)$ in X for a maximal ideal $P \trianglelefteq R$, which means that $P' \cap R = P$.

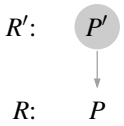


There are four main geometric results on integral ring extensions in the above spirit; they are commonly named Lying Over, Incomparability, Going Up, and Going Down. We will treat them now in turn. The first one, Lying Over, is the simplest of them — it just asserts that for an integral extension $R \subset R'$ every prime ideal in R is the contraction of a prime ideal in R'. In our pictures, we will always use gray color to indicate objects whose existence we are about to prove.

Proposition 9.18 (Lying Over). *Let $R \subset R'$ be a ring extension, and $P \trianglelefteq R$ prime.*

(a) *There is a prime ideal $P' \trianglelefteq R'$ with $P' \cap R = P$ if and only if $PR' \cap R \subset P$.*

(b) *If $R \subset R'$ is integral, then this is always the case.*



We say in this case that P' is lying over P.

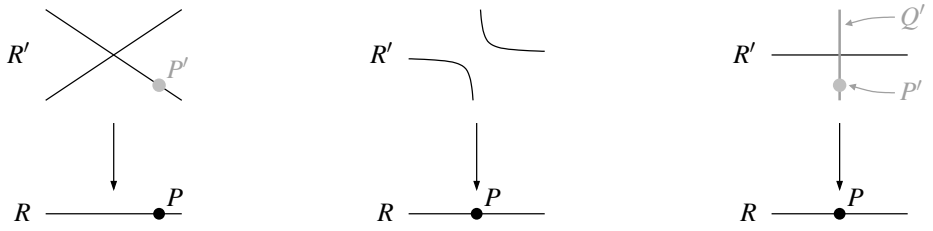
Proof.

(a) "⇒" If $P' \cap R = P$ then $PR' \cap R = (P' \cap R)R' \cap R \subset P'R' \cap R = P' \cap R = P$.

"⇐" Consider the multiplicatively closed set $S = R \setminus P$. As $PR' \cap S = (PR' \cap R) \setminus P = \emptyset$ by assumption, Exercise 6.14 (a) implies that there is a prime ideal $P' \trianglelefteq R'$ with $PR' \subset P'$ and $P' \cap S = \emptyset$. But the former inclusion implies $P \subset PR' \cap R \subset P' \cap R$ and the latter $P' \cap R = P' \cap P \subset P$, so we get $P' \cap R = P$ as desired.

(b) Let $a \in PR' \cap R$. Since $a \in PR'$ it follows from the Cayley-Hamilton theorem as in the first half of the proof of Lemma 9.15 (b) that there is a monic relation $a^n + c_{n-1}a^{n-1} + \dots + c_0 = 0$ with $c_0, \dots, c_{n-1} \in P$. As moreover $a \in R$, this means that $a^n = -c_{n-1}a^{n-1} - \dots - c_0 \in P$, and thus $a \in P$ as P is prime. □

Example 9.19. Let us consider the three ring extensions R' of $R = \mathbb{C}[x]$ from Example 9.4 again.

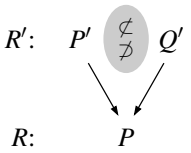


$R' = \mathbb{C}[x,y]/(y^2 - x^2)$ (a) $\quad R' = \mathbb{C}[x,y]/(xy-1)$ (b) $\quad R' = \mathbb{C}[x,y]/(xy)$ (c)

Recall that the extension (a) is integral by Proposition 9.5. Correspondingly, the picture above on the left shows a prime ideal P' lying over P, i. e. such that $P' \cap R = P$. In contrast, in case (b) there is no prime ideal lying over P, which means by Proposition 9.18 (b) that the ring extension $R \subset R'$ cannot be integral (as we have already seen in Example 9.4). In short, over an algebraically closed field the restriction of the Lying Over property to maximal ideals $P \trianglelefteq R$ (i. e. to points of the corresponding variety) means that morphisms of varieties corresponding to integral ring extensions are always surjective.

Of course, a prime ideal $P' \trianglelefteq R'$ lying over a given prime ideal $P \trianglelefteq R$ is in general not unique — there are e. g. two choices for P' in example (a) above. The case (c) is different however: as the fiber over P is one-dimensional, there are not only many choices for prime ideals lying over P, but also such prime ideals P' and Q' with $Q' \subsetneq P'$ as shown in the picture above. We will prove now that such a situation cannot occur for integral ring extensions, which essentially means that the fibers of the corresponding morphisms have to be finite.

Proposition 9.20 (Incomparability). *Let $R \subset R'$ be an integral ring extension. If P' and Q' are distinct prime ideals in R' with $P' \cap R = Q' \cap R$ then $P' \not\subset Q'$ and $Q' \not\subset P'$.*



Proof. Let $P' \cap R = Q' \cap R$ and $P' \subset Q'$. We will prove that $Q' \subset P'$ as well, so that $P' = Q'$.

Assume for a contradiction that there is an element $a \in Q' \setminus P'$. By Lemma 9.7 (a) we know that R'/P' is integral over $R/(P' \cap R)$, so there is a monic relation

$$\overline{a}^n + \overline{c_{n-1}}\,\overline{a}^{n-1} + \cdots + \overline{c_0} = 0 \qquad (*)$$

in R'/P' with $c_0, \ldots, c_{n-1} \in R$. Pick such a relation of minimal degree n. Since $a \in Q'$ this relation implies $\overline{c_0} \in Q'/P'$, but as $c_0 \in R$ too we conclude that $\overline{c_0} \in (Q' \cap R)/(P' \cap R) = (Q' \cap R)/(Q' \cap R) = 0$. Hence $(*)$ has no constant term. But since $\overline{a} \neq 0$ in the integral domain R'/P' we can then divide the relation by $\overline{a}$ to get a monic relation of smaller degree — in contradiction to the choice of n. □

In geometric terms, the following corollary is essentially a restatement of the finite fiber property: it says that in integral ring extensions only maximal ideals can contract to maximal ideals, i. e. that points are the only subvarieties that can map to a single point in the target space.

Corollary 9.21. *Let $R \subset R'$ be an integral ring extension.*

(a) *If R and R' are integral domains then R is a field if and only if R' is a field.*

(b) *A prime ideal $P' \trianglelefteq R'$ is maximal if and only if $P' \cap R$ is maximal.*

Proof.

(a) "⇒" Assume that R is a field, and let $P' \trianglelefteq R'$ be a maximal ideal. Moreover, consider the zero ideal $0 \trianglelefteq R'$, which is prime since R' is an integral domain. Both ideals contract to a prime ideal in R by Exercise 2.9, hence to 0 since R is a field. Incomparability as in Proposition 9.20 then implies that $P' = 0$. So 0 is a maximal ideal of R', and thus R' is a field.

"⇐" Now assume that R is not a field. Then there is a non-zero maximal ideal $P \trianglelefteq R$. By Lying Over as in Proposition 9.18 (b) there is now a prime ideal $P' \trianglelefteq R'$ with $P' \cap R = P$, in particular with $P' \neq 0$. So R' has a non-zero prime ideal, i. e. R' is not a field.

(b) By Lemmas 2.3 (a) and 9.7 (a) we know that $R/(P' \cap R) \subset R'/P'$ is an integral extension of integral domains, so the result follows from (a) with Lemma 2.3 (b). □

Exercise 9.22. Which of the following extension rings R' are integral over $R = \mathbb{C}[x]$?

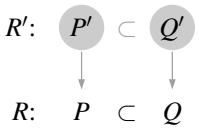
(a) $R' = \mathbb{C}[x,y,z]/(z^2 - xy)$;

(b) $R' = \mathbb{C}[x,y,z]/(z^2 - xy, y^3 - x^2)$;

(c) $R' = \mathbb{C}[x,y,z]/(z^2 - xy, x^3 - yz)$.

17

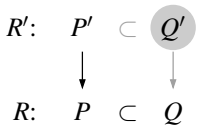
Remark 9.23. In practice, one also needs "relative versions" of the Lying Over property: let us assume that we have an (integral) ring extension $R \subset R'$ and two prime ideals $P \subset Q$ in R. If we are now given a prime ideal P' lying over P or a prime ideal Q' lying over Q, can we fill in the other prime ideal so that $P' \subset Q'$ and we get a diagram as shown on the right?



Although these two questions look symmetric, their behavior is in fact rather different. Let us start with the easier case — the so-called Going Up property — in which we prescribe P' and are looking for the bigger prime ideal Q'. It turns out that this always works for integral extensions. The proof is quite simple: we can reduce it to the standard Lying Over property by first taking the quotient by P' since this keeps only the prime ideals containing P' by Lemma 1.21.

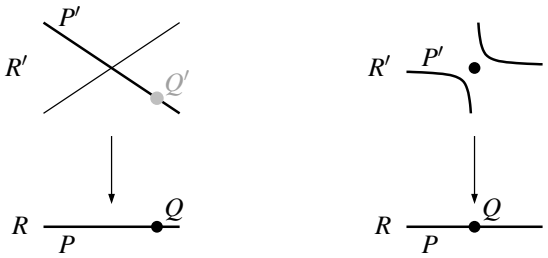
Proposition 9.24 (Going Up). *Let $R \subset R'$ be an integral ring extension. Moreover, let $P, Q \trianglelefteq R$ be prime ideals with $P \subset Q$, and let $P' \trianglelefteq R'$ be prime with $P' \cap R = P$.*

Then there is a prime ideal $Q' \trianglelefteq R'$ with $P' \subset Q'$ and $Q' \cap R = Q$.



Proof. As R' is integral over R, we know that R'/P' is integral over $R/(P' \cap R) = R/P$ by Lemma 9.7 (a). Now Q/P is prime by Corollary 2.4, and hence Lying Over as in Proposition 9.18 (b) implies that there is a prime ideal in R'/P' contracting to Q/P, which must be of the form Q'/P' for a prime ideal $Q' \trianglelefteq R'$ with $P' \subset Q'$ by Lemma 1.21 and Corollary 2.4 again. Now $(Q'/P') \cap R/P = Q/P$ means that $Q' \cap R = Q$, and so the result follows. □

Example 9.25. The ring extension (a) below, which is the same as in Example 9.4 (a), shows a typical case of the Going Up property (note that the correspondence between subvarieties and ideals reverses inclusions, so that the bigger prime ideal Q resp. Q' corresponds to the smaller subvariety).



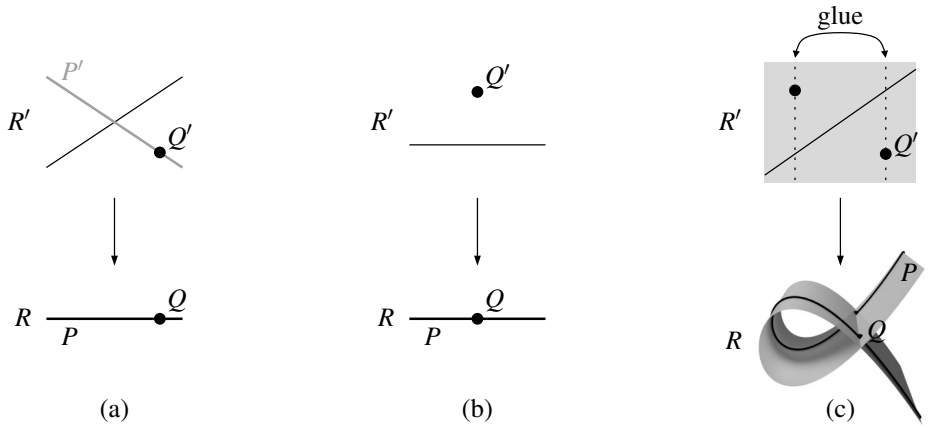
In contrast, in the extension (b) the ring

$$R' = \mathbb{C}[x,y]/I \quad \text{with} \quad I = (xy-1) \cap (x,y)$$

is the coordinate ring of the variety as in Example 9.4 (b) together with the origin. In this case, it is easily seen geometrically that the extension $R \subset R'$ with $R = \mathbb{C}[x]$ satisfies the Lying Over and Incomparability properties as explained in Example 9.19, however not the Going Up property: as in the picture above, the maximal ideal (x,y) of the origin is the only prime ideal in R' lying over Q, but it does not contain the given prime ideal P' lying over P. We see from this example that we can regard the Going Up property as a statement about the "continuity of fibers": if we have a family of points in the base space specializing to a limit point (corresponding to $P \subset Q$ in R) and a corresponding family of points in the fibers (corresponding to P'), then these points in the fiber should "converge" to a limit point Q' over Q. So by Proposition 9.24 the extension (b) above is not integral (which of course could also be checked directly).

Example 9.26. Let us now consider the opposite "Going Down" direction in Remark 9.23, i. e. the question whether we can always construct the smaller prime ideal P' from Q' (for given $P \subset Q$ of course). Picture (a) below shows this for the extension of Example 9.4 (a).

This time the idea to attack this question is to reduce it to Lying Over by localizing at Q' instead of taking quotients, since this keeps exactly the prime ideals contained in Q' by Example 6.8. Surprisingly however, in contrast to Proposition 9.24 the Going Down property does not hold for general integral extensions. The pictures (b) and (c) below show typical examples of this (in both cases it can be checked that the extension is integral):

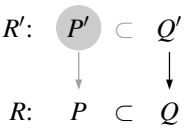


- In case (b) the space corresponding to R' has two components, of which one is only a single point lying over Q. The consequence is that Going Down obviously fails if we choose Q' to be this single point. In order to avoid such a situation we can require R' to be an integral domain, i. e. to correspond to an irreducible space.
- The case (c) is more subtle: the space $R' = \mathbb{C}[x,y]$ corresponds to a (complex) plane, and geometrically R is obtained from this by identifying the two dashed lines in the picture above. In fact, this is just a 2-dimensional version of the situation of Example 9.11 (b) and Exercise 9.12. Although both spaces are irreducible in this case, the singular shape of the base space corresponding to R makes the Going Down property fail for the choices of P, Q, and Q' shown above: note that the diagonal line and the two marked points in the top space are exactly the inverse image of the curve P in the base. As one might expect from Example 9.11 (b), we can avoid this situation by requiring R to be normal.

The resulting proposition is then the following.

Proposition 9.27 (Going Down). *Let $R \subset R'$ be an integral ring extension. Assume that R is a normal and R' an integral domain. Now let $P \subset Q$ be prime ideals in R, and let $Q' \trianglelefteq R'$ be a prime ideal with $Q' \cap R = Q$.*

Then there is a prime ideal $P' \trianglelefteq R'$ with $P' \subset Q'$ and $P' \cap R = P$.

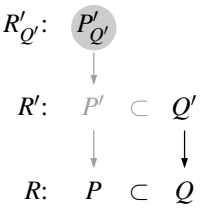


Proof. The natural map $R' \to R'_{Q'}$ is injective since R' is an integral domain. So we can compose it with the given extension to obtain a ring extension $R \subset R'_{Q'}$ as in the picture below on the right. We will show that there is a prime ideal in $R'_{Q'}$ lying over P; by Example 6.8 it must be of the form $P'_{Q'}$ for some prime ideal $P' \subset Q'$. Since $P'_{Q'}$ contracts to P' in R', we then have $P' \cap R = P$ as required.

To prove Lying Over for P in the extension $R \subset R'_{Q'}$ it suffices by Proposition 9.18 (a) to show that $PR'_{Q'} \cap R \subset P$. So let $a \in PR'_{Q'} \cap R$, in particular $a = \frac{p}{s}$ for some $p \in PR'$ and $s \in R' \setminus Q'$. We may assume without loss of generality that $a \neq 0$. As R is normal we can apply Lemma 9.15 (b) to see that the minimal polynomial of p over $K = \operatorname{Quot} R$ is of the form

$$f = x^n + c_{n-1}x^{n-1} + \cdots + c_0$$

for some $c_0, \ldots, c_{n-1} \in P$. Note that as a minimal polynomial f is irreducible over K [G3, Lemma 2.6]. But we also know that $a \in R \subset K$, and hence the polynomial



$$\frac{1}{a^n} f(ax) = x^n + \frac{c_{n-1}}{a} x^{n-1} + \cdots + \frac{c_0}{a^n} =: x^n + c'_{n-1} x^{n-1} + \cdots + c'_0$$

obtained from f by a coordinate transformation is irreducible over K as well. As it is obviously monic and satisfies $\frac{1}{a^n} f(as) = \frac{1}{a^n} f(p) = 0$, it must be the minimal polynomial of s [G3, Lemma 2.6], and so its coefficients $c'_0, \ldots, c'_{n-1}$ lie in R by Proposition 9.15 (a).

Now assume for a contradiction that $a \notin P$. The equations $c'_{n-i} a^i = c_{n-i} \in P$ of elements of R then imply $c'_{n-i} \in P$ for all $i = 1, \ldots, n$ since P is prime. So as $s \in R'$ we see that

$$s^n = -c'_{n-1} s^{n-1} - \cdots - c'_0 \quad \in PR' \subset QR' \subset Q'R' = Q',$$

which means that $s \in Q'$ since Q' is prime. This contradicts $s \in R' \setminus Q'$, and thus $a \in P$ as required. □

Exercise 9.28. Let $R \subset R'$ be an arbitrary ring extension. Show:

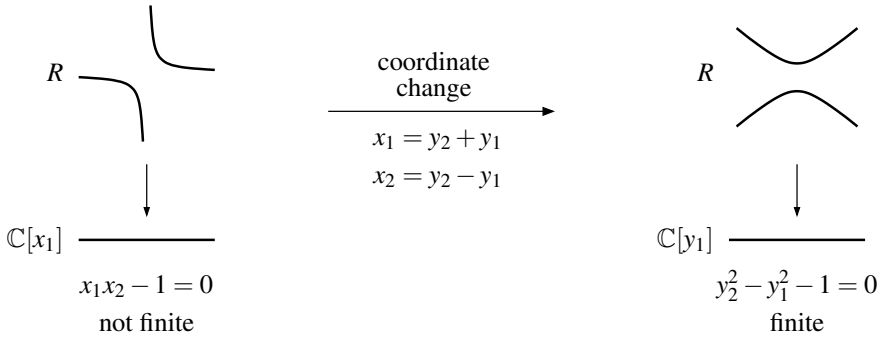
(a) The extension $R \subset R'$ has the Going Up property if and only if for all prime ideals $P' \trianglelefteq R'$ and $P = P' \cap R$ the natural map $\operatorname{Spec}(R'/P') \to \operatorname{Spec}(R/P)$ is surjective.

(b) The extension $R \subset R'$ has the Going Down property if and only if for all prime ideals $P' \trianglelefteq R'$ and $P = P' \cap R$ the natural map $\operatorname{Spec}(R'_{P'}) \to \operatorname{Spec}(R_P)$ is surjective.

10. Noether Normalization and Hilbert's Nullstellensatz

In the last chapter we have gained much understanding for integral and finite ring extensions. We now want to prove an elementary but powerful theorem stating that every finitely generated algebra R over a field K (so in particular every coordinate ring of a variety by Remark 1.31) is a finite extension ring of a polynomial ring $K[z_1,\dots,z_r]$ — and hence of a very simple K-algebra that is easy to deal with. Let us start by giving the geometric idea behind this so-called Noether Normalization theorem, which is in fact very simple.

Example 10.1 (Idea of Noether Normalization)**.** Let $R=\mathbb{C}[x_1,x_2]/(x_1x_2-1)$ be the coordinate ring of the variety $X=V(x_1x_2-1)\subset\mathbb{A}^2_{\mathbb{C}}$ as in Example 9.4 (b). We know already that R is not integral (and hence not finite) over $\mathbb{C}[x_1]$; this is easily seen geometrically in the picture below on the left since this map does not satisfy the Lying Over property for the origin as in Example 9.19.



It is easy to change this however by a linear coordinate transformation: if we set e. g. $x_1=y_2+y_1$ and $x_2=y_2-y_1$ then we can write R also as $R=\mathbb{C}[y_1,y_2]/(y_2^2-y_1^2-1)$, and this is now finite over $\mathbb{C}[y_1]$ by Proposition 9.5 since the polynomial $y_2^2-y_1^2-1$ is monic in y_2. Geometrically, the coordinate transformation has tilted the space X as in the picture above on the right so that e. g. the Lying Over property now obviously holds. Note that this is not special to the particular transformation that we have chosen; in fact, "almost any" linear coordinate change would have worked to achieve this goal.

In terms of geometry, we are therefore looking for a change of coordinates so that a suitable coordinate projection to some affine space $\mathbb{A}^r_K$ then corresponds to a finite ring extension of a polynomial ring over K in r variables. Note that this number r can already be thought of as the "dimension" of X (a concept that we will introduce in Chapter 11) as finite ring extensions correspond to surjective geometric maps with finite fibers by Example 9.19, and thus should not change the dimension (we will prove this in Lemma 11.8).

As we have seen above already, the strategy to achieve our goal is to find a suitable change of coordinates so that the given relations among the variables become monic. The first thing we have to do is therefore to prove that such a change of coordinates is always possible. It turns out that a *linear* change of coordinates works in general only for infinite fields, whereas for arbitrary fields one has to allow more general coordinate transformations.

18

Lemma 10.2. *Let $f\in K[x_1,\dots,x_n]$ be a non-zero polynomial over an infinite field K. Assume that f is homogeneous, i. e. every monomial of f has the same degree (in the sense of Exercise 0.16).*

Then there are $a_1,\dots,a_{n-1}\in K$ such that $f(a_1,\dots,a_{n-1},1)\neq 0$.

Proof. We will prove the lemma by induction on n. The case $n=1$ is trivial, since a homogeneous polynomial in one variable is just a constant multiple of a monomial.

So assume now that $n > 1$, and write f as $f = \sum_{i=0}^{d} f_i x_1^i$ where the $f_i \in K[x_2, \dots, x_n]$ are homogeneous of degree $d - i$. As f is non-zero, at least one f_i has to be non-zero. By induction we can therefore choose $a_2, \dots, a_{n-1}$ such that $f_i(a_2, \dots, a_{n-1}, 1) \neq 0$ for this i. But then $f(\cdot, a_2, \dots, a_{n-1}, 1) \in K[x_1]$ is a non-zero polynomial, so it has only finitely many zeroes. As K is infinite, we can therefore find $a_1 \in K$ such that $f(a_1, \dots, a_{n-1}, 1) \neq 0$. □

Lemma 10.3. *Let $f \in K[x_1, \dots, x_n]$ be a non-zero polynomial over an infinite field K. Then there are $\lambda \in K$ and $a_1, \dots, a_{n-1} \in K$ such that*

$$\lambda\, f(y_1 + a_1 y_n, y_2 + a_2 y_n, \dots, y_{n-1} + a_{n-1} y_n, y_n) \quad \in K[y_1, \dots, y_n]$$

is monic in y_n (i. e. as an element of $R[y_n]$ with $R = K[y_1, \dots, y_{n-1}]$).

Proof. Let d be the degree of f in the sense of Exercise 0.16, and write $f = \sum_{k_1,\dots,k_n} c_{k_1,\dots,k_n} x_1^{k_1} \cdots x_n^{k_n}$ with $c_{k_1,\dots,k_n} \in K$. Then the leading term of

$$\begin{aligned} &\lambda\, f(y_1 + a_1 y_n, y_2 + a_2 y_n, \dots, y_{n-1} + a_{n-1} y_n, y_n) \\ &= \lambda \sum_{k_1,\dots,k_n} c_{k_1,\dots,k_n} (y_1 + a_1 y_n)^{k_1} \cdots (y_{n-1} + a_{n-1} y_n)^{k_{n-1}} y_n^{k_n} \end{aligned}$$

in y_n is obtained by always taking the second summand in the brackets and only keeping the degree-d terms, i. e. it is equal to

$$\lambda \sum_{\substack{k_1,\dots,k_n \\ k_1+\dots+k_n = d}} c_{k_1,\dots,k_n} a_1^{k_1} \cdots a_{n-1}^{k_{n-1}} y_n^{k_1+\dots+k_n} = \lambda\, f_d(a_1, \dots, a_{n-1}, 1)\, y_n^d,$$

where f_d is the (homogeneous) degree-d part of f. Now pick $a_1, \dots, a_{n-1}$ by Lemma 10.2 such that $f_d(a_1, \dots, a_{n-1}, 1) \neq 0$, and set $\lambda = f_d(a_1, \dots, a_{n-1}, 1)^{-1}$. □

Exercise 10.4. Let $f \in K[x_1, \dots, x_n]$ be a non-zero polynomial over an arbitrary field K. Prove that there are $\lambda \in K$ and $a_1, \dots, a_{n-1} \in \mathbb{N}$ such that

$$\lambda\, f(y_1 + y_n^{a_1}, y_2 + y_n^{a_2}, \dots, y_{n-1} + y_n^{a_{n-1}}, y_n) \quad \in K[y_1, \dots, y_n]$$

is monic in y_n.

Proposition 10.5 (Noether Normalization). *Let R be a finitely generated algebra over a field K, with generators $x_1, \dots, x_n \in R$. Then there is an injective K-algebra homomorphism $K[z_1, \dots, z_r] \to R$ from a polynomial ring over K to R that makes R into a finite extension ring of $K[z_1, \dots, z_r]$.*

Moreover, if K is an infinite field the images of $z_1, \dots, z_r$ in R can be chosen to be K-linear combinations of $x_1, \dots, x_n$.

Proof. We will prove the statement by induction on the number n of generators of R. The case $n = 0$ is trivial, as we can then choose $r = 0$ as well.

So assume now that $n > 0$. We have to distinguish two cases:

(a) There is no algebraic relation among the $x_1, \dots, x_n \in R$, i. e. there is no non-zero polynomial f over K such that $f(x_1, \dots, x_n) = 0$ in R. Then we can choose $r = n$ and the map $K[z_1, \dots, z_n] \to R$ given by $z_i \mapsto x_i$ for all i, which is even an isomorphism in this case.

(b) There is a non-zero polynomial f over K such that $f(x_1, \dots, x_n) = 0$ in R. Then we choose λ and $a_1, \dots, a_{n-1}$ as in Lemma 10.3 (if K is infinite) or Exercise 10.4 (for any K) and set

$$\begin{aligned} &y_1 := x_1 - a_1 x_n,\ \dots,\ y_{n-1} := x_{n-1} - a_{n-1} x_n,\ y_n := x_n \\ &\quad \text{(so that } x_1 = y_1 + a_1 y_n,\ \dots,\ x_{n-1} = y_{n-1} + a_{n-1} y_n,\ x_n = y_n\text{)} \\ \text{or} \quad &y_1 := x_1 - x_n^{a_1},\ \dots,\ y_{n-1} := x_{n-1} - x_n^{a_{n-1}},\ y_n := x_n \\ &\quad \text{(so that } x_1 = y_1 + y_n^{a_1},\ \dots,\ x_{n-1} = y_{n-1} + y_n^{a_{n-1}},\ x_n = y_n\text{)}, \end{aligned}$$

respectively. Note that in both cases these relations show that the K-subalgebra $K[y_1,\dots,y_n]$ of R generated by $y_1,\dots,y_n \in R$ is the same as that generated by $x_1,\dots,x_n$, i. e. all of R. Moreover, y_n is integral over the K-subalgebra $K[y_1,\dots,y_{n-1}]$ of R, since

$$\lambda f(y_1 + a_1 y_n, \dots, y_{n-1} + a_{n-1} y_n, y_n) \quad \text{or} \quad \lambda f(y_1 + y_n^{a_1}, \dots, y_{n-1} + y_n^{a_{n-1}}, y_n),$$

respectively, is monic in y_n and equal to $\lambda f(x_1,\dots,x_n) = 0$. Hence $R = K[y_1,\dots,y_n]$ is finite over $K[y_1,\dots,y_{n-1}]$ by Proposition 9.5. In addition, the subalgebra $K[y_1,\dots,y_{n-1}]$ of R is finite over a polynomial ring $K[z_1,\dots,z_r]$ by the induction hypothesis, and thus R is finite over $K[z_1,\dots,z_r]$ by Lemma 9.6 (a).

Moreover, if K is infinite we can always choose the coordinate transformation of Lemma 10.3, and thus $y_1,\dots,y_n$ (i. e. also the images of $z_1,\dots,z_r$ by induction) are linear combinations of $x_1,\dots,x_n$. □

Remark 10.6. Let $R = A(X)$ be the coordinate ring of a variety X over a field K.

(a) In the Noether Normalization of Proposition 10.5, the (images of) $z_1,\dots,z_r$ in R are algebraically independent functions on X in the sense that there is no polynomial relation among them with coefficients in K. On the other hand, every other element of R is algebraically dependent on $z_1,\dots,z_r$, i. e. it satisfies a (monic) polynomial relation with coefficients in $K[z_1,\dots,z_r]$. We can therefore think of r as the "number of parameters" needed to describe X, i. e. as the "dimension" of X as already mentioned in Example 10.1. In fact, we will see in Remark 11.10 that the number r in Proposition 10.5 is uniquely determined to be the dimension of X in the sense of Chapter 11.

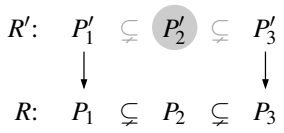
(b) As one would have guessed already from the geometric picture in Example 10.1, the proof of Lemma 10.2 shows that most choices of linear coordinate transformations are suitable to obtain a Noether normalization: in each application of this lemma, only finitely many values of $a_1 \in K$ have to be avoided. Hence we can translate Proposition 10.5 into geometry by saying that a sufficiently general projection to an r-dimensional linear subspace corresponds to a finite ring extension, and hence to a surjective map with finite fibers (where r is the dimension of X as in (a)).

Exercise 10.7. Find a Noether normalization of the $\mathbb{C}$-algebra $\mathbb{C}[x,y,z]/(xy+z^2, x^2y - xy^3 + z^4 - 1)$.

Exercise 10.8. Let $R \subset R'$ be an integral ring extension, and assume that R is a finitely generated algebra over some field K. Moreover, let $P_1 \subsetneq P_3$ be prime ideals in R and $P_1' \subsetneq P_3'$ be prime ideals in R' such that $P_1' \cap R = P_1$ and $P_3' \cap R = P_3$.

(a) Prove: If there is a prime ideal P_2 in R with $P_1 \subsetneq P_2 \subsetneq P_3$, then there is also a prime ideal P_2' in R' with $P_1' \subsetneq P_2' \subsetneq P_3'$.

(b) Can we always find P_2' in (a) such that in addition $P_2' \cap R = P_2$ holds?



As an important application of Noether Normalization we can now give rigorous proofs of some statements in our dictionary between algebra and geometry, namely of the correspondence between (maximal) ideals in the coordinate ring $A(X)$ of a variety X over an algebraically closed field and subvarieties (resp. points) of X. There are various related statements along these lines, and they are all known in the literature by the German name *Hilbert's Nullstellensatz* ("theorem of the zeroes").

Let us start with the simplest instance of this family of propositions. Still very algebraic in nature, it is the statement most closely related to Noether Normalization, from which the geometric results will then follow easily.

Corollary 10.9 (Hilbert's Nullstellensatz, version 1**).** *Let K be a field, and let R be a finitely generated K-algebra which is also a field.*

Then $K \subset R$ is a finite field extension. In particular, if in addition K is algebraically closed then $R = K$.

Proof. By Noether Normalization as in Proposition 10.5 we know that R is finite over a polynomial ring $K[z_1,\dots,z_r]$, and thus also integral over $K[z_1,\dots,z_r]$ by Proposition 9.5. But R is a field, hence $K[z_1,\dots,z_r]$ must be a field as well by Corollary 9.21 (a). This is only the case for $r=0$, and so R is finite over K.

In particular, if K is algebraically closed then there are no algebraic extension fields of K since all zeroes of polynomials over K lie already in K. Hence by Proposition 9.5 there are no finite extensions either in this case, and we must have $R=K$. □

Corollary 10.10 (**Hilbert's Nullstellensatz**, version 2)**.** *Let K be an algebraically closed field. Then all maximal ideals of the polynomial ring $K[x_1,\dots,x_n]$ are of the form*

$$I(a) = (x_1-a_1,\dots,x_n-a_n)$$

for some $a=(a_1,\dots,a_n)\in\mathbb{A}^n_K$.

Proof. Let $P \trianglelefteq K[x_1,\dots,x_n]$ be a maximal ideal. Then $K[x_1,\dots,x_n]/P$ is a field by Lemma 2.3 (b), and in addition certainly a finitely generated K-algebra. Hence $K[x_1,\dots,x_n]/P = K$ by Corollary 10.9, i. e. the natural map $K \to K[x_1,\dots,x_n]/P$, $c \mapsto \overline{c}$ is an isomorphism. Choosing inverse images $a_1,\dots,a_n$ of $\overline{x_1},\dots,\overline{x_n}$ we get $\overline{x_i}=\overline{a_i}$ for all i, and thus $(x_1-a_1,\dots,x_n-a_n)\subset P$. But $(x_1-a_1,\dots,x_n-a_n)$ is a maximal ideal by Example 2.6 (c), and so we must already have $P=(x_1-a_1,\dots,x_n-a_n)=I(a)$. □

Remark 10.11 (Points of a variety correspond to maximal ideals)**.** It is easy to extend Corollary 10.10 to a statement about an arbitrary variety $X\subset\mathbb{A}^n_K$ over an algebraically closed field K: if $R=A(X)$ is the coordinate ring of X we claim that there is a one-to-one correspondence

$$\begin{array}{rcl} \{\text{points of } X\} & \overset{1:1}{\longleftrightarrow} & \{\text{maximal ideals in } A(X)\} \\ a & \longmapsto & I(a) \\ V(I) & \longleftarrow & I. \end{array}$$

In fact, the maximal ideals of $A(X)=K[x_1,\dots,x_n]/I(X)$ are in one-to-one correspondence with maximal ideals $I\trianglelefteq K[x_1,\dots,x_n]$ such that $I\supset I(X)$ by Lemma 1.21 and Corollary 2.4. By Corollary 10.10 this is the same as ideals of the form $I(a)=(x_1-a_1,\dots,x_n-a_n)$ containing $I(X)$. But $I(a)\supset I(X)$ is equivalent to $a\in X$ by Lemma 0.9 (a) and (c), so the result follows.

Remark 10.12 (Zeroes of ideals in $K[x_1,\dots,x_n]$)**.** Another common reformulation of Hilbert's Nullstellensatz is that every proper ideal $I\trianglelefteq K[x_1,\dots,x_n]$ in the polynomial ring over an algebraically closed field K has a zero: by Corollary 2.17 we know that I is contained in a maximal ideal, which must be of the form $I(a)$ by Corollary 10.10. But $I\subset I(a)$ implies $a\in V(I)$ by Lemma 0.9 (a) and (c), and hence $V(I)\neq\emptyset$.

Note that this statement is clearly false over fields that are not algebraically closed, as e. g. (x^2+1) is a proper ideal in $\mathbb{R}[x]$ with empty zero locus in $\mathbb{A}^1_{\mathbb{R}}$.

19

In order to extend the correspondence between points and maximal ideals to arbitrary subvarieties we need another algebraic preliminary result first: recall that in any ring R the radical of an ideal I equals the intersection of all prime ideals containing I by Lemma 2.21. We will show now that it is in fact sufficient to intersect all maximal ideals containing I if R is a finitely generated algebra over a field.

Corollary 10.13 (**Hilbert's Nullstellensatz**, version 3)**.** *For every ideal I in a finitely generated algebra R over a field K we have*

$$\sqrt{I} = \bigcap_{\substack{P \text{ maximal} \\ P\supset I}} P.$$

Proof. The inclusion "$\subset$" follows immediately from Lemma 2.21, since every maximal ideal is prime.

For the opposite inclusion "$\supset$", let $f \in R$ with $f \notin \sqrt{I}$; we have to find a maximal ideal $P \supset I$ with $f \notin P$. Consider the multiplicatively closed set $S = \{f^n : n \in \mathbb{N}\}$. As $f \notin \sqrt{I}$ implies $I \cap S = \emptyset$, we get by Exercise 6.14 (a) a prime ideal $P \trianglelefteq R$ with $P \supset I$ and $P \cap S = \emptyset$, in particular with $f \notin P$. Moreover, we can assume by Exercise 6.14 (a) that $S^{-1}P$ is maximal. It only remains to show that P is maximal.

To do this, consider the ring extension $K \to R/P \to (R/P)_f = R_f/P_f$, where the subscript f denotes localization at S as in Example 6.5 (c). Note that the second map is in fact an inclusion since R/P is an integral domain, and the stated equality holds by Corollary 6.22 (b). Moreover, R_f/P_f is a field since P_f is maximal, and finitely generated as a K-algebra (as generators we can choose the classes of generators for R together with $\frac{1}{f}$). So $K \subset R_f/P_f$ is a finite field extension by Corollary 10.9, and hence integral by Proposition 9.5. But then $R/P \subset R_f/P_f$ is integral as well, which means by Corollary 9.21 (a) that R/P is a field since R_f/P_f is. Hence P is maximal by Lemma 2.3 (b). □

Corollary 10.14 (Hilbert's Nullstellensatz, version 4**).** *Let $X \subset \mathbb{A}^n_K$ be a variety over an algebraically closed field K. Then for every ideal $I \trianglelefteq A(X)$ we have $I(V(I)) = \sqrt{I}$.*

In particular, there is a one-to-one correspondence

$$\begin{array}{rcl} \{\textit{subvarieties of } X\} & \overset{1:1}{\longleftrightarrow} & \{\textit{radical ideals in } A(X)\} \\ Y & \longmapsto & I(Y) \\ V(I) & \longmapsfrom & I. \end{array}$$

Proof. Let us first prove the equality $I(V(I)) = \sqrt{I}$.

"$\subset$": Assume that $f \notin \sqrt{I}$. By Corollary 10.13 there is then a maximal ideal $P \trianglelefteq A(X)$ with $P \supset I$ and $f \notin P$. But by Remark 10.11 this maximal ideal has to be of the form $I(a) = (x_1 - a_1, \ldots, x_n - a_n)$ for some point $a \in X$. Now $I(a) \supset I$ implies $a \in V(I)$ by Lemma 0.9 (a) and (c), and $f \notin I(a)$ means $f(a) \neq 0$. Hence $f \notin I(V(I))$.

"$\supset$": Let $f \in \sqrt{I}$, i. e. $f^n \in I$ for some $n \in \mathbb{N}$. Then $(f(a))^n = 0$, and hence $f(a) = 0$, for all $a \in V(I)$. This means that $f \in I(V(I))$.

The one-to-one correspondence now follows immediately from what we already know: the two maps are well-defined since $I(Y)$ is always radical by Remark 1.10, and they are inverse to each other by Lemma 0.9 (c) and the statement $I(V(I)) = \sqrt{I}$ proven above. □

11. Dimension

We have already met several situations in this course in which it seemed to be desirable to have a notion of *dimension* (of a variety, or more generally of a ring): for example, in the geometric interpretation of the Hilbert Basis Theorem in Remark 7.15, or of the Noether Normalization in Remark 10.6. We have also used the term "curve" several times already to refer to a "one-dimensional variety", and even if we have not defined this rigorously yet it should be intuitively clear what this means. But although this concept of dimension is very intuitive and useful, it is unfortunately also one of the most complicated subjects in commutative algebra when it comes to actually proving theorems. We have now developed enough tools however to be able to study some basic dimension theory in this chapter without too many complications. Let us start with the definition of dimension, whose idea is similar to that of the length of a module in Definition 3.18.

Definition 11.1 (Dimension)**.** Let R be a ring.

(a) The **(Krull) dimension** $\dim R$ of R is the maximum number $n \in \mathbb{N}$ such that there is a chain of prime ideals
$$P_0 \subsetneq P_1 \subsetneq \cdots \subsetneq P_n$$
of length n in R.

In order to distinguish this notion of dimension from that of a K-vector space V, we will write the latter always as $\dim_K V$ in these notes.

(b) The **codimension** or **height** of a prime ideal P in R is the maximum number $n \in \mathbb{N}$ such that there is a chain as in (a) with $P_n \subset P$. We denote it by $\operatorname{codim}_R P$, or just $\operatorname{codim} P$ if the ring is clear from the context.

(c) The **dimension** $\dim X$ of a variety X is defined to be the dimension of its coordinate ring $A(X)$. A 1-dimensional variety is called a **curve**. The **codimension** of an irreducible subvariety Y in X is defined to be the codimension of the prime ideal $I(Y)$ in $A(X)$ (see Remark 2.7 (b)); we denote it by $\operatorname{codim}_X Y$ or just $\operatorname{codim} Y$.

In all cases, we set the dimension resp. codimension formally to ∞ if there is no bound on the length of the chains considered. Also, note that in all cases there is at least one such chain (by Corollary 2.17 in (a), and the trivial length-0 chain with $P_0 = P$ in (b)), hence the definitions above make sense.

Remark 11.2 (Geometric interpretation of dimension)**.** Let X be a variety over an algebraically closed field. By Remark 2.7 (b) and Corollary 10.14, the prime ideals in the coordinate ring $A(X)$ are in one-to-one correspondence with non-empty irreducible subvarieties of X. As this correspondence reverses inclusions, Definition 11.1 says that the dimension of X — or equivalently of its coordinate ring $A(X)$ — is equal to the biggest length n of a chain
$$X_0 \supsetneq X_1 \supsetneq \cdots \supsetneq X_n \neq \emptyset$$
of irreducible subvarieties of X. Now as in Remark 7.15 the geometric idea behind this definition is that making an *irreducible* subvariety smaller is only possible by reducing its dimension, so that in a maximal chain as above the dimension of X_i should be $\dim X - i$, with X_n being a point and X_0 an irreducible component of X.

Similarly, the codimension of an irreducible subvariety Y in X is the biggest length n of a chain
$$X_0 \supsetneq X_1 \supsetneq \cdots \supsetneq X_n \supset Y.$$
Again, for a maximal chain X_0 should be an irreducible component of X, and the dimension should drop by 1 in each inclusion in the chain. Moreover, we will have $X_n = Y$ in a maximal chain, so that we can think of n as $\dim X - \dim Y$, and hence as what one would expect geometrically to be the codimension of Y in X (see Example 11.13 (a)).

Example 11.3.

(a) Every field has dimension 0, since the zero ideal is the only prime ideal in this case.

(b) More generally, the dimension of a ring R is 0 if and only if there are no strict inclusions among prime ideals of R, i. e. if and only if all prime ideals are already maximal. So we can e. g. rephrase the theorem of Hopkins in Proposition 7.17 as

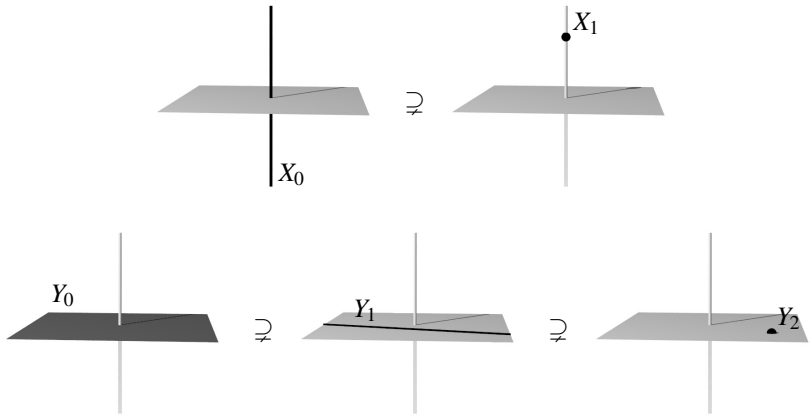
$$R \text{ is Artinian} \quad \Leftrightarrow \quad R \text{ is Noetherian and } \dim R = 0.$$

Note that this fits well with the geometric interpretation of Remark 11.2, since we have already seen in Remark 7.15 that Artinian rings correspond to finite unions of points.

(c) Let R be a principal ideal domain which is not a field. Then except for the zero ideal (which is prime but not maximal) the notions of prime and maximal ideals agree by Example 2.6 (b). So the longest chains of prime ideals in R are all of the form $0 \subsetneq P$ for a maximal ideal P. It follows that $\dim R = 1$.

In particular, the ring $\mathbb{Z}$ and polynomial rings $K[x]$ over a field K have dimension 1. Geometrically, this means that the variety $\mathbb{A}^1_K$ (with coordinate ring $K[x]$) has dimension 1.

Remark 11.4 (Maximal chains of prime ideals can have different lengths)**.** One of the main obstacles when dealing with dimensions is that, in general, maximal chains of prime ideals as in Definition 11.1 (in the sense that they cannot be extended to a longer chain by inserting more prime ideals) do not necessarily all have the same length. This is easy to see in geometry, where the corresponding statement is roughly that a space can be made up of components of different dimension. Consider e. g. the union $X = V(x_1x_3, x_2x_3) \subset \mathbb{A}^3_{\mathbb{R}}$ of a line and a plane as in Example 0.4 (e), and the following two chains of irreducible subvarieties in X of lengths 1 and 2, respectively.



The first chain $X_0 \supsetneq X_1$ is maximal by Example 11.3 (c), since the line has dimension 1. Nevertheless, the second chain is longer (and in fact also maximal, since the plane has dimension 2 as we will see in Proposition 11.9). Hence, due to the two components of X (of different dimension) the maximal chains in X have different lengths. Similarly, the same chains above show that the codimension of the point X_1 is 1, whereas the codimension of the point Y_2 is 2.

Let us state a few properties of dimension that are immediately obvious from the definition.

Remark 11.5 (First properties of dimension)**.** Let R be a ring.

(a) For any prime ideal $P \trianglelefteq R$, the prime ideals of R contained in P are in one-to-one correspondence with prime ideals in the localization R_P by Example 6.8. In other words, we always have $\operatorname{codim} P = \dim R_P$.

(b) Again let $P \trianglelefteq R$ be a prime ideal, and set $n = \dim R/P$ and $m = \operatorname{codim} P = \dim R_P$. Then by Lemma 1.21 there are chains of prime ideals

$$P_0 \subsetneq \cdots \subsetneq P_m \subset P \quad \text{and} \quad P \subset Q_0 \subsetneq \cdots \subsetneq Q_n$$

in R that can obviously be glued to a single chain of length $m+n$. Hence we conclude that

$$\dim R \geq \dim R/P + \operatorname{codim} P.$$

Geometrically, this means for an irreducible subvariety Y of a variety X that

$$\dim X \geq \dim Y + \operatorname{codim}_X Y,$$

since $A(Y) \cong A(X)/I(Y)$ by Lemma 0.9 (d). Note that we do not have equality in general, since e. g. $\dim X_1 = 0$ and $\operatorname{codim} X_1 = 1$ in the example of Remark 11.4.

(c) Dimension is a "local concept": we claim that

$$\dim R = \sup\{\dim R_P : P \text{ maximal ideal of } R\} = \sup\{\operatorname{codim} P : P \text{ maximal ideal of } R\}.$$

In fact, if $P_0 \subsetneq \cdots \subsetneq P_n$ is a chain of prime ideals in R then by Example 6.8 the corresponding localized chain is a chain of primes ideals of the same length in R_P, where P is any maximal ideal containing P_n. Conversely, any chain of prime ideals in a localization R_P corresponds to a chain of prime ideals (contained in P) of the same length in R.

Geometrically, we can think of this as the statement that the dimension of a variety is the maximum of the "local dimensions" at every point — so that e. g. the union of a line and a plane in Remark 11.4 has dimension 2.

In the favorable case when all maximal chains of prime ideals do have the same length, the properties of Remark 11.5 hold in a slightly stronger version. We will see in Corollary 11.12 that this always happens e. g. for coordinate rings of irreducible varieties. In this case, the dimension and codimension of a subvariety always add up to the dimension of the ambient variety, and the local dimension is the same at all points.

Lemma 11.6. *Let R be a ring of finite dimension in which all maximal chains of prime ideals have the same length. Moreover, let $P \trianglelefteq R$ be a prime ideal. Then:*

(a) *The quotient R/P is also a ring of finite dimension in which all maximal chains of prime ideals have the same length;*

(b) $\dim R = \dim R/P + \operatorname{codim} P$*;*

(c) $\dim R_P = \dim R$ *if P is maximal.*

Proof. By Lemma 1.21 and Corollary 2.4, a chain of prime ideals in R/P corresponds to a chain $Q_0 \subsetneq \cdots \subsetneq Q_r$ of prime ideals in R that contain P. In particular, the length of such chains is bounded by $\dim R < \infty$, and hence $\dim R/P < \infty$ as well. Moreover, if the chain is maximal we must have $Q_0 = P$, and thus we can extend it to a maximal chain

$$P_0 \subsetneq \cdots \subsetneq P_m = P = Q_0 \subsetneq \cdots \subsetneq Q_r$$

of prime ideals in R that includes P. The chains $P_0 \subsetneq \cdots \subsetneq P_m$ and $Q_0 \subsetneq \cdots \subsetneq Q_r$ then mean that $\operatorname{codim} P \geq m$ and $\dim R/P \geq r$. Moreover, we have $m + r = \dim R$ by assumption. Hence

$$\dim R \geq \dim R/P + \operatorname{codim} P \geq \dim R/P + m \geq r + m = \dim R$$

by Remark 11.5 (b), and so equality holds. This shows $r = \dim R/P$ and hence (a), and of course also (b). The statement (c) follows from (b), since for maximal P we have $\dim R/P = 0$ by Example 11.3 (a), and $\dim R_P = \operatorname{codim} P$ by Remark 11.5 (a). □

20

Exercise 11.7. Let $I = Q_1 \cap \cdots \cap Q_n$ be a primary decomposition of an ideal I in a Noetherian ring R. Show that

$$\dim R/I = \max\{\dim R/P : P \text{ is an isolated prime ideal of } I\}.$$

What is the geometric interpretation of this statement?

Let us now show that, in a Noether normalization $K[z_1, \ldots, z_r] \to R$ of a finitely generated algebra R over a field K as in Proposition 10.5, the number r is uniquely determined to be $\dim R$ — as already motivated in Remark 10.6. More precisely, this will follow from the following two geometrically intuitive facts:

- Integral extensions preserve dimension: by Example 9.19, they correspond to surjective maps of varieties with finite fibers, and thus the dimension of the source and target of the map should be the same.
- The dimension of the polynomial ring $K[x_1,\dots,x_n]$ (and thus of the variety $\mathbb{A}^n_K$) is n.

We will start with the proof of the first of these statements.

Lemma 11.8 (Invariance of dimension under integral extensions)**.** *For any integral ring extension $R \subset R'$ we have* $\dim R = \dim R'$.

Proof.

"$\leq$" Let $P_0 \subsetneq \cdots \subsetneq P_n$ be a chain of prime ideals in R. By Lying Over (for P_0) and Going Up (successively for $P_1,\dots,P_n$) as in Propositions 9.18 and 9.24 we can find a corresponding chain $P'_0 \subsetneq \cdots \subsetneq P'_n$ of the same length in R' (where the inclusions have to be strict again since $P'_i \cap R = P_i$ for all i).

"$\geq$" Now let $P'_0 \subsetneq \cdots \subsetneq P'_n$ be a chain of prime ideals in R'. Intersecting with R we get a chain of prime ideals $P_0 \subsetneq \cdots \subsetneq P_n$ in R by Exercise 2.9 (b), where the inclusions are strict again by Incomparability as in Proposition 9.20. □

For the statement that $\dim K[x_1,\dots,x_n] = n$ we can actually right away prove the stronger result that every chain of prime ideals has length n, and thus e. g. by Lemma 11.6 (c) that the local dimension of $K[x_1,\dots,x_n]$ is the same at all maximal ideals. This is not too surprising if K is algebraically closed, since then all maximal ideals of $K[x_1,\dots,x_n]$ are of the form $(x_1 - a_1,\dots,x_n - a_n)$ by Corollary 10.10, and thus can all be obtained from each other by translations. But for general fields there will be more maximal ideals in the polynomial ring, and thus the statement that all such localizations have the same dimension is far less obvious.

Proposition 11.9 (Dimension of polynomial rings)**.** *Let K be a field, and let $n \in \mathbb{N}$.*

(a) $\dim K[x_1,\dots,x_n] = n$.

(b) *All maximal chains of prime ideals in $K[x_1,\dots,x_n]$ have length n.*

Proof. We will prove both statements by induction on n, with the case $n = 0$ being obvious. (In fact, we also know the statement for $n = 1$ already by Example 11.3 (c)).

So let $n \geq 1$, and let $P_0 \subsetneq \cdots \subsetneq P_m$ be a chain of prime ideals in $K[x_1,\dots,x_n]$. We have to show that $m \leq n$, and that equality always holds for a maximal chain. By possibly extending the chain we may assume without loss of generality that $P_0 = 0$, P_1 is a minimal non-zero prime ideal, and P_m is a maximal ideal. Then $P_1 = (f)$ for some non-zero polynomial f by Exercise 8.32 (b), since $K[x_1,\dots,x_n]$ is a unique factorization domain by Remark 8.6.

By a change of coordinates as in the proof of the Noether normalization in Proposition 10.5 we can also assume without loss of generality that f is monic in x_n, and hence that $K[x_1,\dots,x_n]/P_1 = K[x_1,\dots,x_n]/(f)$ is integral over $K[x_1,\dots,x_{n-1}]$ by Proposition 9.5. We can now transfer our chain of prime ideals from $K[x_1,\dots,x_n]$ to $K[x_1,\dots,x_{n-1}]$ as in the diagram below: after dropping the first prime ideal P_0, first extend the chain by the quotient map $K[x_1,\dots,x_n] \to K[x_1,\dots,x_n]/P_1$, and then contract it by the integral ring extension $K[x_1,\dots,x_{n-1}] \to K[x_1,\dots,x_n]/P_1$.

$$\begin{array}{rccccc}
K[x_1,\dots,x_n]: & P_1 & \subsetneq & \cdots & \subsetneq & P_m \\
& \big\downarrow \text{extension} & & & & \big\downarrow \text{extension} \\
K[x_1,\dots,x_n]/P_1: & P_1/P_1 & \subsetneq & \cdots & \subsetneq & P_m/P_1 \\
& \big\downarrow \text{contraction} & & & & \big\downarrow \text{contraction} \\
K[x_1,\dots,x_{n-1}]: & (P_1/P_1)\cap K[x_1,\dots,x_{n-1}] & \subsetneq & \cdots & \subsetneq & (P_m/P_1)\cap K[x_1,\dots,x_{n-1}]
\end{array}$$

We claim that both these steps preserve prime ideals and their strict inclusions, and transfer maximal chains to maximal chains. In fact, for the extension along the quotient map this follows from the one-to-one correspondence between prime ideals in rings and their quotients as in Lemma 1.21 and Corollary 2.4. The contraction then maps prime ideals to prime ideals by Exercise 2.9 (b), and keeps the strict inclusions by the Incomparability property of Proposition 9.20. Moreover, it also preserves maximal chains: in the bottom chain of the above diagram the first entry is 0, the last one is a maximal ideal by Corollary 9.21 (b), and if we could insert another prime ideal at any step in the chain we could do the same in the middle chain as well by Exercise 10.8 (a).

Now by the induction hypothesis the length $m-1$ of the bottom chain is at most $n-1$, and equal to $n-1$ if the chain is maximal. Hence we always have $m \leq n$, and $m = n$ if the original chain was maximal. □

Remark 11.10 (Noether normalization and dimension)**.** Let R be a finitely generated algebra over a field K, and let $K[z_1,\dots,z_n] \to R$ be a Noether normalization as in Proposition 10.5. Using Lemma 11.8 and Proposition 11.9 we now see rigorously that then the number n is uniquely determined to be $n = \dim K[z_1,\dots,z_n] = \dim R$, as already expected in Remark 10.6. In particular, it follows that a finitely generated algebra over a field (and hence a variety) is always of finite dimension — a statement that is not obvious from the definitions! In fact, the following exercise shows that the finite dimension and Noetherian conditions are unrelated, although rings that meet one of these properties but not the other do not occur very often in practice.

Exercise 11.11 (Finite dimension $\not\Longleftrightarrow$ noetherian)**.** Give an example of a non-Noetherian ring of finite dimension.

In fact, there are also Noetherian rings which do not have finite dimension, but these are hard to construct [E, Exercise 9.6].

Corollary 11.12. *Let R be a finitely generated algebra over a field K, and assume that R is an integral domain. Then every maximal chain of prime ideals in R has length* $\dim R$.

Proof. By Lemma 1.30 and Lemma 2.3 (a) we can write $R = K[x_1,\dots,x_n]/P$ for a prime ideal P in a polynomial ring $K[x_1,\dots,x_n]$. Thus the statement follows from Proposition 11.9 (b) with Lemma 11.6 (a). □

Example 11.13. Let Y be an irreducible subvariety of an irreducible variety X. Then by Corollary 11.12 every maximal chain of prime ideals in $A(X)$ has the same length, and consequently Lemma 11.6 for $R = A(X)$ implies that

(a) $\dim X = \dim Y + \operatorname{codim}_X Y$ for every irreducible subvariety Y of X (since then $P = I(Y)$ is a prime ideal and $R/P \cong A(Y)$);

(b) the local dimension of X is $\dim X$ at every point, i. e. all localizations of $A(X)$ at maximal ideals have dimension $\dim X$.

Next, let us study how the codimension of a prime ideal is related to the number of generators of the ideal. Geometrically, one would expect that an irreducible subvariety given by n equations has codimension at most n, with equality holding if the equations are "independent" in a suitable sense. More generally, if the zero locus of the given equations is not irreducible, each of its irreducible components should have codimension at most n.

Algebraically, if $I = (a_1,\dots,a_n)$ is an ideal in a coordinate ring generated by n elements, the irreducible components of $V(I)$ are the maximal irreducible subvarieties of $V(I)$ and thus correspond to the minimal prime ideals over I as in Exercise 2.23. So our geometric idea above leads us to the expectation that a minimal prime ideal over an ideal generated by n elements should have codimension at most n.

We will now prove this statement by induction on n for any Noetherian ring. For the proof we need the following construction of the so-called symbolic powers $P^{(k)}$ of a prime ideal P in a ring R. Their behavior is very similar to that of the ordinary powers P^k; in fact the two notions agree after

localization at P as we will see in the next lemma. The main advantage of the symbolic powers is that they are always primary (in contrast to the ordinary powers, see Example 8.13).

Lemma 11.14 (Symbolic Powers)**.** *Let R be a ring. For a prime ideal $P \trianglelefteq R$ and $n \in \mathbb{N}$ consider the ideal*

$$P^{(n)} := \{a \in R : ab \in P^n \text{ for some } b \in R \backslash P\}$$

called the n-th symbolic power of P. Then:

(a) $P^n \subset P^{(n)} \subset P$;

(b) $P^{(n)}$ *is P-primary;*

(c) $P^{(n)} R_P = P^n R_P$.

Proof.

(a) The first inclusion is obvious (take $b = 1$). For the second, let $a \in P^{(n)}$, hence $ab \in P^n \subset P$ for some $b \in R \backslash P$. But then $a \in P$ since P is prime.

(b) Taking radicals in (a) we see that $P = \sqrt{P} = \sqrt{P^n} \subset \sqrt{P^{(n)}} \subset \sqrt{P} = P$, so $\sqrt{P^{(n)}} = P$. To see that $P^{(n)}$ is P-primary, let $ab \in P^{(n)}$, i. e. $abc \in P^n$ for some $c \in R \backslash P$. Then if $b \notin \sqrt{P^{(n)}} = P$ we also have $bc \notin P$ and thus by definition $a \in P^{(n)}$. Hence $P^{(n)}$ is P-primary.

(c) For the inclusion "$\subset$", let $\frac{b}{s} \in P^{(n)} R_P$, i. e. $bc \in P^n$ for some $s, c \in R \backslash P$. Then $\frac{b}{s} = \frac{bc}{sc} \in P^n R_P$. The other inclusions is obvious since $P^{(n)} \supset P^n$. □

Proposition 11.15 (**Krull's Principal Ideal Theorem**)**.** *Let R be a Noetherian ring, and let $a \in R$. Then every minimal prime ideal P over (a) satisfies* $\operatorname{codim} P \leq 1$.

Proof. Let $Q' \subset Q \subsetneq P$ be a chain of prime ideals in R; we have to prove that $Q' = Q$. By the one-to-one correspondence of Example 6.8 between prime ideals in R and in its quotients resp. localizations we can take the quotient by Q' and localize at P and prove the statement in the resulting ring, which we will again call R for simplicity. Note that this new ring is also still Noetherian by Remark 7.8 (b) and Exercise 7.23. In this new situation we then have:

- by taking the quotient we achieved that $Q' = 0$ and R is an integral domain;
- by localizing we achieved that R is local, with unique maximal ideal P;
- we have to prove that $Q = 0$.

Let us now consider the symbolic powers $Q^{(n)}$ of Lemma 11.14. Obviously, we have $Q^{(n+1)} \subset Q^{(n)}$ for all n. Moreover:

(a) $Q^{(n)} \subset Q^{(n+1)} + (a)$ for some n: The ring $R/(a)$ is Noetherian by Remark 7.8 (b) and of dimension 0 since the unique maximal ideal $P/(a)$ of $R/(a)$ is also minimal by assumption. Hence $R/(a)$ is Artinian by Hopkins as in Example 11.3 (b). This means that the descending chain

$$(Q^{(0)} + (a))/(a) \supsetneq (Q^{(1)} + (a))/(a) \supsetneq \cdots$$

of ideals in $R/(a)$ becomes stationary. Hence $Q^{(n)} + (a) = Q^{(n+1)} + (a)$ for some n, which implies $Q^{(n)} \subset Q^{(n+1)} + (a)$.

(b) $Q^{(n)} = Q^{(n+1)} + P Q^{(n)}$: The inclusion "$\supset$" is clear, so let us prove "$\subset$". If $b \in Q^{(n)}$ then $b = c + ar$ for some $c \in Q^{(n+1)}$ and $r \in R$ by (a). So $ar = b - c \in Q^{(n)}$. But $a \notin Q$ (otherwise P would not be minimal over (a)) and $Q^{(n)}$ is Q-primary by Lemma 11.14 (b), and hence $r \in Q^{(n)}$. But this means that $b = c + ar \in Q^{(n+1)} + P Q^{(n)}$.

21

Taking the quotient by $Q^{(n+1)}$ in the equation of (b) we see that $Q^{(n)}/Q^{(n+1)} = P Q^{(n)}/Q^{(n+1)}$, and hence that $Q^{(n)}/Q^{(n+1)} = 0$ by Nakayama's Lemma as in Exercise 6.16 (a). This means that $Q^{(n)} = Q^{(n+1)}$. By Lemma 11.14 (c), localization at Q now gives $Q^n R_Q = Q^{n+1} R_Q$, hence $Q^n R_Q = (Q R_Q) Q^n R_Q$ by Exercise 1.19 (c), and so $Q^n R_Q = 0$ by Exercise 6.16 (a) again. But as R is an integral domain, this is only possible if $Q = 0$. □

Exercise 11.16. Let $n \in \mathbb{N}_{>0}$, and let $P_0 \subsetneq P_1 \subsetneq \cdots \subsetneq P_n$ be a chain of prime ideals in a Noetherian ring R. Moreover, let $a \in P_n$. Prove:

(a) There is a chain of prime ideals $P_0' \subsetneq P_1' \subsetneq \cdots \subsetneq P_{n-1}' \subsetneq P_n$ (i. e. in the given chain we may change all prime ideals except the last one) such that $a \in P_1'$.

(b) There is in general no such chain with $a \in P_0'$.

Corollary 11.17. *Let R be a Noetherian ring, and let $a_1, \ldots, a_n \in R$. Then every minimal prime ideal P over $(a_1, \ldots, a_n)$ satisfies* $\operatorname{codim} P \leq n$.

Proof. We will prove the statement by induction on n; the case $n = 1$ is Proposition 11.15. So let $n \geq 2$, and let $P_0 \subsetneq \cdots \subsetneq P_s$ be a chain of prime ideals in P. After possibly changing some of these prime ideals (except the last one), we may assume by Exercise 11.16 (a) that $a_n \in P_1$. But then

$$P_1/(a_n) \subsetneq \cdots \subsetneq P_s/(a_n)$$

is a chain of prime ideals of length $s-1$ in $P/(a_n)$. As $P/(a_n)$ is minimal over $(\overline{a_1}, \ldots, \overline{a_{n-1}})$, we now know by induction that $s-1 \leq \operatorname{codim} P/(a_n) \leq n-1$, and hence that $s \leq n$. As the given chain was arbitrary, this means that $\operatorname{codim} P \leq n$. □

Remark 11.18 (Geometric interpretation of Krull's Principal Ideal Theorem)**.** Let R be the coordinate ring of a variety $X \subset \mathbb{A}^n_K$ over K, and let $I = (f_1, \ldots, f_r) \trianglelefteq R$ be an ideal generated by r elements. As mentioned above, the minimal prime ideals over I correspond to the irreducible components of $V(I)$. Hence, Corollary 11.17 states geometrically that every irreducible component of $V(I)$ has codimension at most r, and thus by Example 11.13 (a) dimension at least $n-r$.

For the case of a principal ideal (a), we can even give an easy criterion for when equality holds in Proposition 11.15. Geometrically, intersecting a variety X with the zero locus of a polynomial f reduces the dimension if the polynomial does not vanish identically on any irreducible component of X, i. e. if f is not a zero divisor in $A(X)$:

Corollary 11.19. *Let R be a Noetherian ring, and let $a \in R$ not be a zero-divisor. Then for every minimal prime ideal P over (a) we have* $\operatorname{codim} P = 1$.

Proof. Let $P_1, \ldots, P_n$ be the minimal prime ideals over the zero ideal as in Corollary 8.30. By Proposition 8.27 they can be written as $P_i = \sqrt{0 : b_i}$ for some non-zero $b_1, \ldots, b_n \in R$.

We claim that $a \notin P_i$ for all $i = 1, \ldots, n$. In fact, if $a \in P_i$ for some i then $a \in \sqrt{0 : b_i}$, and hence $a^r b_i = 0$ for some $r \in \mathbb{N}_{>0}$. But if we choose r minimal then $a \cdot (a^{r-1} b_i) = 0$ although $a^{r-1} b_i \neq 0$, i. e. a would have to be a zero-divisor in contradiction to our assumption.

So $a \notin P_i$ for all i. But on the other hand $a \in P$, and hence P cannot be any of the minimal prime ideals $P_1, \ldots, P_n$ of R. So P must strictly contain one of the $P_1, \ldots, P_n$, which means that $\operatorname{codim} P \geq 1$. The equality $\operatorname{codim} P = 1$ now follows from Proposition 11.15. □

Example 11.20. Let $X = V(x^2 + y^2 - 1) \subset \mathbb{A}^2_{\mathbb{R}}$ be the unit circle. As its coordinate ring is $R = \mathbb{R}[x,y]/(x^2 + y^2 - 1)$ and the ideal $(x^2 + y^2 - 1)$ is prime, we get as expected that X has dimension

$$\begin{aligned} \dim X = \dim R &= \dim \mathbb{R}[x,y] - \operatorname{codim}(x^2 + y^2 - 1) && \text{(Lemma 11.6 (b) and Proposition 11.9 (b))} \\ &= 2 - 1 && \text{(Proposition 11.9 (a) and Corollary 11.19)} \\ &= 1. \end{aligned}$$

With an analogous computation, note that the ring $\mathbb{R}[x,y]/(x^2 + y^2 + 1)$ has dimension 1 as well, although $V(x^2 + y^2 + 1) = \emptyset$ in $\mathbb{A}^2_{\mathbb{R}}$.

Exercise 11.21. Compute the dimension and all maximal ideals of $\mathbb{C}[x,y,z]/(x^2 - y^2, z^2 x - z^2 y)$.

Exercise 11.22. Show:

(a) Every maximal ideal in the polynomial ring $\mathbb{Z}[x]$ is of the form (p, f) for some prime number $p \in \mathbb{Z}$ and a polynomial $f \in \mathbb{Z}[x]$ whose class in $\mathbb{Z}_p[x] = \mathbb{Z}[x]/(p)$ is irreducible.

(b) $\dim \mathbb{Z}[x] = 2$.

Exercise 11.23. For each of the two ring extensions $R \subset R'$

$$\text{(i) } \mathbb{R}[y] \subset \mathbb{R}[x,y]/(x^2-y) \qquad \text{(ii) } \mathbb{Z} \subset \mathbb{Z}[x]/(x^2-3)$$

find:

(a) a non-zero prime ideal $P \trianglelefteq R$ with a unique prime ideal $P' \trianglelefteq R'$ lying over P, and $R'/P' \cong R/P$;

(b) a non-zero prime ideal $P \trianglelefteq R$ with a unique prime ideal $P' \trianglelefteq R'$ lying over P, and $R'/P' \ncong R/P$;

(c) a non-zero prime ideal $P \trianglelefteq R$ such that there is more than one prime ideal in R' lying over P.

(In these three cases the prime ideal P is then called *ramified*, *inert*, and *split*, respectively. Can you see the reason for these names?)

Exercise 11.24. Let $R = \mathbb{C}[x,y,z,w]/I$ with $I = (x,y) \cap (z,w)$, and let $a = \overline{x+y+z+w} \in R$. By Krull's Principal Ideal Theorem we know that every isolated prime ideal of (a) in R has codimension at most 1. However, show now that (a) has an embedded prime ideal of codimension 2.

We have now studied the Krull dimension of rings as in Definition 11.1 in some detail. In the rest of this chapter, we want to discuss two more approaches to define the notion of dimension. Both of them will turn out to give the same result as the Krull dimension in many (but not in all) cases.

The first construction is based on the idea that the dimension of a variety X over a field K can be thought of as the number of independent variables needed to describe its points. In more algebraic terms, this means that $\dim X$ should be the maximum number of functions on X that are "algebraically independent" in the sense that they do not satisfy a polynomial relation with coefficients in K. This notion of algebraic independence is best-behaved for extensions of fields, and so we will restrict ourselves to irreducible varieties: in this case $A(X)$ is a finitely generated K-algebra that is an integral domain, and hence we can consider its quotient field $\operatorname{Quot} R$ as an extension field of K (i. e. consider rational instead of polynomial functions on X).

Definition 11.25 (Algebraic dependence, transcendence bases)**.** Let $K \subset L$ be a field extension, and let B be a subset of L. As usual, we denote by $K(B)$ the smallest subfield of L containing K and B [G3, Definition 1.13].

(a) The subset B is called **algebraically dependent** over K if there is a non-zero polynomial $f \in K[x_1,\dots,x_n]$ with $f(b_1,\dots,b_n) = 0$ for some distinct $b_1,\dots,b_n \in B$. Otherwise B is called **algebraically independent**.

(b) The subset B is called a **transcendence basis** of L over K if it is algebraically independent and L is algebraic over $K(B)$.

Example 11.26.

(a) Let $K \subset L$ be a field extension, and let $a \in L$. Then $\{a\}$ is algebraically dependent over K if and only if a is algebraic over K.

(b) Let $R = K[x_1,\dots,x_n]$ be the polynomial ring in n variables over a field K, and let

$$L = \operatorname{Quot} R = \left\{ \frac{f}{g} : f,g \in K[x_1,\dots,x_n], g \neq 0 \right\}$$

be its quotient field, i. e. the field of formal rational functions in n variables over K.

This field is usually denoted by $K(x_1,\dots,x_n)$, and thus by the same round bracket notation as for the smallest extension field of K containing given elements of a larger field — which can be described explicitly as the set of all rational functions in these elements with coefficients in K. Note that this ambiguity is completely analogous to the square bracket notation $K[x_1,\dots,x_n]$ which is also used to denote both polynomial expressions in formal variables (to obtain the polynomial ring) or in given elements of an extension ring (to form the subalgebra generated by these elements).

The set $B = \{x_1, \dots, x_n\}$ is clearly a transcendence basis of L over K: these elements are obviously algebraically independent by definition, and the smallest subfield of L containing K and $x_1, \dots, x_n$ is just L itself.

Remark 11.27. In contrast to what one might expect, the definition of a transcendence basis B of a field extension $K \subset L$ does not state that L is generated by B in some sense — it is only generated by B up to algebraic extensions. For example, the empty set is a transcendence basis of every algebraic field extension.

It is true however that a transcendence basis is a "maximal algebraically independent subset": any element $a \in L \backslash B$ is algebraic over $K(B)$, and thus $B \cup \{a\}$ is algebraically dependent.

As expected, we can now show that transcendence bases always exist, and that all transcendence bases of a given field extension have the same number of elements.

Lemma 11.28 (Existence of transcendence bases)**.** *Let $K \subset L$ be a field extension. Moreover, let $A \subset L$ be algebraically independent over K, and let $C \subset L$ be a subset such that L is algebraic over $K(C)$. Then A can be extended by a suitable subset of C to a transcendence basis of L over K.*

In particular, every field extension has a transcendence basis.

Proof. Let

$$M = \{B : B \text{ is algebraically independent over } K \text{ with } A \subset B \subset A \cup C\}.$$

Then every totally ordered subset $N \subset M$ has an upper bound B: if $N = \emptyset$ we can take $A \in M$, otherwise $\bigcup_{B \in N} B$. Note that the latter is indeed an element of M:

- Assume that $\bigcup_{B \in N} B$ is algebraically dependent. Then $f(b_1, \dots, b_n) = 0$ for a non-zero polynomial f over K and distinct elements $b_1, \dots, b_n \in \bigcup_{B \in N} B$. But as N is totally ordered we must already have $b_1, \dots, b_n \in B$ for some $B \in N$, in contradiction to B being algebraically independent over K.
- We have $A \subset \bigcup_{B \in N} B \subset A \cup C$ since $A \subset B \subset A \cup C$ for all $B \in N$.

Hence M has a maximal element B. We claim that B is a transcendence basis of L over K. In fact, as an element of M we know that B is algebraically independent. Moreover, since B is maximal every element of C is algebraic over $K(B)$, hence $K(C)$ is algebraic over $K(B)$ [G3, Exercise 2.27 (a)], and so L is algebraic over $K(B)$ by Lemma 9.6 (b) since L is algebraic over $K(C)$ by assumption. □

Proposition and Definition 11.29 (Transcendence degree)**.** *Let $K \subset L$ be a field extension. If L has a finite transcendence basis over K, then all such transcendence bases are finite and have the same number n of elements. In this case we call this number n the* ***transcendence degree*** $\operatorname{tr}\deg_K L$ *of L over K. Otherwise, we set formally* $\operatorname{tr}\deg_K L = \infty$.

Proof. Let B and C be transcendence bases of L over K, and assume that $B = \{b_1, \dots, b_n\}$ has n elements. By symmetry, it suffices to show that $|C| \leq |B|$. We will prove this by induction on n, with the case $n = 0$ being obvious since then $K \subset L$ is algebraic, and hence necessarily $C = \emptyset$.

So assume now that $n > 0$. As we are done if $C \subset B$, we can pick an element $c \in C \backslash B$, which is necessarily transcendental. By Lemma 11.28 we can extend it to a transcendence basis B' by elements of B. Note that B' can have at most n elements: the set $\{c, b_1, \dots, b_n\}$ is algebraically dependent since B is a transcendence basis.

So B' and C are two transcendence bases of L over K that both contain c. Hence $B' \backslash \{c\}$ and $C \backslash \{c\}$ are two transcendence bases of L over $K(c)$. By induction this means that $|B' \backslash \{c\}| = |C \backslash \{c\}|$, and hence that $|C| = |C \backslash \{c\}| + 1 = |B' \backslash \{c\}| + 1 \leq n = |B|$ as desired. □

Example 11.30.

(a) A field extension is algebraic if and only if it has transcendence degree 0.

(b) Given that π is transcendental over $\mathbb{Q}$, the field extensions $\mathbb{Q} \subset \mathbb{Q}(\pi)$ and $\mathbb{Q} \subset \mathbb{Q}(\pi, \sqrt{2})$ both have transcendence degree 1, with $\{\pi\}$ as a transcendence basis. The field extension $\mathbb{Q} \subset \mathbb{C}$ has infinite transcendence degree, since the set of all rational functions in finitely many elements with coefficients in $\mathbb{Q}$ is countable, whereas $\mathbb{C}$ is uncountable.

(c) The quotient field $K(x_1, \ldots, x_n)$ of the polynomial ring $K[x_1, \ldots, x_n]$ has transcendence degree n over K, since $\{x_1, \ldots, x_n\}$ is a transcendence basis by Example 11.26 (b).

With the notion of transcendence degree we can now give another interpretation of the dimension of an irreducible variety — or more generally of a finitely generated algebra over a field which is a domain.

Proposition 11.31 (Transcendence degree as dimension)**.** *Let R be a finitely generated algebra over a field K, and assume that R is an integral domain. Then* $\dim R = \operatorname{tr\,deg}_K \operatorname{Quot} R$.

Proof. Let $K[z_1, \ldots, z_n] \to R$ be a Noether normalization as in Proposition 10.5, where $n = \dim R$ by Remark 11.10. Then $\operatorname{Quot} R$ is algebraic over $\operatorname{Quot} K[z_1, \ldots, z_n] = K(z_1, \ldots, z_n)$: if $a, b \in R$ with $b \neq 0$ then a and b are integral over $K[z_1, \ldots, z_n]$, hence algebraic over $K(z_1, \ldots, z_n)$. Therefore $K(z_1, \ldots, z_n)(a, b)$ is algebraic over $K(z_1, \ldots, z_n)$ [G3, Exercise 2.27 (a)], which implies that $\frac{a}{b} \in \operatorname{Quot} R$ is algebraic over $K(z_1, \ldots, z_n)$.

But this means that the transcendence basis $\{z_1, \ldots, z_n\}$ of $K(z_1, \ldots, z_n)$ over K (see Example 11.26 (b)) is also a transcendence basis of $\operatorname{Quot} R$ over K, i. e. that $\operatorname{tr\,deg}_K \operatorname{Quot} R = n = \dim R$. □

Exercise 11.32. Let K be a field, and let R be a K-algebra which is an integral domain. If R is finitely generated we know by Proposition 11.31 that $\dim R = \operatorname{tr\,deg}_K \operatorname{Quot} R$. Now drop the assumption that R is finitely generated and show:

(a) $\dim R \leq \operatorname{tr\,deg}_K \operatorname{Quot} R$.

(b) The inequality in (a) may be strict.

Exercise 11.33. Let $X \subset \mathbb{A}^n_K$ and $Y \subset \mathbb{A}^m_K$ be varieties over a field K. Prove:

(a) $\dim(X \times Y) = \dim X + \dim Y$.

(b) If $n = m$ then every irreducible component of $X \cap Y$ has dimension at least $\dim X + \dim Y - n$.

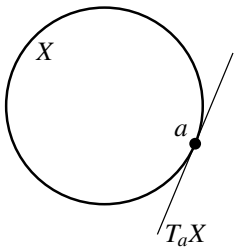
22

Finally, another approach to the notion of dimension is given by linearization. If a is a point of a variety $X \subset \mathbb{A}^n_K$ over a field K we can try to approximate X around a by an affine linear space in $\mathbb{A}^n_K$, and consider its dimension to be some sort of local dimension of X at a. Let us describe the precise construction of this linear space.

Construction 11.34 (Tangent spaces)**.** Let us consider a variety $X \subset \mathbb{A}^n_K$ with coordinate ring $A(X) = K[x_1, \ldots, x_n]/I(X)$, and let $a = (a_1, \ldots, a_n) \in X$. By a linear change of coordinates $y_i := x_i - a_i$ we can shift a to the origin.

Consider the "differential" map

$$d : K[x_1, \ldots, x_n] \to K[y_1, \ldots, y_n], \quad f \mapsto \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \cdot y_i,$$



where $\frac{\partial f}{\partial x_i}$ denotes the formal derivative of f with respect to x_i [G1, Exercise 9.10]. It can be thought of as assigning to a polynomial f its linear term in the Taylor expansion at the point a. We then define the *tangent space* to X at a to be

$$T_aX := V(\{df : f \in I(X)\}) \quad \subset K^n, \qquad (*)$$

i. e. we take all equations describing X, linearize them at the point a, and take the common zero locus of these linearizations. As in the picture above we can therefore think of T_aX as the linear space that

gives the best approximation of X at a. Note that it suffices in $(*)$ to let f range over a generating set for $I(X)$, since for $f,g \in I(X)$ and $h \in K[x_1,\dots,x_n]$ we have

$$d(f+g) = df+dg \quad \text{and} \quad d(hf) = h(a)\,df + f(a)\,dh = h(a)\,df.$$

In order to transfer these ideas to other non-geometric cases we need an alternative description of the tangent space: let $P = I(a) = (\overline{y_1},\dots,\overline{y_n}) \trianglelefteq A(X)$ be the (maximal) ideal of the point a, and consider the K-linear map

$$\varphi : P \to \mathrm{Hom}_K(T_aX, K), \quad \overline{f} \mapsto df|_{T_aX}$$

(that is well-defined by the definition $(*)$ of T_aX). It is clearly surjective since every linear map on T_aX must be a linear combination of the coordinate functions. Moreover, we claim that $\ker\varphi = P^2$:

"$\subset$" Let $\varphi(\overline{f}) = df|_{T_aX} = 0$. Note that $L := \{dg : g \in I(X)\}$ is a linear subspace of $\langle y_1,\dots,y_n\rangle$ of some dimension k. Its zero locus T_aX then has dimension $n-k$. Hence the space of all linear forms vanishing on T_aX has dimension k again and clearly contains L, and thus must be equal to L. As df lies in this space, we conclude that $df = dg$ for some $g \in I(X)$. Then $d(f-g) = 0$, and thus the Taylor expansion of $f-g$ at a does not contain constant or linear terms. Hence $\overline{f} = \overline{f-g} \in P^2$.

"$\supset$" The K-vector space P^2 is generated by products $\overline{fg}$ for $f,g \in P = I(a)$, and we clearly have $\varphi(\overline{fg}) = f(a)\,dg|_{T_aX} + g(a)\,df|_{T_aX} = 0$ since $f(a) = g(a) = 0$.

So by the homomorphism theorem we get a natural isomorphism $P/P^2 \to \mathrm{Hom}_K(T_aX,K)$ of K-vector spaces. In other words, the tangent space T_aX is naturally the dual vector space of P/P^2. In particular, its dimension is $\dim_K T_aX = \dim_K P/P^2$.

Remark 11.35 (Tangent spaces are local)**.** Let P be a maximal ideal in a ring R, and let $K = R/P$. Then the R-module P/P^2 is also a K-vector space (with the same multiplication), and the classes of elements of $S = R\backslash P$ in K are invertible. Hence localizing P/P^2 at S is an isomorphism of K-vector spaces, i. e. we get

$$P/P^2 \;\cong\; S^{-1}(P/P^2) \;\cong\; S^{-1}P/S^{-1}P^2$$

by Corollary 6.22 (b). So the tangent space T_aX of a variety X at a point $a \in X$ as in Construction 11.34 can equally well be obtained from the local ring of X at a.

This observation allows us to assign to any local ring R (with maximal ideal P and $K = R/P$) a number $\dim_K P/P^2$ that can be interpreted as the dimension of the tangent space if R is the ring of local functions on a variety at a point. Let us see how this number compares to the actual dimension $\dim R$ of R. For simplicity, we will restrict our attention to Noetherian rings.

Lemma 11.36. *Let R be a local Noetherian ring with maximal ideal P, and set $K = R/P$.*

(a) *The number $\dim_K P/P^2$ is the minimum number of generators for the ideal P.*

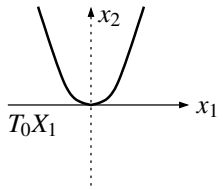
(b) $\dim R \le \dim_K P/P^2 < \infty$.

Proof. Of course, the ideal P is finitely generated since R is Noetherian. Let n be the minimum number of generators for P.

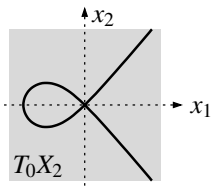
(a) Let $P = (a_1,\dots,a_n)$. Then $\overline{a_1},\dots,\overline{a_n}$ generate P/P^2 as an R/P-vector space, and thus $\dim_K P/P^2 \le n$. Moreover, if $\overline{a_1},\dots,\overline{a_n}$ were linearly dependent then after relabeling $\overline{a_1},\dots,\overline{a_{n-1}}$ would still generate P/P^2 — but then Nakayama's Lemma as in Exercise 6.16 (b) (for $M = I = P$) implies that $a_1,\dots,a_{n-1}$ generate P as an R-module, in contradiction to the minimality of n. Hence $\dim_K P/P^2 = n$.

(b) By Remark 11.5 (c), Remark 11.5 (a), and Corollary 11.17 we have $\dim R = \dim R_P = \mathrm{codim}\, P \le n < \infty$. Hence the result follows since $n = \dim_K P/P^2$ by (a). □

Example 11.37. Consider the curves X_1 and X_2 in $\mathbb{A}^2_{\mathbb{R}}$ with ideals $I(X_1) = (x_2 - x_1^2)$ and $I(X_2) = (x_2^2 - x_1^2 - x_1^3)$, respectively, as in the picture below.



$X_1 = V(x_2 - x_1^2)$



$X_2 = V(x_2^2 - x_1^2 - x_1^3)$

By Construction 11.34, their tangent spaces T_0X_1 and T_0X_2 at the origin are the zero loci of the linear parts of their defining equations, i. e.

$$\begin{aligned} T_0X_1 &= V(x_2) \quad \text{is the } x_1\text{-axis, and} \\ T_0X_2 &= V(0) \quad \text{is all of } \mathbb{R}^2. \end{aligned}$$

Hence for X_1, the tangent space T_0X_1 has the same dimension as X_1, i. e. the first inequality in Lemma 11.36 (b) is an equality. Geometrically, this means that X_1 has a "good" approximation at 0 by a linear space of the same dimension — which is the usual understanding of a tangent space. In the case of X_2 however, there is no reasonable approximation at 0 by a linear space of dimension 1. Consequently, the dimension of the tangent space is bigger than that of X, i. e. the first inequality of Lemma 11.36 (b) is strict. This is usually expressed by saying that the origin is a *singular point* of X_2, whereas it is a *regular* or *smooth* point of X_1. The precise definition of these terms is as follows.

Definition 11.38 (Regular local rings, regular points of a variety)**.**

(a) Let R be a local ring with maximal ideal P, and set $K = R/P$. We say that R is **regular** if it is Noetherian and $\dim R = \dim_K P/P^2$.

(b) Let X be a variety. A point $a \in X$ is called **regular** or **smooth** if its ring of local functions $A(X)_{I(a)}$ as in Example 6.5 (d) is regular, i. e. by Construction 11.34 if $\dim X = \dim T_aX$ (note that $A(X)_{I(a)}$ is always Noetherian by Remark 7.15 and Exercise 7.23). Otherwise we say that a is a **singular** point of X.

Example 11.39.

(a) Any field is a regular local ring (of dimension 0).

(b) Let $X \subset \mathbb{A}^n_K$ be a variety over a field K, with ideal $I(X) = (f_1, \dots, f_r) \trianglelefteq K[x_1, \dots, x_n]$. By Construction 11.34, the tangent space T_aX to X at a point $a \in X$ is the kernel of the Jacobian matrix

$$Jf(a) := \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{\substack{1 \le i \le r \\ 1 \le j \le n}}.$$

Hence a is a regular point of X if and only if $\dim X = \dim_K \ker Jf(a)$. In particular, this is the case if $Jf(a)$ has maximal row rank r, because then

$$\begin{aligned} \dim X &\le \dim_K \ker Jf(a) \quad \text{(Lemma 11.36 (b))} \\ &= n - r \\ &\le \dim X, \end{aligned}$$

since by Remark 11.18 every irreducible component of $X = V(f_1, \dots, f_r)$ has dimension at least $n - r$. Note that this fits well with the implicit function theorem [G2, Proposition 27.9] in analysis, which states that (in the case $K = \mathbb{R}$) a Jacobi matrix with maximal row rank at a point a guarantees that X is locally the graph of a continuously differentiable function, and thus that X is regular at a in the sense of Example 11.37.

Explicitly, the standard parabola X_1 in Example 11.37 is regular at every point since its Jacobian matrix $(-2x_1 \;\; 1)$ has full row rank everywhere, whereas the curve X_2 has a singular point at the origin as its Jacobian matrix $(-2x_1 - 3x_1^2 \;\; 2x_2)$ vanishes there.

Finally, we will prove the important result that any regular local ring is an integral domain — a statement that is easy to understand geometrically since a variety that is regular at a point in the sense of Example 11.37 should not consist of several components that meet in this point.

Proposition 11.40. *Any regular local ring is an integral domain.*

Proof. Let R be a regular local ring of dimension n with maximal ideal P, and set $K = R/P$. We will prove the statement of the proposition by induction on n. For $n = 0$ we have $\dim_K P/P^2 = 0$, hence $P = P^2$, so $P = 0$ by Nakayama's Lemma as in Exercise 6.16 (a). But if the zero ideal is maximal then R is a field, and hence obviously an integral domain.

So let now $n > 0$, and let $P_1, \dots, P_r$ be the minimal primes over the zero ideal as in Corollary 8.30. Note that $P \not\subset P^2 \cup P_1 \cup \dots \cup P_r$ since otherwise by Exercise 2.10 (c) we would have $P \subset P^2$ (hence $P = P^2$, and thus $n = 0$ by Lemma 11.36 (b)) or $P \subset P_i$ for some i (hence $P_j \subset P \subset P_i$ for all j, so $r = 1$ and P is the unique maximal and the unique minimal prime, which means again that $n = 0$). We can therefore find an element $a \in P$ with $a \notin P^2$ and $a \notin P_i$ for all i.

Consider the ring $R/(a)$, which is again local (with maximal ideal $P/(a)$, and $(P/(a))/(P/(a))^2 \cong P/(P^2 + (a))$) and Noetherian (by Remark 7.8 (b)). We will show that it is regular of dimension $n-1$:

- As $a \notin P^2$, we can extend the element $\overline{a} \in P/P^2$ to a basis $(\overline{a}, \overline{a_2}, \dots, \overline{a_n})$ of P/P^2. Then $\overline{a_2}, \dots, \overline{a_n}$ generate the vector space $P/(P^2 + (a))$, and hence $\dim P/(P^2 + (a)) \leq n - 1$.
- Let $P_i = Q_0 \subsetneq \dots \subsetneq Q_n = P$ be a maximal chain of prime ideals in R; note that it has to start with a minimal prime ideal P_i and end with P. As $a \in P$, we can arrange this by Exercise 11.16 (a) so that $a \in Q_1$. Then $Q_1/(a) \subsetneq \dots \subsetneq Q_n/(a)$ is a chain of prime ideals of length $n-1$ in $R/(a)$, and so $\dim R/(a) \geq n - 1$.

Together with Lemma 11.36 (b) we conclude that

$$n - 1 \leq \dim R/(a) \leq \dim P/(P^2 + (a)) \leq n - 1,$$

and hence that $R/(a)$ is regular of dimension $n-1$.

By induction, this means that $R/(a)$ is an integral domain, and hence by Lemma 2.3 (a) that (a) is a prime ideal. Hence we must have $P_i \subset (a)$ for some i. This means for any $b \in P_i$ that $b = ac$ for some $c \in R$, hence $c \in P_i$ since $a \notin P_i$ and P_i is prime, and therefore $b \in P \cdot P_i$. In total, this means that $P_i = P \cdot P_i$, and so $P_i = 0$ by Nakayama as in Exercise 6.16 (a). Thus the zero ideal is prime, i. e. R is an integral domain. □

23

12. Valuation Rings

In the remaining two chapters of this class we will restrict our attention to 1-dimensional rings, i. e. geometrically to curves. More precisely, we will only deal with 1-dimensional rings whose localizations at all maximal ideals are regular. For a curve X over an algebraically closed field, when the maximal ideals of $A(X)$ are in one-to-one correspondence with points of X by Remark 10.11, this means precisely that we will require X to be smooth, i. e. that all points of X are smooth in the sense of Example 11.37 and Definition 11.38.

In the current chapter we will consider such rings and varieties from a local point of view, postponing their global study to Chapter 13. Such 1-dimensional regular local rings are probably the "nicest" rings (that are not fields) — they have a very special structure as we will see in the following key lemma and its geometric interpretation.

Lemma 12.1. *Let R be a 1-dimensional regular local ring.*

(a) *The maximal ideal P of R is a non-zero principal ideal.*

(b) *For every element $a \in R\backslash\{0\}$ there is a unique number $n \in \mathbb{N}$ such that $(a) = P^n$. We will call it the valuation of a (see Remark 12.15 (a)).*

Proof.

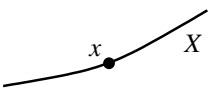
(a) By assumption, the vector space dimension of P/P^2 over R/P is 1. But this is also the minimum number of generators for P by Lemma 11.36 (a), hence P is principal and non-zero.

(b) By Proposition 11.40 we know that R is an integral domain. So it has a unique minimal prime ideal 0 and a unique maximal ideal P, and since $\dim R = 1$ these two prime ideals are in fact the only ones. Hence by Lemma 2.21 we know that $\sqrt{(a)}$ is equal to P or R. In both cases it follows that $P^n \subset (a)$ for some $n \in \mathbb{N}$ by Exercise 7.22 (b). Since P is principal by (a), we can write this equivalently as $t^n = ba$ for some $b \in R$ and a generator t of P.

Now choose n minimal with this property. Then b must be a unit: otherwise (b) is contained in a maximal ideal by Corollary 2.17, which must be P. Hence $b = b't$ for some $b' \in R$, so $t^n = b'ta$, and thus $t^{n-1} = b'a$ since R is an integral domain — in contradiction to the minimality of n.

So with this choice of n we get $t^n = ba$ for a unit b, which means that $(a) = (t^n) = P^n$. Moreover, no other choice of exponent would work: if we had $t^k = ca$ for a unit c and $k < n$ (resp. $k > n$), then this would imply that $t^{n-k} = bc^{-1}$ (resp. $t^{k-n} = cb^{-1}$) is a unit, in contradiction to $t \in P$. □

Remark 12.2 (Geometric interpretation: orders of vanishing)**.** Let x be a fixed smooth point on a curve X, and let R be the ring of local functions on X at x, i. e. the localization of $A(X)$ at the maximal ideal $I(x)$ as in Example 6.5 (d). Then R is a 1-dimensional regular local ring.



The maximal ideal P of R then consists of all local functions that vanish at x, and a generator t for P as in Lemma 12.1 (a) can be thought of as a local coordinate for X in a neighborhood of x (that has the value 0 at x, and vanishes to order 1 there). Consequently, if $a \in R$ is a local function on X at x the equation $(a) = P^n$ of Lemma 12.1 (b), i. e. $a = ct^n$ for a unit c, means that a vanishes to order n at x. We can therefore think of the valuation constructed in Lemma 12.1 as an order of vanishing of a local function on a curve at a point.

Actually, the full algebraic notion of a valuation is more general than above: one considers valuations on general integral domains and in fact also on their quotient fields, and allows their values to be in any ordered Abelian group instead of just in $\mathbb{N}$ or $\mathbb{Z}$. Let us give the precise definition now, returning to our special case of 1-dimensional regular local rings later in Proposition 12.14.

Definition 12.3 (Valuations and valuation rings)**.**

(a) An **ordered group** is an Abelian group G (usually written additively) with a total order $\leq$ such that $m \leq n$ implies $m+k \leq n+k$ for all $m,n,k \in G$.

(b) Let K be a field; as usual we write K^* for $K \backslash \{0\}$. A **valuation** on K is a map $\nu : K^* \to G$ for an ordered group G such that for all $a,b \in K^*$ we have

(i) $\nu(ab) = \nu(a) + \nu(b)$ (i. e. ν is a homomorphism of groups); and

(ii) $\nu(a+b) \geq \min(\nu(a), \nu(b))$ if $a+b \neq 0$.

We extend this map formally to all of K by setting $\nu(0) := \infty$, so that (i) and (ii) hold for all $a,b \in K$.

(c) For a valuation $\nu : K^* \to G$ the subgroup $\nu(K^*) \leq G$ is called the **value group**, and

$$R_\nu := \{a \in K : \nu(a) \geq 0\}$$

the **valuation ring** of ν. Note that it is indeed a ring: we clearly have $\nu(0) = \infty \geq 0$ and $\nu(1) = 0$, and for $a,b \in K$ with $\nu(a) \geq 0$ and $\nu(b) \geq 0$ it follows immediately from (i) and (ii) above that $\nu(a+b) \geq 0$ and $\nu(ab) \geq 0$ as well.

Example 12.4.

(a) Any field K allows the *trivial valuation* $\nu : K^* \to \{0\}$, $a \mapsto 0$. Its value group is $\{0\}$, and its valuation ring is K itself. Note that even in this case we still have $\nu(0) = \infty$ by definition.

(b) The groups $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are all ordered. In fact, the most common non-trivial value group in our examples will be $\mathbb{Z}$ (see Proposition 12.13). But for general valuations other groups may occur as well, as we will see e. g. in (e) below, and in the proof of Proposition 12.8.

(c) The most important example of a valuation is the following. Fix a prime element p in a unique factorization domain R, and let $K = \operatorname{Quot} R$. Note that every non-zero element of R can be written uniquely as $a\,p^n$ for some $n \in \mathbb{N}$ and $a \in R$ with $p \nmid a$, and consequently every element of K^* can be written uniquely as $a\,p^n$ for some $n \in \mathbb{Z}$ and $a \in K^*$ that can be written as a quotient of elements of R that do not contain p as a factor. With this representation there is an obvious map

$$\nu : K^* \to \mathbb{Z}, \quad a\,p^n \mapsto n$$

that assigns to every element the exponent of p in its prime factorization. It is clearly well-defined and a homomorphism of groups. It also satisfies condition (ii) of Definition 12.3 (b): for any two such elements $a\,p^n$ and $b\,p^m$ (with $a,b \in K^*$ not containing the prime factor p, and $n,m \in \mathbb{Z}$), without loss of generality with $n \geq m$, we have

$$\nu(a\,p^n + b\,p^m) = \nu\big(p^m(a\,p^{n-m} + b)\big) = m + \nu(a\,p^{n-m} + b) \overset{(*)}{\geq} m = \nu(b\,p^m),$$

where $(*)$ follows since $a\,p^{n-m} + b$ does not contain the factor p in its denominator. Hence ν is a valuation on K. Its value group is obviously $\mathbb{Z}$, and its valuation ring is

$$R_\nu = \{a\,p^n : a \in K^*, n \in \mathbb{N}, a \text{ does not contain } p\} \cup \{0\},$$

i. e. precisely the localization $R_{(p)}$ of R at the prime ideal (p) as in Example 6.5 (d).

(d) Let L be a field, and let

$$K = \left\{ \sum_{n \in \mathbb{Z}} a_n t^n : a_n \in L \text{ for all } n, \text{ the set } \{n \in \mathbb{Z} : a_n \neq 0\} \text{ is bounded below} \right\}$$

be the set of all formal "power series with integer exponents" in one variable t. As in the case of the usual power series ring $L[\![t]\!]$, there are no convergence requirements on the coefficients of the series [G1, Definition 9.1]. We do require however that for a non-zero element $f \in$

K there is a minimum exponent of t occurring in f, which implies that we can write f uniquely as $f = t^n g$ for some $n \in \mathbb{Z}$ and $g \in L[\![t]\!]$ with non-zero constant coefficient. With this definition K is in fact a field: it is obvious that it is a ring, and since a power series with non-zero constant coefficient is invertible [G1, Exercise 9.12], a multiplicative inverse of $t^n g$ as above is $t^{-n} g^{-1}$. We call K the field of (formal) *Laurent series* over L.

It is now obvious that

$$\nu : K^* \to \mathbb{Z}, \quad \sum_{n \in \mathbb{Z}} a_n t^n \mapsto \min\{n \in \mathbb{Z} : a_n \neq 0\}$$

is a valuation on K. Its valuation ring is simply the ring $L[\![t]\!]$ of power series over L. In fact, this construction is a special case of (c) since one can show that $L[\![t]\!]$ is a unique factorization domain with $t \in L[\![t]\!]$ prime, and $K = \mathrm{Quot}(L[\![t]\!])$.

(e) The idea of (d) can be generalized to construct a valuation with value group $\mathbb{Q}$: for a field L we now consider the set

$$K = \left\{ \sum_{q \in \mathbb{Q}} a_q t^q : a_q \in L \text{ for all } q, \text{ the set } \{q \in \mathbb{Q} : a_q \neq 0\} \text{ is bounded below and has bounded denominators} \right\}$$

of all formal "power series with rational exponents", and the same map

$$\nu : K^* \to \mathbb{Q}, \quad \sum_{q \in \mathbb{Q}} a_q t^q \mapsto \min\{q \in \mathbb{Q} : a_q \neq 0\}$$

as above. It can then be checked that K is again a field — we will not do this here as we will not need this field again — and that ν is a valuation with value group $\mathbb{Q}$. The field K is called the field of (formal) *Puiseux series* over L.

Remark 12.5.

(a) We will see in Propositions 12.13 and 12.14 that a 1-dimensional regular local ring is a unique factorization domain, and that in this case the construction of Example 12.4 (c) specializes to our original situation of Lemma 12.1 and Remark 12.2. For the moment it suffices to note that, in the geometric situation, we should imagine K in Definition 12.3 to be the field of rational functions on a variety X around a smooth point a. The valuation $\nu(f) \in \mathbb{Z}$ of an element $f \in K^*$ can then be thought of as the order of the zero (if $\nu(f) > 0$) or pole (if $\nu(f) < 0$) of f at $a \in X$.

(b) It is clear from Definition 12.3 that every valuation ring R_ν of a valuation ν is an integral domain, since it is a subring of a field. In fact, valuation rings have many more nice properties. Let us prove the most important ones now.

Lemma 12.6 (Properties of valuation rings)**.** *Let $R = R_\nu$ be the valuation ring of a valuation $\nu : K^* \to G$ as in Definition 12.3.*

(a) *For all $a \in K^*$ we have $a \in R$ or $a^{-1} \in R$.*

(b) *For all $a, b \in R$ we have $\nu(a) \leq \nu(b)$ if and only if $b \in (a)$.*

(c) *The group of units of R is $R^* = \{a \in R : \nu(a) = 0\}$.*

(d) *R is a local ring with maximal ideal $P = \{a \in R : \nu(a) > 0\}$.*

(e) *R is a normal domain.*

Proof.

(a) Since $\nu(a) + \nu(a^{-1}) = \nu(aa^{-1}) = \nu(1) = 0$ we must have $\nu(a) \geq 0$ or $\nu(a^{-1}) \geq 0$, and thus $a \in R$ or $a^{-1} \in R$.

(b) The statement is obvious for $a = 0$, so let us assume that $a \neq 0$. If $\nu(a) \leq \nu(b)$ then $\nu(\frac{b}{a}) = \nu(b) - \nu(a) \geq 0$, hence $\frac{b}{a} \in R$, and thus $b = \frac{b}{a} \cdot a \in (a)$. Conversely, if $b \in (a)$ then $b = ca$ for some $c \in R$, and therefore $\nu(b) = \nu(c) + \nu(a) \geq \nu(a)$.

(c) Let $a \in R$ with $a \neq 0$, so $\nu(a) \geq 0$. Then $a \in R^*$ if and only if $a^{-1} \in R$, i. e. $\nu(a^{-1}) = -\nu(a) \geq 0$ as well, which is the case if and only if $\nu(a) = 0$.

(d) It is clear that P is an ideal, since $a, b \in P$ and $r \in R$ imply $\nu(a+b) \geq \min(\nu(a), \nu(b)) > 0$ and $\nu(ra) = \nu(r) + \nu(a) > 0$. Moreover, any bigger ideal contains an element with valuation 0, i. e. a unit by (c), and thus is equal to R.

(e) Let $a \in K^*$ be integral over R. Then $a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0$ for some $n \in \mathbb{N}$ and $c_0, \ldots, c_{n-1} \in R$. Assume that $a \notin R$. Then $a^{-1} \in R$ by (a), which leads to the contradiction

$$a = -c_{n-1} - \cdots - c_0\, a^{-n+1} \quad \in R.$$

Hence we must have $a \in R$ as claimed. □

So far, we have always started with a valuation ν on a field K, and then constructed a valuation ring $R_\nu \subset K$ from it. We will now show that this process can be reversed: the valuation ring R_ν itself contains in fact enough information to reconstruct the valuation ν from it.

Construction 12.7 (Reconstruction of the valuation from its valuation ring)**.** Let $R = R_\nu$ be the valuation ring of a valuation $\nu : K^* \to G$ as in Definition 12.3. Then ν can be recovered completely from R up to isomorphisms in the following sense:

(a) We must have $K = \operatorname{Quot} R$: the inclusion "$\supset$" is obvious, and "$\subset$" follows from Lemma 12.6 (a), since both $a \in R$ and $a^{-1} \in R$ imply $a \in \operatorname{Quot} R$.

(b) The group homomorphism $\nu : K^* \to G$ has kernel R^* by Lemma 12.6 (c), and image $\nu(K^*)$. Hence the homomorphism theorem implies that the value group $\nu(K^*)$ is isomorphic to K^*/R^* (which is uniquely determined by R due to (a)), and that with this isomorphism the valuation map $\nu : K^* \to K^*/R^*$ is just the quotient map.

(c) The order on $\nu(K^*)$ is determined by R since for $a, b \in K^*$ we have $\nu(a) \leq \nu(b)$ if and only if $\nu(\frac{b}{a}) \geq 0$, i. e. if $\frac{b}{a} \in R$.

So we can say that every valuation ring belongs to a unique valuation (if we assume the valuation to be surjective, i. e. we can of course not recover the whole group G from R, but only the valuation subgroup $\nu(K^*)$ of G). In fact, there is even an easy criterion to determine whether a given ring is such a valuation ring:

Proposition 12.8 (Alternative characterization of valuation rings)**.** *For a ring R the following statements are equivalent:*

(a) *R is the valuation ring of a valuation $\nu : K^* \to G$ (which is then unique up to isomorphisms by Construction 12.7).*

(b) *R is an integral domain, and for all $a \in \operatorname{Quot} R \setminus \{0\}$ we have $a \in R$ or $a^{-1} \in R$.*

Proof. The implication (a) $\Rightarrow$ (b) is just Lemma 12.6 (a). For the opposite direction, let R be an integral domain, set $K = \operatorname{Quot}(R)$ and $G = K^*/R^*$. Note that, in contrast to Definition 12.3 (a), the group G will be written *multiplicatively* in this proof. Now

$$\overline{a} \leq \overline{b} \;:\Leftrightarrow\; \frac{b}{a} \in R$$

is a well-defined relation on G (for $a, b \in K^*$ and $c, d \in R^*$ with $\frac{b}{a} \in R$ we have $\frac{bd}{ac} \in R$) and makes G into an ordered group (for $a, b, c \in K^*$ with $\frac{b}{a} \in R$ we have $\frac{bc}{ac} \in R$). Let $\nu : K^* \to G$ be the quotient map, so that $\nu(a) \leq \nu(b)$ for $a, b \in K^*$ if and only if $\frac{b}{a} \in R$. Then ν is a valuation:

(i) The relation $\nu(ab) = \nu(a)\,\nu(b)$ is obvious.

(ii) Let $a, b \in K^*$. By assumption we have $\frac{a}{b} \in R$ or $\frac{b}{a} \in R$. In the first case $\frac{a+b}{b} = \frac{a}{b} + 1 \in R$ and thus $\nu(a+b) \geq \nu(b)$, whereas in the second case we get similarly $\nu(a+b) \geq \nu(a)$. Hence $\nu(a+b) \geq \min(\nu(a), \nu(b))$.

Its valuation ring R_ν is now exactly R, since we have $\nu(a) \geq 0 = \nu(1)$ for $a \in K^*$ if and only if $\frac{a}{1} \in R$. □

Remark 12.9. Proposition 12.8 states that we could alternatively *define* a valuation ring to be an integral domain R such that $a \in R$ or $a^{-1} \in R$ for all $a \in \operatorname{Quot} R \setminus \{0\}$. In fact, one will often find this definition in the literature, and in the following we will also often talk about a valuation ring R without mentioning a valuation first. Our results above show that even then the ring R always has a unique valuation $\nu : K \to G$ associated to it such that $R = R_\nu$ and ν is surjective.

24

This alternative characterization of valuation rings allows us to prove some more of their properties that would be harder to see using the original Definition 12.3. The following lemma and proposition give two examples of this.

Lemma 12.10 (Enlarging valuation rings)**.** *Let R be valuation ring, and let R' be another ring with $R \subset R' \subset \operatorname{Quot} R$. Then R' is also a valuation ring, with $\operatorname{Quot} R' = \operatorname{Quot} R$.*

Proof. Taking quotient fields in the inclusion $R \subset R' \subset \operatorname{Quot} R$ we see that $\operatorname{Quot} R' = \operatorname{Quot} R$. Now if $a \in \operatorname{Quot} R' = \operatorname{Quot} R$ we have $a \in R$ or $a^{-1} \in R$ by Proposition 12.8, and hence also $a \in R'$ or $a^{-1} \in R'$. Therefore R' is a valuation ring by Proposition 12.8 again. □

Proposition 12.11 (Integral closure from valuation rings)**.** *Let R be an integral domain, and let $\overline{R}$ be its integral closure in* $\operatorname{Quot} R$*. Then*

$$\overline{R} = \bigcap_{\substack{R \subset R' \subset \operatorname{Quot} R \\ R' \text{ valuation ring}}} R'.$$

Proof.

"$\subset$" Let $x \in \overline{R}$, and let R' be a valuation ring with $R \subset R' \subset \operatorname{Quot} R$. So $x \in \operatorname{Quot} R = \operatorname{Quot} R'$ is integral over R, and thus also over R'. But then $x \in R'$ since R' is normal by Lemma 12.6 (e).

"$\supset$" Let $x \notin \overline{R}$. We will construct a valuation ring R' with $R \subset R' \subset \operatorname{Quot} R$ and $x \notin R'$.

Note that $x \notin R[\frac{1}{x}]$, since otherwise there would be a relation $x = a_0 + a_1 x^{-1} + \cdots + a_n x^{-k}$ with $a_0, \ldots, a_k \in R$, which means that $x^{k+1} - a_0 x^k - a_1 x^{k-1} - \cdots - a_k = 0$, and thus that x would be integral over R in contradiction to $x \notin \overline{R}$. So the set

$$M = \{R' : R' \text{ is a ring with } R[\tfrac{1}{x}] \subset R' \subset \operatorname{Quot} R \text{ and } x \notin R'\}$$

is non-empty since it contains $R[\frac{1}{x}]$. By Zorn's Lemma as in Proposition 2.16 it contains a maximal element R'. Of course we then have $R \subset R' \subset \operatorname{Quot} R$ and $x \notin R'$, so it only remains to show that R' is a valuation ring.

We will prove this using the criterion of Proposition 12.8 (b). So let $a \in \operatorname{Quot} R' = \operatorname{Quot} R$, and assume for a contradiction that $a \notin R'$ and $a^{-1} \notin R'$. Then $R'[a]$ and $R'[a^{-1}]$ strictly contain R', and thus by maximality of R' in M we conclude that $x \in R'[a]$ and $x \in R'[a^{-1}]$. So we can write

$$x = \sum_{i=0}^{n} r_i a^i = \sum_{i=0}^{m} s_i a^{-i}$$

for some $n, m \in \mathbb{N}$ and $r_0, \ldots, r_n, s_0, \ldots, s_m \in R'$. We can choose n and m minimal, and assume without loss of generality that $m \leq n$. Then

$$\begin{aligned} x &= s_0 + (x - s_0) x^{-1} \sum_{i=0}^{n} r_i a^i \\ &= s_0 + (x - s_0) x^{-1} \sum_{i=0}^{n-1} r_i a^i + \sum_{i=1}^{m} s_i a^{-i} x^{-1} r_n a^n \end{aligned}$$

is a polynomial expression of degree less than n in a with coefficients in R', in contradiction to the minimality of n. Hence our assumption $a \notin R'$ and $a^{-1} \notin R'$ must have been wrong, and we conclude that R' is a valuation ring. □

Exercise 12.12. Let R be a valuation ring with value group G. Let us call a subset $A \subset G$ *non-negative* if $a \geq 0$ for all $a \in G$, and *saturated* if for all $a, b \in G$ with $a \leq b$ and $a \in A$ we have $b \in A$. Show:

(a) There is a natural inclusion-preserving one-to-one correspondence between ideals of R and non-negative saturated subsets of G.

(b) For any two ideals $I, J \subset R$ we have $I \subset J$ or $J \subset I$.

So far we have considered arbitrary valuation rings. However, as most rings occurring in practice are Noetherian, we now want to specialize to this case. Surprisingly, this seemingly small additional assumption has far-reaching consequences: it fixes the value group to be $\mathbb{Z}$, restricts the possibilities to the 1-dimensional regular local rings of Lemma 12.1, and leads to many other nice properties as the following two propositions show.

Proposition and Definition 12.13 (Discrete valuation rings)**.** *For a valuation ring R with unique maximal ideal P (see Lemma 12.6 (d)) the following statements are equivalent:*

(a) *R is Noetherian, but not a field.*

(b) *R is a principal ideal domain, but not a field.*

(c) *The value group of R is $\mathbb{Z}$.*

Moreover, in this case the valuation $v(a) \in \mathbb{Z}$ of an element $a \in R \backslash \{0\}$ is the unique natural number n such that $(a) = P^n$.

*If the above equivalent conditions hold, we say that R is a **discrete valuation ring** (short: **DVR**).*

Proof.

(a) $\Rightarrow$ (b): As R is Noetherian, every ideal I in R can be written as $I = (a_1, \dots, a_r)$ for some $a_1, \dots, a_r$ by Lemma 7.4 (c). Now if a_i has minimal valuation among these elements then $a_j \in (a_i)$ for all $j = 1, \dots, r$ by Lemma 12.6 (b), and hence $I = (a_i)$.

(b) $\Rightarrow$ (c): The maximal ideal P must be non-zero as otherwise R would be a field. So by assumption and Example 2.6 (a) it is of the form $P = (t)$ for a prime element $t \in R$. Localizing at (t) again does not change R, and so $R = R_{(t)}$ is the localization of the unique factorization domain R (see Example 8.3 (a)) at the ideal (t). By Example 12.4 (c) it is thus a valuation ring with value group $\mathbb{Z}$, which shows (c).

Moreover, every non-zero element $a \in R$ can be written uniquely as $a = ct^n$ for some $n \in \mathbb{N}$ and $c \in R \backslash P$. Then c is a unit by Lemma 12.6 (c) and (d), and hence $(a) = (t^n) = P^n$. Example 12.4 (c) shows that $v(a) = n$ in this case, proving the additional statement of the proposition.

(c) $\Rightarrow$ (a): It is clear that R cannot be a field, since otherwise its valuation group would be $(\operatorname{Quot} R)^* / R^* = \{1\}$ by Construction 12.7.

Now let $I \trianglelefteq R$ be a non-zero ideal. As the value group of R is $\mathbb{Z}$, there is a non-zero element $a \in I$ with minimal valuation $v(a)$. But then $I = (a)$, since for every element $b \in I$ we have $v(b) \geq v(a)$, and thus $b \in (a)$ by Lemma 12.6 (b). □

Proposition 12.14 (Equivalent conditions for discrete valuation rings)**.** *For a Noetherian local ring R the following statements are equivalent:*

(a) *R is a discrete valuation ring.*

(b) *R is a principal ideal domain, but not a field.*

(c) *R is a 1-dimensional unique factorization domain.*

(d) *R is a 1-dimensional normal domain.*

(e) *R is a 1-dimensional regular ring.*

Proof.

(a) $\Rightarrow$ (b) follows from Proposition 12.13.

(b) $\Rightarrow$ (c) holds by Example 8.3 (a) and Example 11.3 (c).

(c) $\Rightarrow$ (d) is Example 9.10.

(d) $\Rightarrow$ (e): As R is a local 1-dimensional domain, its maximal ideal P and the zero ideal are the only prime ideals in R. So if we choose $a \in P \backslash \{0\}$ then $\sqrt{(a)} = P$ by Lemma 2.21, and thus $P^n \subset (a)$ for some $n \in \mathbb{N}$ by Exercise 7.22 (b). Let us choose n minimal. Note that certainly $n > 0$ since otherwise a would have to be a unit. So $P^{n-1} \not\subset (a)$, and we can pick $b \in P^{n-1} \backslash (a)$.

We set $t = \frac{a}{b} \in \operatorname{Quot} R$ and claim that $t \in R$ and $P = (t)$. To see this, note first that

$$\frac{1}{t} P = \frac{b}{a} P \subset \frac{1}{a} P^n \subset \frac{1}{a} (a) \subset R,$$

so $\frac{1}{t} P$ is an ideal in R. If it was contained in the maximal ideal P, the R-module homomorphism $P \to P$, $x \mapsto \frac{1}{t} x$ would satisfy a monic polynomial with coefficients in R by Proposition 3.25. Thus $\frac{1}{t} \in \operatorname{Quot} R$ would be integral over R. As R is assumed to be normal, this would imply $\frac{1}{t} \in R$ and hence $b = \frac{1}{t} a \in (a)$, in contradiction to our choice of b. So the ideal $\frac{1}{t} P$ is not contained in the unique maximal ideal P, which means that $\frac{1}{t} P = R$. In other words we have $P = (t)$, so P is principal and hence R is a regular 1-dimensional ring by Lemma 11.36 (a).

(e) $\Rightarrow$ (a): By Lemma 12.1 we know that the maximal ideal of R is generated by one element t, and that every $a \in R \backslash \{0\}$ can be written as a unit times t^n for some $n \in \mathbb{N}$. Consequently, every element $a \in \operatorname{Quot} R \backslash \{0\}$ can be written as a unit times t^n for $n \in \mathbb{Z}$. But then $a \in R$ (if $n \geq 0$) or $a^{-1} \in R$ (if $n \leq 0$), and thus R is a valuation ring. It is discrete by Proposition 12.13 as it is Noetherian by assumption, and not a field since $\dim R = 1$. □

Remark 12.15.

(a) By Proposition 12.14, the 1-dimensional regular local rings considered in Lemma 12.1 are exactly the discrete valuation rings. Moreover, the additional statement in Proposition 12.13 shows that the "valuation" constructed in Lemma 12.1 is in fact the valuation in the sense of Definition 12.3 and Proposition 12.8.

(b) In this course we have met two conditions on rings that correspond geometrically to some sort of "non-singularity" statement: normality in Example 9.11, and regularity in Example 11.37. Proposition 12.14 states that these two conditions agree in the 1-dimensional case. One can show however that this is no longer true in higher dimensions: regular local rings are always normal, but the converse is in general false.

Example 12.16.

(a) As in Remark 12.2, every ring of local functions at a smooth point on a curve is a 1-dimensional regular local ring, and thus by Proposition 12.14 a discrete valuation ring. By Remark 12.15 (a) its valuation can be interpreted as orders of vanishing as in Remark 12.2.

(b) The power series ring $L[\![t]\!]$ over a field L is a discrete valuation ring by Proposition 12.13, since it is a valuation ring with valuation group $\mathbb{Z}$ by Example 12.4 (d).

Our results above allow to give a very easy description of the ideals in a discrete valuation ring.

Corollary 12.17 (Ideals in a discrete valuation ring)**.** *Let R be a discrete valuation ring with unique maximal ideal P. Then the non-zero ideals of R are exactly the powers P^n for $n \in \mathbb{N}$. They form a strictly decreasing chain*

$$R = P^0 \supsetneq P^1 \supsetneq P^2 \supsetneq \cdots .$$

Proof. By Proposition 12.13, any non-zero ideal in R is principal and of the form P^n for some $n \in \mathbb{N}$. Moreover, these powers form a strict chain of ideals since $P^n = P^{n+1} = P \cdot P^n$ for some n would imply $P^n = 0$ by Nakayama's Lemma as in Example 6.16 (a), and thus the contradiction $P = 0$. □

Finally, the valuation in a discrete valuation ring can also be expressed in terms of the length of modules introduced in Definition 3.18.

Corollary 12.18 (Valuation and length)**.** *Let R be a discrete valuation ring, and let $a \in R \backslash \{0\}$. Then $\nu(a) = l_R(R/(a))$.*

Proof. By Proposition 12.13 (b) the maximal ideal P of R can be generated by one element t. Moreover, Proposition 12.13 shows that $(a) = P^{\nu(a)} = (t^{\nu(a)})$. Now

$$0 = (t^{\nu(a)})/(a) \subsetneq (t^{\nu(a)-1})/(a) \subsetneq \cdots \subsetneq (t^0)/(a) = R/(a) \qquad (*)$$

is a chain of submodules of $R/(a)$ with successive quotients $(t^{i-1})/(t^i)$ for $i = 1, \ldots, \nu(a)$ by Proposition 3.10 (b). But

$$R/(t) \to (t^{i-1})/(t^i), \ \overline{c} \mapsto \overline{ct^{i-1}}$$

is an isomorphism for all i, and thus $(*)$ is a composition series for $R/(a)$ as an R-module. Hence we conclude that $l_R(R/(a)) = \nu(a)$. □

13. Dedekind Domains

In the last chapter we have mainly studied 1-dimensional regular local rings, i. e. geometrically the local properties of smooth points on curves. We now want to patch these local results together to obtain global statements about 1-dimensional rings (resp. curves) that are "locally regular". The corresponding notion is that of a Dedekind domain.

Definition 13.1 (Dedekind domains)**.** An integral domain R is called **Dedekind domain** if it is Noetherian of dimension 1, and for all maximal ideals $P \trianglelefteq R$ the localization R_P is a regular local ring.

Remark 13.2 (Equivalent conditions for Dedekind domains)**.** As a Dedekind domain R is an integral domain of dimension 1, its prime ideals are exactly the zero ideal and all maximal ideals. So every localization R_P for a maximal ideal P is a 1-dimensional local ring. As these localizations are also Noetherian by Exercise 7.23, we can replace the requirement in Definition 13.1 that the local rings R_P are regular by any of the equivalent conditions in Proposition 12.14. For example, a Dedekind domain is the same as a 1-dimensional Noetherian domain such that all localizations at maximal ideals are discrete valuation rings.

This works particularly well for the normality condition as this is a local property and can thus be transferred to the whole ring:

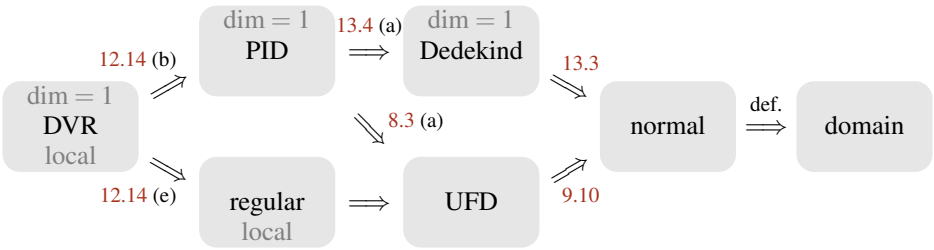
Lemma 13.3. *A 1-dimensional Noetherian domain is a Dedekind domain if and only if it is normal.*

Proof. By Remark 13.2 and Proposition 12.14, a 1-dimensional Noetherian domain R is a Dedekind domain if and only if all localizations R_P at a maximal ideal P are normal. But by Exercise 9.13 (c) this is equivalent to R being normal. □

25

Example 13.4.

(a) By Lemma 13.3, any principal ideal domain which is not a field is a Dedekind domain: it is 1-dimensional by Example 11.3 (c), clearly Noetherian, and normal by Example 9.10 since it is a unique factorization domain by Example 8.3 (a). For better visualization, the following diagram shows the implications between various properties of rings for the case of integral domains that are not fields. Rings that are always 1-dimensional and / or local are marked as such. It is true that every regular local ring is a unique factorization domain, but we have not proven this here since this requires more advanced methods — we have only shown in Proposition 11.40 that any regular local ring is an integral domain.



(b) Let X be an irreducible curve over an algebraically closed field. Assume that X is smooth, i. e. that all points of X are smooth in the sense of Example 11.37 and Definition 11.38. Then the coordinate ring $A(X)$ is a Dedekind domain: it is an integral domain by Lemma 2.3 (a) since $I(X)$ is a prime ideal by Remark 2.7 (b). It is also 1-dimensional by assumption and

Noetherian by Remark 7.15. Moreover, by Hilbert's Nullstellensatz as in Remark 10.11 the maximal ideals in $A(X)$ are exactly the ideals of points, and so our smoothness assumption is the same as saying that all localizations at maximal ideals are regular.

In fact, irreducible smooth curves over algebraically closed fields are the main geometric examples for Dedekind domains. However, there is also a large class of examples in number theory, which explains why the concept of a Dedekind domain is equally important in number theory and geometry: it turns out that the ring of integral elements in a number field, i. e. in a finite field extension of $\mathbb{Q}$, is always a Dedekind domain. Let us prove this now.

Proposition 13.5 (Integral elements in number fields)**.** *Let $\mathbb{Q} \subset K$ be a finite field extension, and let R be the integral closure of $\mathbb{Z}$ in K. Then R is a Dedekind domain.*

Proof. As a subring of a field, R is clearly an integral domain. Moreover, by Example 11.3 (c) and Lemma 11.8 we have $\dim R = \dim \mathbb{Z} = 1$. It is also easy to see that R is normal: if $a \in \operatorname{Quot} R \subset K$ is integral over R it is also integral over $\mathbb{Z}$ by transitivity as in Lemma 9.6 (b), so it is contained in the integral closure R of $\mathbb{Z}$ in K. Hence by Lemma 13.3 it only remains to show that R is Noetherian — which is in fact the hardest part of the proof. We will show this in three steps.

(a) We claim that $|R/pR| < \infty$ for all prime numbers $p \in \mathbb{Z}$.

Note that R/pR is a vector space over $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$. It suffices to show that $\dim_{\mathbb{Z}_p} R/pR \leq \dim_{\mathbb{Q}} K$ since this dimension is finite by assumption. So let $\overline{a_1}, \dots, \overline{a_n} \in R/pR$ be linearly independent over $\mathbb{Z}_p$. We will show that $a_1, \dots, a_n \in K$ are also independent over $\mathbb{Q}$, so that $n \leq \dim_{\mathbb{Q}} K$. Otherwise there are $\lambda_1, \dots, \lambda_n \in \mathbb{Q}$ not all zero with $\lambda_1 a_1 + \dots + \lambda_n a_n = 0$. After multiplying these coefficients with a common scalar we may assume that all of them are integers, and not all of them are divisible by p. But then $\overline{\lambda_1}\,\overline{a_1} + \dots + \overline{\lambda_n}\,\overline{a_n} = 0$ is a non-trivial relation in R/pR with coefficients in $\mathbb{Z}_p$, in contradiction to $\overline{a_1}, \dots, \overline{a_n}$ being independent over $\mathbb{Z}_p$.

(b) We will show that $|R/mR| < \infty$ for all $m \in \mathbb{Z}\backslash\{0\}$.

In fact, this follows by induction on the number of prime factors in m: for one prime factor the statement is just that of (a), and for more prime factors it follows from the exact sequence of Abelian groups

$$0 \longrightarrow R/m_1R \xrightarrow{\cdot m_2} R/m_1m_2R \longrightarrow R/m_2R \longrightarrow 0,$$

since this means that $|R/m_1m_2R| = |R/m_1R| \cdot |R/m_2R| < \infty$.

(c) Now let $I \trianglelefteq R$ be any non-zero ideal. We claim that $m \in I$ for some $m \in \mathbb{Z}\backslash\{0\}$.

Otherwise we would have

$$\dim R/I = \dim \mathbb{Z}/(I \cap \mathbb{Z}) = \dim \mathbb{Z} = 1$$

by Lemma 11.8, since R/I is integral over $\mathbb{Z}/(I \cap \mathbb{Z})$ by Lemma 9.7 (a). But $\dim R$ has to be bigger than $\dim R/I$, since a chain of prime ideals in R/I corresponds to a chain of prime ideals in R containing I, which can always be extended to a longer chain by the zero ideal since R is an integral domain. Hence $\dim R > 1$, a contradiction.

Putting everything together, we can choose a non-zero $m \in I \cap \mathbb{Z}$ by (c), so that $mR \trianglelefteq I$. Hence $|I/mR| \leq |R/mR| < \infty$ by (b), so $I/mR = \{\overline{a_1}, \dots, \overline{a_n}\}$ for some $a_1, \dots, a_n \in I$. But then the ideal $I = (a_1, \dots, a_n, m)$ is finitely generated, and hence R is Noetherian. □

Example 13.6. Consider again the ring $R = \mathbb{Z}[\sqrt{5}\,i]$ of Example 8.3 (b). By Example 9.16, it is the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{5}\,i)$. Hence Proposition 13.5 shows that R is a Dedekind domain.

We see from this example that a Dedekind domain is in general not a unique factorization domain, as e. g. by Example 8.3 (b) the element 2 is irreducible, but not prime in R, so that it does not have a factorization into prime elements. However, we will prove now that a Dedekind domain always has an analogue of the unique factorization property for *ideals*, i. e. every non-zero ideal can be written uniquely as a product of non-zero prime ideals (which are then also maximal since Dedekind

domains are 1-dimensional). In fact, this is the most important property of Dedekind domains in practice.

Proposition 13.7 (Prime factorization of ideals in Dedekind domains)**.** *Let R be a Dedekind domain.*

(a) *Let $P \trianglelefteq R$ be a maximal ideal, and let $Q \trianglelefteq R$ be any ideal. Then*

$$Q \text{ is } P\text{-primary} \quad \Leftrightarrow \quad Q = P^k \text{ for some } k \in \mathbb{N}_{>0}.$$

Moreover, the number k is unique in this case.

(b) *Any non-zero ideal $I \trianglelefteq R$ has a "prime factorization"*

$$I = P_1^{k_1} \cdot \cdots \cdot P_n^{k_n}$$

with $k_1, \dots, k_n \in \mathbb{N}_{>0}$ and distinct maximal ideals $P_1, \dots, P_n \trianglelefteq R$. It is unique up to permutation of the factors, and $P_1, \dots, P_n$ are exactly the associated prime ideals of I.

Proof.

(a) The implication "$\Leftarrow$" holds in arbitrary rings by Lemma 8.12 (b), so let us show the opposite direction "$\Rightarrow$". Let Q be P-primary, and consider the localization map $R \to R_P$. Then Q^e is a non-zero ideal in the localization R_P, which is a discrete valuation ring by Remark 13.2. So by Corollary 12.17 we have $Q^e = (P^e)^k$ for some k, and hence $Q^e = (P^k)^e$ as extension commutes with products by Exercise 1.19 (c). Contracting this equation now gives $Q = P^k$ by Lemma 8.33, since Q and P^k are both P-primary by Lemma 8.12 (b).

The number k is unique since $P^k = P^l$ for $k \neq l$ would imply $(P^e)^k = (P^e)^l$ by extension, in contradiction to Corollary 12.17.

(b) As R is Noetherian, the ideal I has a minimal primary decomposition $I = Q_1 \cap \cdots \cap Q_n$ by Corollary 8.21. Since I is non-zero, the corresponding associated prime ideals $P_1, \dots, P_n$ of these primary ideals are distinct and non-zero, and hence maximal as $\dim R = 1$. In particular, there are no strict inclusions among the ideals $P_1, \dots, P_n$, and thus all of them are minimal over I. By Proposition 8.34 this means that the ideals $Q_1, \dots, Q_n$ in our decomposition are unique.

Now by (a) we have $Q_i = P_i^{k_i}$ for unique $k_i \in \mathbb{N}_{>0}$ for $i = 1, \dots, n$. This gives us a unique decomposition $I = P_1^{k_1} \cap \cdots \cap P_n^{k_n}$, and thus also a unique factorization $I = P_1^{k_1} \cdot \cdots \cdot P_n^{k_n}$ by Exercise 1.8 since the ideals $P_1^{k_1}, \dots, P_n^{k_n}$ are pairwise coprime by Exercise 2.24. □

Example 13.8. Recall from Examples 8.3 (b) and 13.6 that the element 2 in the Dedekind domain $R = \mathbb{Z}[\sqrt{5}\,i]$ does not admit a factorization into prime elements. But by Proposition 13.7 (b) the ideal (2) must have a decomposition as a product of maximal ideals (which cannot all be principal, as otherwise we would have decomposed the number 2 into prime factors). Concretely, we claim that this decomposition is

$$(2) = (2, 1 + \sqrt{5}\,i)^2.$$

To see this, note first that the ideal $(2, 1 + \sqrt{5}\,i)$ is maximal by Lemma 2.3 (b) since the quotient

$$\mathbb{Z}[\sqrt{5}\,i]/(2, 1 + \sqrt{5}\,i) \;\cong\; \mathbb{Z}/(2) \;\cong\; \mathbb{Z}_2$$

is a field. Moreover, we have $(2) \subset (2, 1 + \sqrt{5}\,i)^2$ since

$$2 = (1 + \sqrt{5}\,i)^2 - 2^2 - 2\sqrt{5}\,i(1 + \sqrt{5}\,i) \quad \in (2, 1 + \sqrt{5}\,i)^2,$$

and $(2, 1 + \sqrt{5}\,i)^2 \subset (2)$ as

$$2^2 \in (2), \quad 2(1 + \sqrt{5}\,i) \in (2), \quad \text{and} \;\; (1 + \sqrt{5}\,i)^2 = -4 + 2\sqrt{5}\,i \in (2).$$

To understand the geometric meaning of the prime factorization of ideals we need a lemma first.

Lemma 13.9 (Ideals in Dedekind domains)**.** *Let R be a Dedekind domain.*

(a) *For all distinct maximal ideals $P_1,\dots,P_n$ of R and $k_1,\dots,k_n,l_1,\dots,l_n \in \mathbb{N}$ we have*

$$P_1^{k_1} \cdot \dots \cdot P_n^{k_n} \subset P_1^{l_1} \cdot \dots \cdot P_n^{l_n} \quad \Leftrightarrow \quad l_i \leq k_i \text{ for all } i = 1,\dots,n.$$

(b) *For any $a \in R\backslash\{0\}$ we have*

$$(a) = P_1^{\nu_1(a)} \cdot \dots \cdot P_n^{\nu_n(a)},$$

where $P_1,\dots,P_n$ are the associated prime ideals of (a), and ν_i denotes the valuation of the discrete valuation ring R_{P_i} (restricted to R).

Proof. By Exercise 6.29 (a) ideal containment is a local property, i. e. we can check it on all localizations R_P for maximal ideals P. Moreover, products commute with localization by Exercise 1.19 (c), and the localization of P_i at a maximal ideal $P \neq P_i$ is the unit ideal by Example 6.25 (a). Hence:

(a) We have

$$\begin{aligned} P_1^{k_1} \cdot \dots \cdot P_n^{k_n} \subset P_1^{l_1} \cdot \dots \cdot P_n^{l_n} &\Leftrightarrow (P_i^e)^{k_i} \subset (P_i^e)^{l_i} \text{ in } R_{P_i} \text{ for all } i \\ &\Leftrightarrow l_i \leq k_i \text{ for all } i. \qquad \text{(Corollary 12.17)} \end{aligned}$$

(b) By Proposition 13.7 (b) we know that $(a) = P_1^{k_1} \cdot \dots \cdot P_n^{k_n}$ for suitable $k_1,\dots,k_n \in \mathbb{N}$ if $P_1,\dots,P_n$ are the associated prime ideals of (a). To determine the exponent k_i for $i = 1,\dots,n$, we localize at R_{P_i} to get $(a) = (P_i^e)^{k_i}$ in the discrete valuation ring R_{P_i}, and use Proposition 12.13 to conclude from this that $k_i = \nu_i(a)$. □

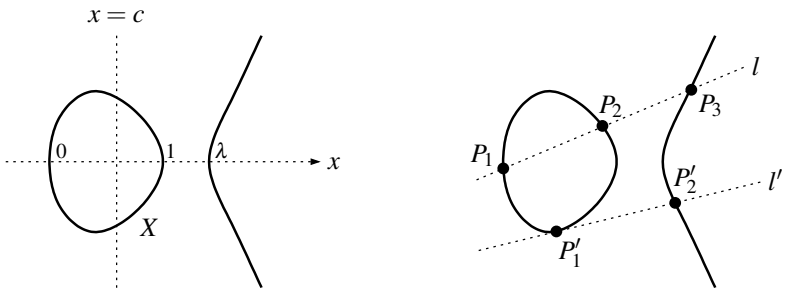
Remark 13.10. If a is a non-zero element in a Dedekind domain R and $P \trianglelefteq R$ a maximal ideal that is not an associated prime ideal of (a), the same argument as in the proof of Lemma 13.9 (b) shows that the valuation of a in the discrete valuation ring R_P is 0. Hence the ideals $P_1,\dots,P_n$ in the statement of this lemma are exactly the maximal ideals of R so that the valuation of a in the corresponding discrete valuation ring is non-zero.

Remark 13.11 (Geometric interpretation of the prime factorization of ideals)**.** Let X be an irreducible smooth curve over an algebraically closed field, so that its coordinate ring $R = A(X)$ is a Dedekind domain by Example 13.4 (b). Now let $f \in R$ be a non-zero polynomial function on X, and let $a_1,\dots,a_n \in X$ be the zeroes of f, with corresponding maximal ideals $P_1,\dots,P_n \trianglelefteq R$. Moreover, for $i = 1,\dots,n$ let k_i be the order of vanishing of f at a_i as in Remark 12.2, i. e. the valuation of f in the ring R_{P_i} of local functions on X at a_i. Then Lemma 13.9 (b) (together with Remark 13.10) states that

$$(f) = P_1^{k_1} \cdot \dots \cdot P_n^{k_n}.$$

In other words, the prime factorization of a principal ideal (f) in the coordinate ring of X encodes the orders of vanishing of the function f at all points of X. Here is a concrete example of this construction that will also be used later on in Example 13.29.

Example 13.12. Consider the complex plane cubic curve $X = V(y^2 - x(x-1)(x-\lambda)) \subset \mathbb{A}^2_{\mathbb{C}}$ for some $\lambda \in \mathbb{C}\backslash\{0,1\}$. The picture (a) below shows approximately the real points of X in the case $\lambda \in \mathbb{R}_{>1}$: the vertical line $x = c$ for $c \in \mathbb{R}$ intersects X in two real points symmetric with respect to this axis if $0 < c < 1$ or $c > \lambda$, in exactly the point $(c,0)$ if $c \in \{0,1,\lambda\}$, and in no real point in all other cases.



(a) (b)

It is easy to check as in Example 11.39 (b) that all points of X are smooth, so that the coordinate ring $R = \mathbb{C}[x,y]/(y^2 - x(x-1)(x-\lambda))$ of X is a Dedekind domain by Example 13.4 (b).

Now let $l \in R$ be a general linear function as in picture (b) above. On the curve X it vanishes at three points (to order 1) since the cubic equation $y^2 = x(x-1)(x-\lambda)$ together with a general linear equation in x and y will have three solutions. If P_1, P_2, and P_3 are the maximal ideals in R corresponding to these points, Remark 13.11 shows that $(l) = P_1 \cdot P_2 \cdot P_3$ in R.

Note that for special linear functions it might happen that some of these points coincide, as in the case of l' above which vanishes to order 2 at the point P_1'. Consequently, in this case we get $(l') = P_1'^2 \cdot P_2$.

For computational purposes, the unique factorization property for ideals allows us to perform calculations with ideals in Dedekind domains very much in the same way as in principal ideal domains. For example, the following Proposition 13.13 is entirely analogous (both in its statement and in its proof) to Example 1.4.

Proposition 13.13 (Operations on ideals in Dedekind domains)**.** *Let I and J be two non-zero ideals in a Dedekind domain, with prime factorizations*

$$I = P_1^{k_1} \cdot \cdots \cdot P_n^{k_n} \quad \text{and} \quad J = P_1^{l_1} \cdot \cdots \cdot P_n^{l_n}$$

as in Proposition 13.7, where $P_1, \dots, P_n$ are distinct maximal ideals and $k_1, \dots, k_n, l_1, \dots, l_n \in \mathbb{N}$. Then

$$\begin{aligned} I + J &= P_1^{m_1} \cdot \cdots \cdot P_n^{m_n} && \text{with } m_i = \min(k_i, l_i), \\ I \cap J &= P_1^{m_1} \cdot \cdots \cdot P_n^{m_n} && \text{with } m_i = \max(k_i, l_i), \\ I \cdot J &= P_1^{m_1} \cdot \cdots \cdot P_n^{m_n} && \text{with } m_i = k_i + l_i \end{aligned}$$

for $i = 1, \dots, n$. In particular, $I \cdot J = (I+J) \cdot (I \cap J)$.

Proof. By Proposition 13.7 (b) we can write all three ideals as $P_1^{m_1} \cdot \cdots \cdot P_n^{m_n}$ for suitable $m_1, \dots, m_n$ (if we possibly enlarge the set of maximal ideals occurring in the factorizations). So it only remains to determine the numbers $m_1, \dots, m_n$ for the three cases.

The ideal $I + J$ is the smallest ideal containing both I and J. By Lemma 13.9 (a) this means that m_i is the biggest number less than or equal to both k_i and l_i, i. e. $\min(k_i, l_i)$. Analogously, the intersection $I \cap J$ is the biggest ideal contained in both I and J, so in this case m_i is the smallest number greater than or equal to both k_i and l_i, i. e. $\max(k_i, l_i)$. The exponents $m_i = k_i + l_i$ for the product are obvious. □

As a Dedekind domain is in general not a unique factorization domain, it clearly follows from Example 8.3 (a) that it is usually not a principal ideal domain either. However, a surprising result following from the computational rules in Proposition 13.13 is that every ideal in a Dedekind domain can be generated by *two* elements. In fact, there is an even stronger statement:

Proposition 13.14. *Let R be a Dedekind domain, and let a be a non-zero element in an ideal $I \trianglelefteq R$. Then there is an element $b \in R$ such that $I = (a, b)$.*

In particular, every ideal in R can be generated by two elements.

Proof. By assumption $(a) \subset I$ are non-zero ideals, so we know by Proposition 13.7 (b) and Lemma 13.9 (a) that

$$(a) = P_1^{k_1} \cdot \dots \cdot P_n^{k_n} \qquad \text{and} \qquad I = P_1^{l_1} \cdot \dots \cdot P_n^{l_n}$$

for suitable distinct maximal ideals $P_1, \dots, P_n \trianglelefteq R$ and natural numbers $l_i \leq k_i$ for all $i = 1, \dots, n$. By the uniqueness part of Proposition 13.7 (b) we can pick elements

$$b_i \in P_1^{l_1+1} \cdot \dots \cdot P_i^{l_i} \cdot \dots \cdot P_n^{l_n+1} \setminus P_1^{l_1+1} \cdot \dots \cdot P_i^{l_i+1} \cdot \dots \cdot P_n^{l_n+1} \quad \subset I$$

for all i. Then $b_i \in P_j^{l_j+1}$ for all $j \neq i$, but $b_i \notin P_i^{l_i+1}$, since otherwise by Proposition 13.13

$$b_i \in P_i^{l_i+1} \cap \left(P_1^{l_1+1} \cdot \dots \cdot P_i^{l_i} \cdot \dots \cdot P_n^{l_n+1}\right) = P_1^{l_1+1} \cdot \dots \cdot P_n^{l_n+1}$$

in contradiction to our choice of b_i. Hence

$$b := b_1 + \dots + b_n \quad \notin P_i^{l_i+1}$$

for all i, but certainly $b \in I$. We now claim that $I = (a,b)$. To see this, note first that by Proposition 13.13 the prime factorization of $(a,b) = (a) + (b)$ can contain at most the maximal ideals occurring in (a), so we can write $(a,b) = P_1^{m_1} \cdot \dots \cdot P_n^{m_n}$ for suitable $m_1, \dots, m_n$. But by Lemma 13.9 (a) we see that:

- $l_i \leq m_i$ for all i since $(a,b) \subset I$;
- $m_i \leq l_i$ for all i since $b \notin P_i^{l_i+1}$, and hence $(a,b) \not\subset P_i^{l_i+1}$.

Therefore we get $m_i = l_i$ for all i, which means that $I = (a,b)$. □

26

We have now studied prime factorizations of ideals in Dedekind domains in some detail. However, recall that the underlying valuations on the local rings are defined originally not only on these discrete valuation rings, but also on their quotient field. Geometrically, this means that we can equally well consider orders of rational functions, i. e. quotients of polynomials, at a smooth point of a curve. These orders can then be positive (if the function has a zero), negative (if it has a pole), or zero (if the function has a non-zero value at the given point). Let us now transfer this extension to the quotient field to the global case of a Dedekind domain R. Instead of ideals we then have to consider corresponding structures (i. e. R-submodules) that do not lie in R itself, but in its quotient field $\operatorname{Quot} R$.

Definition 13.15 (Fractional ideals)**.** Let R be an integral domain with quotient field $K = \operatorname{Quot} R$.

(a) A **fractional ideal** of R is an R-submodule I of K such that $aI \subset R$ for some $a \in R \setminus \{0\}$.

(b) For $a_1, \dots, a_n \in K$ we set as expected

$$(a_1, \dots, a_n) := Ra_1 + \dots + Ra_n$$

and call this the fractional ideal generated by $a_1, \dots, a_n$ (note that this is in fact a fractional ideal since we can take for a in (a) the product of the denominators of $a_1, \dots, a_n$).

Example 13.16.

(a) A subset I of an integral domain R is a fractional ideal of R if and only if it is an ideal in R (the condition in Definition 13.15 (a) that $aI \subset R$ for some a is vacuous in this case since we can always take $a = 1$).

(b) $\left(\frac{1}{2}\right) = \frac{1}{2}\mathbb{Z} \subset \mathbb{Q}$ is a fractional ideal of $\mathbb{Z}$. In contrast, the localization $\mathbb{Z}_{(2)} \subset \mathbb{Q}$ of Example 6.5 (d) is not a fractional ideal of $\mathbb{Z}$: it is a $\mathbb{Z}$-submodule of $\mathbb{Q}$, but there is no non-zero integer a such that $a\mathbb{Z}_{(2)} \subset \mathbb{Z}$.

Remark 13.17. Let R be an integral domain with quotient field K.

(a) Let I be a fractional ideal of R. The condition $aI \subset R$ of Definition 13.15 (a) ensures that I is finitely generated if R is Noetherian: as aI is an R-submodule in R it is actually an ideal in R, and hence of the form $aI = (a_1, \dots, a_n)$ for some $a_1, \dots, a_n \in R$. But then also $I = \left(\frac{a_1}{a}, \dots, \frac{a_n}{a}\right)$ is finitely generated.

(b) The standard operations on ideals of Construction 1.1 can easily be extended to fractional ideals, or more generally to R-submodules of K. In the following, we will mainly need products and quotients: for two R-submodules I and J of K we set

$$IJ := \left\{ \sum_{i=1}^{n} a_i b_i : n \in \mathbb{N},\ a_1,\dots,a_n \in I,\ b_1,\dots,b_n \in J \right\},$$
$$I :_K J := \{a \in K : aJ \subset I\}.$$

Note that IJ is just the smallest R-submodule of K containing all products ab for $a \in I$ and $b \in J$, as expected. The index K in the notation of the quotient $I :_K J$ distinguishes this construction from the ordinary ideal quotient $I : J = \{a \in R : aJ \subset I\}$ — note that both quotients are defined but different in general if both I and J are ordinary ideals in R.

Exercise 13.18. Let K be the quotient field of a Noetherian integral domain R. Prove that for any two fractional ideals I and J of R and any multiplicatively closed subset $S \subset R$ we have:

(a) $S^{-1}(IJ) = S^{-1}I \cdot S^{-1}J$,

(b) $S^{-1}(I :_K J) = S^{-1}I :_K S^{-1}J$.

Our goal in the following will be to check whether the multiplication as in Remark 13.17 (b) defines a group structure on the set of all non-zero fractional ideals of an integral domain R. As associativity and the existence of the neutral element R are obvious, the only remaining question is the existence of inverse elements, i. e. whether for a given non-zero fractional ideal I there is always another fractional ideal J with $IJ = R$. We will see now that this is indeed the case for Dedekind domains, but not in general integral domains.

Definition 13.19 (Invertible and principal ideals)**.** Let R be an integral domain, and let I be an R-submodule of $K = \operatorname{Quot} R$.

(a) I is called an **invertible ideal** or **(Cartier) divisor** if there is an R-submodule of J of K such that $IJ = R$.

(b) I is called **principal** if $I = (a)$ for some $a \in K$.

Lemma 13.20. *As above, let K be the quotient field of an integral domain R, and let I be an R-submodule of K.*

(a) *If I is an invertible ideal with $IJ = R$, then $J = R :_K I$.*

(b) *We have the implications*

$$I \text{ non-zero principal} \quad \Rightarrow \quad I \text{ invertible} \quad \Rightarrow \quad I \text{ fractional}.$$

Proof.

(a) By definition of the quotient, $IJ = R$ implies $R = IJ \subset I(R :_K I) \subset R$, so we have equality $IJ = I(R :_K I)$. Multiplication by J now gives the desired result $J = R :_K I$.

(b) If $I = (a)$ is principal with $a \in K^*$ then $(a) \cdot \left(\frac{1}{a}\right) = R$, hence I is invertible.

Now let I be an invertible ideal, i. e. $I(R :_K I) = R$ by (a). This means that

$$\sum_{i=1}^{n} a_i b_i = 1$$

for some $n \in \mathbb{N}$ and $a_1,\dots,a_n \in I$ and $b_1,\dots,b_n \in R :_K I$. Then we have for all $b \in I$

$$b = \sum_{i=1}^{n} a_i \underbrace{b_i b}_{\in R}.$$

So if we let $a \in R$ be the product of the denominators of $a_1,\dots,a_n$, we get $ab \in R$. Therefore $aI \subset R$, i. e. I is fractional. □

Example 13.21.

(a) Let R be a principal ideal domain. Then every non-zero fractional ideal I of R is principal: we have $aI \subset R$ for some $a \in \operatorname{Quot} R \backslash \{0\}$. This is an ideal in R, so of the form (b) for some $b \in R \backslash \{0\}$. It follows that $I = \left(\frac{b}{a}\right)$, i. e. I is principal.

In particular, Lemma 13.20 (b) implies that the notions of principal, invertible, and fractional ideals all agree for non-zero ideals in a principal ideal domain.

(b) The ideal $I = (x,y)$ in the ring $R = \mathbb{R}[x,y]$ is not invertible: setting $K = \operatorname{Quot} R = \mathbb{R}(x,y)$ we have

$$R :_K I = \{f \in \mathbb{R}(x,y) : xf \in \mathbb{R}[x,y] \text{ and } yf \in \mathbb{R}[x,y]\} = \mathbb{R}[x,y].$$

But $I(R :_K I) = (x,y)\mathbb{R}[x,y] \neq R$, and hence I is not invertible by Lemma 13.20 (a).

Proposition 13.22 (Invertible = fractional ideals in Dedekind domains)**.** *In a Dedekind domain, every non-zero fractional ideal is invertible.*

Proof. Let I be a non-zero fractional ideal of a Dedekind domain R. Assume that I is not invertible, which means by Lemma 13.20 (a) that $I(R :_K I) \neq R$. As the inclusion $I(R :_K I) \subset R$ is obvious, this means that $I(R :_K I)$ is a proper ideal of R. It must therefore be contained in a maximal ideal P by Corollary 2.17.

Extending this inclusion by the localization map $R \to R_P$ then gives $I^e(R_P :_K I^e) \subset P^e$ by Exercise 13.18. This means by Lemma 13.20 (a) that I^e is not invertible in R_P. But R_P is a discrete valuation ring by Remark 13.2, hence a principal ideal domain by Proposition 12.14, and so I^e cannot be a fractional ideal either by Example 13.21 (a). This is clearly a contradiction, since I is assumed to be fractional. □

Remark 13.23. By construction, the invertible ideals of an integral domain R form an Abelian group under multiplication, with neutral element R. As expected, we will write the inverse $R :_K I$ of an invertible ideal I as in Lemma 13.20 (a) also as I^{-1}. Proposition 13.22 tells us that for Dedekind domains this group of invertible ideals can also be thought of as the group of non-zero fractional ideals.

Moreover, it is obvious that the non-zero principal fractional ideals form a subgroup:

- every non-zero principal fractional ideal is invertible by Lemma 13.20 (b);
- the neutral element $R = (1)$ is principal;
- for two non-zero principal fractional ideals (a) and (b) their product (ab) is principal;
- for any non-zero principal fractional ideal (a) its inverse (a^{-1}) is also principal.

So we can define the following groups that are naturally attached to any integral domain.

Definition 13.24 (Ideal class groups)**.** Let R be an integral domain.

(a) The group of all invertible ideals of R (under multiplication) is called the **ideal group** or **group of (Cartier) divisors** of R. We denote it by $\operatorname{Div} R$.

(b) We denote by $\operatorname{Prin} R \leq \operatorname{Div} R$ the subgroup of (non-zero) principal ideals.

(c) The quotient $\operatorname{Pic} R := \operatorname{Div} R / \operatorname{Prin} R$ of all invertible ideals modulo principal ideals is called the **ideal class group**, or **group of (Cartier) divisor classes**, or **Picard group** of R.

Let us restrict the study of these groups to Dedekind domains. In this case, the structure of the ideal group is easy to understand with the following proposition.

Proposition 13.25 (Prime factorization for invertible ideals)**.** *Let I be an invertible ideal in a Dedekind domain R. Then $I = P_1^{k_1} \cdot \dots \cdot P_n^{k_n}$ for suitable distinct maximal ideals $P_1, \dots, P_n$ and $k_1, \dots, k_n \in \mathbb{Z}$, and this representation is unique up to permutation of the factors.*

Proof. By Lemma 13.20 (b) we know that aI is an ideal in R for a suitable $a \in R\backslash\{0\}$. Now by Proposition 13.7 (b) we have $aI = P_1^{r_1} \cdot \dots \cdot P_n^{r_n}$ and $(a) = P_1^{s_1} \cdot \dots \cdot P_n^{s_n}$ for suitable distinct maximal ideals $P_1, \dots, P_n$ and $r_1, \dots, r_n, s_1, \dots, s_n \in \mathbb{N}$, and so we get a factorization

$$I = (a)^{-1} \cdot aI = P_1^{r_1 - s_1} \cdot \dots \cdot P_n^{r_n - s_n}$$

as desired. Moreover, if we have two such factorizations $P_1^{k_1} \cdot \dots \cdot P_n^{k_n} = P_1^{l_1} \cdot \dots \cdot P_n^{l_n}$ with $k_1, \dots, k_n, l_1, \dots, l_n \in \mathbb{Z}$, we can multiply this equation with suitable powers of $P_1, \dots, P_n$ so that the exponents become non-negative. The uniqueness statement then follows from the corresponding one in Proposition 13.7 (b). □

Remark 13.26. Let R be a Dedekind domain. Proposition 13.25 states that the ideal group $\operatorname{Div} R$ is in fact easy to describe: we have an isomorphism

$$\operatorname{Div} R \to \{\varphi : \operatorname{mSpec} R \to \mathbb{Z} : \varphi \text{ is non-zero only on finitely many maximal ideals}\}$$

sending any invertible ideal $P_1^{k_1} \cdot \dots \cdot P_n^{k_n}$ to the map $\varphi : \operatorname{mSpec} R \to \mathbb{Z}$ with only non-zero values $\varphi(P_i) = k_i$ for $i = 1, \dots, n$. This is usually called the *free Abelian group* generated by $\operatorname{mSpec} R$ (since the maximal ideals generate this group, and there are no non-trivial relations among these generators).

The group $\operatorname{Div} R$ is therefore very "big", and also at the same time not very interesting since its structure is so simple. In contrast, the ideal class group $\operatorname{Pic} R = \operatorname{Div} R / \operatorname{Prin} R$ is usually much smaller, and contains a lot of information on R. It is of great importance both in geometry and number theory. It is out of the scope of this course to study it in detail, but we will at least give one interesting example in each of these areas. But first let us note that the ideal class group can be thought of as measuring "how far away R is from being a principal ideal domain":

Proposition 13.27. *For a Dedekind domain R the following statements are equivalent:*

(a) *R is a principal ideal domain.*

(b) *R is a unique factorization domain.*

(c) *$\operatorname{Pic} R$ is the trivial group, i. e. $|\operatorname{Pic} R| = 1$.*

Proof.

(a) $\Rightarrow$ (b) is Example 8.3 (a).

(b) $\Rightarrow$ (c): By Exercise 8.32 (b) every maximal ideal $P \trianglelefteq R$ (which is also a minimal non-zero prime ideal as $\dim R = 1$) is principal. But these maximal ideals generate the group $\operatorname{Div} R$ by Proposition 13.25, and so we have $\operatorname{Prin} R = \operatorname{Div} R$, i. e. $|\operatorname{Pic} R| = 1$.

(c) $\Rightarrow$ (a): By Definition 13.24 the assumption $|\operatorname{Pic} R| = 1$ means that every invertible ideal is principal. But every non-zero ideal of R is invertible by Proposition 13.22, so the result follows. □

Example 13.28 (A non-trivial Picard group in number theory)**.** Let $R = \mathbb{Z}[\sqrt{5}i]$ be the integral closure of $\mathbb{Z}$ in $K = \mathbb{Q}(\sqrt{5}i)$ as in Examples 8.3 (b) and 13.6. We have seen there already that R is a Dedekind domain but not a principal ideal domain: the ideal $I_1 := (2, 1 + \sqrt{5}i)$ is not principal. In particular, the class of I_1 in the Picard group $\operatorname{Pic} R$ is non-trivial. We will now show that this is the only non-trivial element in $\operatorname{Pic} R$, i. e. that $|\operatorname{Pic} R| = 2$ and thus necessarily $\operatorname{Pic} R \cong \mathbb{Z}/2\mathbb{Z}$ as a group, with the two elements given by the classes of the ideals $I_0 := (1)$ and I_1. Unwinding Definition 13.24, we therefore claim that every invertible ideal of R is of the form aI_0 or aI_1 for some $a \in K^*$.

So let I be an invertible ideal of R. Then I is also a non-zero fractional ideal by Lemma 13.20 (b), and thus there is a number $b \in K^*$ with $bI \subset R$. But note that

$$R = \mathbb{Z}[\sqrt{5}i] = \{m + n\sqrt{5}i : m, n \in \mathbb{Z}\}$$

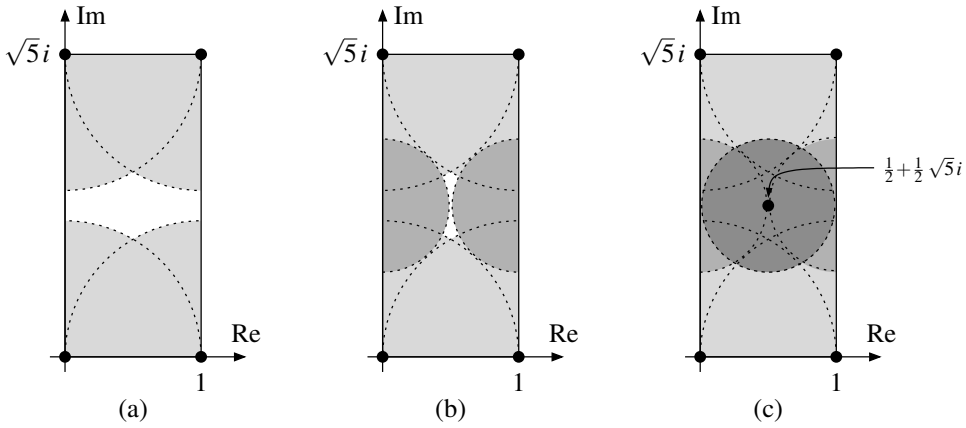
is just a rectangular lattice in the complex plane, and so there is an element $c \in bI \subset K$ of minimal non-zero absolute value. Replacing I by $\frac{b}{c}I$ (which is an equivalent element in the Picard group) we

can therefore assume that I is an invertible ideal with $1 \in I$, hence $R \subset I$, and that 1 is a non-zero element of minimal absolute value in I.

Let us find out whether I can contain more elements except the ones of R. To do this, it suffices to consider points in the rectangle with corners 0, 1, $\sqrt{5}i$, and $1+\sqrt{5}i$ shown in the pictures below: I is an additive subgroup of $\mathbb{C}$ containing R, and so the complete set of points in I will just be an R-periodic repeated copy of the points in this rectangle.

To figure out if I contains more points in this rectangle, we proceed in three steps illustrated below.

(a) By construction, I contains no points of absolute value less than 1 except the origin, i. e. no points in the open disc $U_1(0)$ with radius 1 and center point 0. Likewise, because of the R-periodicity of I, the ideal also does not contain any points in the open unit discs around the other corners of the rectangle, except these corner points themselves. In other words, the shaded area in picture (a) below cannot contain any points of I except the corner points.



(b) Now consider the open disc $U_{\frac{1}{2}}(\frac{1}{2}\sqrt{5}i)$, whose intersection with our rectangle is the left dark half-circle in picture (b) above. Again, it cannot contain any points of I except its center: as I is an R-module, any non-center point $a \in I$ in this disc would lead to a non-center point $2a \in I$ in the disc $U_1(\sqrt{5}i)$, which we excluded already in (a). But in fact the center point $\frac{1}{2}\sqrt{5}i$ cannot lie in I either, since then we would have

$$\sqrt{5}i \cdot \frac{1}{2}\sqrt{5}i + 3 \cdot 1 = \frac{1}{2} \in I$$

as well, in contradiction to (a). In the same way we see that the open disc $U_{\frac{1}{2}}(1+\frac{1}{2}\sqrt{5}i)$ does not contain any points of I either. Hence the complete shaded area in picture (b) is excluded now for points of I.

(c) Finally, consider the open disc $U_{\frac{1}{2}}(\frac{1}{2}+\frac{1}{2}\sqrt{5}i)$, shown in picture (c) above in dark color. For the same reason as in (b), no point in this disc except the center can lie in I. As our discs now cover the complete rectangle, this means that the only point in our rectangle except the corners that can be in I is $\frac{1}{2}+\frac{1}{2}\sqrt{5}i$. This leads to exactly two possibilities for I:

$$\text{either} \quad I = R = I_0 \qquad \text{or} \quad I = \left(1, \frac{1}{2}+\frac{1}{2}\sqrt{5}i\right) = \frac{1}{2}I_1.$$

In fact, we know already that this last case $\frac{1}{2}I_1$ is an invertible ideal of R, so that this time (in contrast to (b) above) it does not lead to a contradiction if the center point of the disc lies in I.

Altogether, we thus conclude that $|\operatorname{Pic} R| = 2$, i. e. $\operatorname{Pic} R \cong \mathbb{Z}/2\mathbb{Z}$, with the class of I_0 being neutral and I_1 being the unique other element. We have indeed also checked already that I_1 is its own inverse in $\operatorname{Pic} R$, since by Example 13.8

$$I_1 \cdot I_1 = (2)$$

is principal, and hence the neutral element in $\operatorname{Pic} R$.

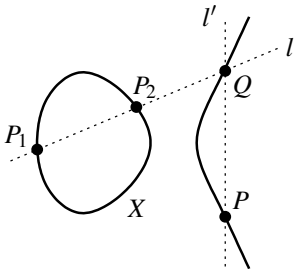
Example 13.29 (A non-trivial Picard group in geometry)**.** Consider again the complex plane cubic curve $X = V(y^2 - x(x-1)(x-\lambda)) \subset \mathbb{A}^2_{\mathbb{C}}$ for some $\lambda \in \mathbb{C}\backslash\{0,1\}$ as in Example 13.12. We have already seen that its coordinate ring $R = A(X)$ is a Dedekind domain. Let us now study its ideal class group $\operatorname{Pic} R$.

Note that there is an obvious map

$$\varphi : X \to \operatorname{Pic} R, \;\; a \mapsto \overline{I(a)}$$

that assigns to each point of X the class of its maximal ideal in $\operatorname{Pic} R = \operatorname{Div} R / \operatorname{Prin} R$. The surprising fact is that this map φ is injective with its image equal to $\operatorname{Pic} R \backslash \{\overline{(1)}\}$, i. e. to $\operatorname{Pic} R$ without its neutral element. We can therefore make it bijective by adding a "point at infinity" to X that is mapped to the missing point $\overline{(1)}$ of $\operatorname{Pic} R$. This is particularly interesting as we then have a bijection between two completely different algebraic structures: X is a variety (but a priori not a group) and $\operatorname{Pic} R$ is a group (but a priori not a variety). So we can use the bijection φ to make the cubic curve $X \cup \{\infty\}$ into a group, and the group $\operatorname{Pic} R$ into a variety.

We cannot prove this statement or study its consequences here since this would require methods that we have not covered in this course — this is usually done in the "Algebraic Geometry" class. However, we can use our definition of the Picard group and the map φ above to describe the group structure on the curve $X \cup \{\infty\}$ explicitly. To do this, consider two points on the curve with maximal ideals P_1 and P_2, as shown in the picture on the right. Draw the line through these two points; it will intersect X in one more point Q since the cubic equation $y^2 = x(x-1)(x-\lambda)$ together with a linear equation in x and y will have three solutions. By Example 13.12, this means algebraically that there is a linear polynomial $l \in R$ (whose zero locus is this line) such that $(l) = P_1 \cdot P_2 \cdot Q$.



Next, draw the vertical line through Q. By the symmetry of X, it will intersect X in one more point P. Similarly to the above, it follows that there is a linear polynomial l' such that $(l') = Q \cdot P$. But in the Picard group $\operatorname{Pic} R$ this means that

$$\overline{P_1} \cdot \overline{P_2} \cdot \overline{Q} = \overline{(l)} = \overline{(l')} = \overline{Q} \cdot \overline{P}$$

(note that (l) and (l') both define the neutral element in $\operatorname{Pic} R$), and thus that $\overline{P_1} \cdot \overline{P_2} = \overline{P}$. Hence the above geometric construction of P from P_1 and P_2 describes the group structure on $X \cup \{\infty\}$ mentioned above.

Note that it is quite obvious without much theory behind it that this geometric two-line construction can be used to associate to any two points on X (corresponding to P_1 and P_2 above) a third point on X (corresponding to P). The surprising statement here (which is very hard to prove without using Picard groups) is that this gives rise to a group structure on $X \cup \{\infty\}$; in particular that this operation is associative. The neutral element is the additional point ∞ (corresponding to the class of principal ideals in $\operatorname{Pic} R$), and the inverse of a point in X is the other intersection point of the vertical line through this point with X (so that e. g. in the above picture we have $\overline{P}^{-1} = \overline{Q}$).

References

[AM] M. Atiyah, I. MacDonald, *Introduction to Commutative Algebra*, Addison Wesley (2004)

[E] D. Eisenbud, *Commutative Algebra with a View towards Algebraic Geometry*, Springer (2004)

[G1] A. Gathmann, *Algebraische Strukturen*, class notes TU Kaiserslautern (2019/20), https://www.mathematik.uni-kl.de/~gathmann/ags

[G2] A. Gathmann, *Grundlagen der Mathematik*, class notes TU Kaiserslautern (2018/19), https://www.mathematik.uni-kl.de/~gathmann/gdm

[G3] A. Gathmann, *Einführung in die Algebra*, class notes TU Kaiserslautern (2010), https://www.mathematik.uni-kl.de/~gathmann/algebra

[HM] S. Hampe, T. Markwig, *Commutative Algebra*, class notes TU Kaiserslautern (2012), www.math.uni-tuebingen.de/~keilen/download/Lehre/MGSS09/CommutativeAlg.pdf

[K] T. Keilen, *Algebraic Structures*, class notes TU Kaiserslautern (2009), www.math.uni-tuebingen.de/~keilen/download/LectureNotes/algebraicstructures.pdf

[M] T. Markwig, *Elementare Zahlentheorie*, class notes TU Kaiserslautern (2010), www.math.uni-tuebingen.de/user/keilen/download/LectureNotes/zahlentheorie.pdf

[S] W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, SINGULAR — a computer algebra system for polynomial computations, https://www.singular.uni-kl.de

Index