14 lectures on

COMMUTATIVE ALGEBRA

an introduction

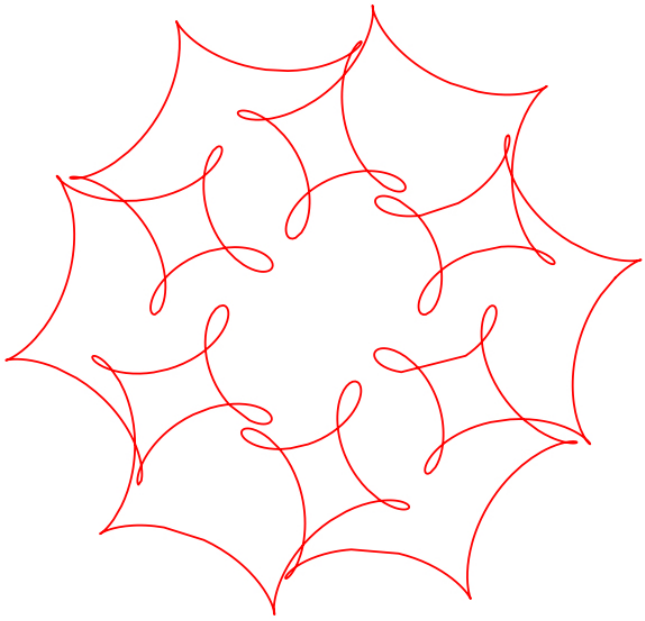
GEIR ELLINGSRUD

Geir Ellingsrud

16 lectures on

COMMUTATIVE ALGEBRA

An introduction



Fell Forlag

Contents

Lecture 0

Intro

Warning: This is a preliminary version. I am working on them and new (and hopefully) better versions will surface from time to time. They still suffer from several shortcomings and are prone to errors (not so many and not so serious, I hope), but they will (hopefully) improve! Some sections are thoroughly checked while others are raw and under construction and they are marked with a warning sign: "careful—construction!".
Version 4.1 (run 193)—14th June 2021 at 10:26am

These notes grew out my giving the introductory course in commutative algebra at UiO at several occasions during the last ten years. This is course where the students meet serious commutative algebra for the first time. Their backgrounds are diverse. They know some linear algebra, but mostly not from a theoretical standpoint, and very few have come as far as the Cayley–Hamilton Theorem. They have had a rudimentary experience in commutative algebra, and have heard about rings and ideals and have seen some examples, but to indicate their level, most do not know Gauss' lemma. Most have followed a course in group and Galois theory, which occasionally goes as far as the Sylow's theorems, and which include basic Galois theory. Given these conditions, the notes starts at the very beginning with the very basic properties of rings and ideals

With that starting point the theory is developed introducing the fundamental concepts and techniques; in short, a guide to a beginners tools necessary to start off practising commutative algebra; And of course, subsequently this leads to the usual collection of the "great theorems" of David Hilbert, Emmy Noether and Wolfgang Krull; the corner stones of the whole theory. A primary function of the course is to prepare the ground for studies in algebraic geometry and number theory,

Being a preparatory course, there is a risk it leaves you with a lurking melancholy as expresses in the lyrics from Leonard Bernstein's song *Some Other Time*: "just when the fun is starting comes the time for parting"(so beautifully performed by Monica Zetterlund and Bill Evans). But there is a cure: Do more mathematics!

The notes are written for the students. The style is rather ample with detailed

explanations, which makes the text rather long. But redundancy of the language is an important factor in making a text accessible and easy to read, however redundancy without variation is futile if not contra productive. Remember the french saying: when you complain you don't understand what the British say, they just repeat the phrase but louder. There is also a gradient in the redundancy—as the course evolves more details are left to the students.

Categories and functors entered mathematics in the 1940's work of Eilenberg and Maclane, and as Peter Freud states "in a fairly explosively manner functors and natural transformations permeated a wide variety of subjects". To give a master course in algebra today—about eighty years later—in an attitude that the word *category* is a slip of the tongue, would be close to a heresy. But categories and functors do not enter the presentation in a substantial way, they only appear as notational devices, except at a few placed a mild use of easy categorical techniques will be convenient and clarifying. And for the benefit of the students a very short appendix is included with the rudimental definitions.

Mathematical theories are not linearly ordered, but rather constitute some kind of intricate graph with nodes being statements and edges being implications. But time is linear, and a challenge for a lecturer is to find a path through this mathematical skein valid both scientifically and pedagogically. And speaking about time, to reach through the curriculum before the term ends, can be sever. The notes suffer from the common syndrome of a never ending expansion, and threaten to end up obese, so any lecturer that might chose to them is obliged to make a reasonable choice among the chapters.

Giving examples is an important part of the teaching, establishing a broad background for the students intuition. So examples abound, some are mainstream situations, but others function as eyeopeners: they are meant to illustrate what delicate situation one risks finding oneself in and what denizens one risks meeting when venturing the stormy waters where the standard hypotheses of the theory no more comply. It is also important to understand why the specific hypotheses of a result are required, and often this is best illustrated through examples.

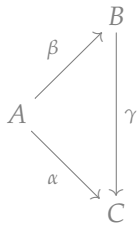
Doing exercise are as well a fundamental when learning mathematics; so we include almost four hundred (397 to be exact)—some are easy and some more demanding. Solutions are provided for many (for the moment only 101 exercises from the first chapters are solved; the solution part is still a construction cite), and they are indicated ✲ by a golden star in the margin. To jump forth and back between an exercise and its solution just click on the numbers. A habit of many authors is to bundle up parts of the theory with the exercises, not (always) out of laziness, but most often as an attempt to limit the number of pages. Anyhow, from the pedagogical angle it is sound practice to force students to participate actively in developing the matter. So, also in these notes some exercises are part of the theory. Ideally, solutions should always be provided for

these exercises (and eventually will).

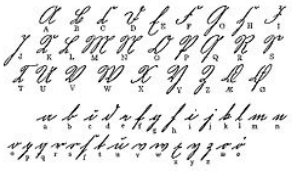
Definitions are not, as in many texts, formatted in special typography to stand out, but are indicated in the margin in blue with the Norwegian version of the name in red.

Definition (definisjon)

An insurmountable means of keeping track of maps and equalities between compositions, is to draw diagrams; a simple one is shown in the margin . Diagrams come all kinds of shapes and can be utterly intricate, but for us they will always be simple and mostly triangular or square. One says that a diagram, or a part of a diagram, is *commutative* if possible equalities hidden in the geometry of the diagram, in fact are equalities; so for example $\alpha = \gamma \circ \beta$ in the marginal example. Almost every diagram we shall draw will be commutative, and we shall tacitly assumed they are.



So a few words about the gothic alphabet, which poperly should be called the *blackletter* typefaces. Mathematicians are always in shortage of alphabets and letters, and tend to use all kinds of creative decorations to have enough glyphs. To overcome this typographical shortfall we have decided to designate ideals by a blackletter typeface. There is certainly a very strong historical evidence—a reminder of the deep German roots of algebra—and it is also convenient in many ways: like a team in uniforms you recognize ideals immediately. However, the inconvenience is that blackletter letters can be difficult to write by hand, and handwritten ones difficult to read, especially the $\mathfrak{p}$ and the $\mathfrak{q}$ stand out in this respect. We confined the use to the lower case letters $\mathfrak{a}$, $\mathfrak{b}$ $\mathfrak{c}$, $\mathfrak{m}$, $\mathfrak{n}$, $\mathfrak{p}$, $\mathfrak{q}$, and may be occasionally an $\mathfrak{r}$. Letters like I and J mostly denote sets of indices, A, B and C and from time to time R and S, are rings and while M and N will be modules.



Lecture 1

Rings

The starring role in commutative algebra is played by the commutative rings and their ideals—they are even the main targets of the investigations. In this chapter we become acquainted with rings, and ideals will be introduced in the next chapter.

Commutative rings come in a great variety of flavours, and the sources where they arise are as diverse. Some rings are best though of as "number systems" as the ring $\mathbb{Z}$ of integers and its well-known larger siblings the field of rationals $\mathbb{Q}$, the field of real numbers $\mathbb{R}$ and the field of complex numbers $\mathbb{C}$ (this suite may be brought at least two steps further, but in a non-commutative way; the next two members are called the quaternions and the octonions). There are also some ubiquitous "little brothers"; the rings $\mathbb{Z}/n\mathbb{Z}$ of integral residue classes modulo a natural number n. Among them we find the finite fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ with p elements, p being a prime, and there are also the other finite fields $\mathbb{F}_q$ with q elements, q being a prime power $q = p^r$. And there are naturally also some "big brothers"; for instance, the field $\bar{\mathbb{Q}}$ consisting of the complex numbers that are roots of polynomials with rational coefficients; and inside $\bar{\mathbb{Q}}$ we find the the ring $\bar{\mathbb{Z}}$ of algebraic integers; the complex numbers being roots of *monic* polynomials with integral coefficients.

The earliest systematic study of commutative rings was of various "generalized number systems"; certain subrings of the ring of algebraic numbers. Already Gauss undertook such studies, but they really sparked off in the nineteenth century with the work of Kronecker and Dedekind.

Other commutative rings resemble rings of functions on different kinds of spaces, like continuous functions on topological spaces (with real or complex values) or holomorphic functions in open domains in the complex plain, but the rings most relevant in our context arise in algebraic geometry. These are rings of polynomial functions with values in a field k defined on so-called varieties, which are vanishing loci in k^n for sets of polynomials.

The development took a new direction around the middle of the twentieth century,

when mathematicians like Zariski and Weil strived for establishing a sound foundation of algebraic geometry, and the recognition of the power algebraic geometric methods have in number theory, eventually lead to the happy marriage of algebraic geometry and number theory — consummated by Grothendieck and his invention of schemes.

1.1 Rings

(1.1) Recall that a ring A is an algebraic structure consisting of a set endowed with two binary operations; an *addition* which makes A an abelian group, and a *multiplication*. The multiplication is assumed to be distributive over the addition, and in this course it will always be associative and commutative (or at least almost always). There are of course both many non-commutative rings and non-associative rings that are extremely interesting, but this course is dedicated to rings that are associative and commutative.

The sum of two elements will naturally be denoted as $a + b$, and the product will we indicated in the traditional way by a dot or simply by juxtaposition; that is, as $a \cdot b$ or just as ab. The *left distributive law* asserts that $a(b + c) = ab + bc$, and since rings for us are commutative, it follows that the *right distributive law* $(b + c)a = ba + ca$ holds as well.

We shall also assume that all rings have a *unit element*; that is, an element 1_A such that $1_A \cdot a = a \cdot 1_A = a$ for all members a of the ring. At most occasions the reference to A will be dropped and the unit element written as 1 whatever the ring is.

EXAMPLE 1.1 The simplest of all rings are the ring $\mathbb{Z}$ of integers and the rings $\mathbb{Z}/n\mathbb{Z}$ of residue classes of integers modulo n. The traditional numbers systems of rational numbers $\mathbb{Q}$, of real numbers $\mathbb{R}$ and of complex numbers $\mathbb{C}$ are well-known rings. ☆

Zero divisors and nilpotents

Zero divisors (nulldivisorer)

Integral domains (integritetsområder)

Non-zero divisors (ikke-nulldivisorer)

Regular elements (regulære elementer)

(1.2) Elements in general rings can behave quite differently from what we are used to in a classical setting of real and complex numbers. It might very well happen that $ab = 0$ without neither a nor b being zero. Such elements are called *zero divisors*. Be aware that the familiar *cancellation law* does not hold in a ring with zero divisors in that $ab = ac$ not necessarily implies that $b = c$. Rings without zero divisors are called *integral domains* or, for short, *domains*. Obviously, elements that are not zero divisors are called *non-zero divisors*, another name being *regular elements*. A regular element a has the virtue that $xa = 0$ implies that $x = 0$ and can therefore be cancelled from equalities like $ab = ac$ (the difference $b - c$ is killed by a and hence vanishes). So in an integral domain, the cancellation law is in force.

For instance, the rings $\mathbb{Z}/n\mathbb{Z}$ have zero divisors whenever n is a composite number: if n factors as $n = pq$ with p and q natural numbers both different from n, it holds true

that $pq = 0$ in $\mathbb{Z}/n\mathbb{Z}$, and p and q are both non-zero in $\mathbb{Z}/n\mathbb{Z}$ neither having n as a factor.

A more geometric example could be the ring of continuous functions on the space X which is the union of the x-axis and the y-axis in the plane. On X the function xy vanishes identically, but neither x nor y does; x does not vanish on the y-axis and y not on the x-axis.

(1.3) It might also happen that powers of non-zero elements vanish, *i. e.* one has $a^n = 0$ for some natural number n but with $a \neq 0$. For instance, in the ring $\mathbb{Z}/p^2\mathbb{Z}$ it holds true that $p^2 = 0$, but $p \neq 0$. Such elements are called *nilpotent*. Rings deprived of nilpotent elements are said to be *reduced*.

Nilpotent elements (nilpotente elementer)

Reduced rings (reduserte ringer)

Units and fields

(1.4) Division by non-zero elements is generally not possible in rings and non-zero elements are not in general invertible. For instance, if p and q are two different primes in $\mathbb{Z}$, the fraction p/q is not an integer and does not lie in $\mathbb{Z}$. Elements in a ring A that are invertible, *i. e.* ring-elements a for which there is an element a^{-1} in A with $aa^{-1} = 1$, are called *units*. They form an abelian group under multiplication, which we shall denote by A^*.

Units (enheter)

Rings A all whose non-zero members are invertible; that is, which satisfy $A^* = A\backslash\{0\}$, are called *fields*. In fields division by non-zero elements can be performed unconditionally.

Fields (kropper)

EXAMPLE 1.2 Well-known fields are the fields of rational numbers $\mathbb{Q}$, of real numbers $\mathbb{R}$ and of complex numbers $\mathbb{C}$. If p is a prime number, the ring $\mathbb{Z}/p\mathbb{Z}$ of integers modulo p is a field, usually denoted by $\mathbb{F}_p$. It is a finite field having p elements. ☆

Examples

(1.3) We do not assume that $1 \neq 0$ although it holds in all but one ring. The exceptional ring is the so-called *null-ring*. When $0 = 1$, it follows that $a = a \cdot 1 = a \cdot 0 = 0$, so zero will be the sole element. The only role the null-ring plays, and the only reason not to throw it over board, is that it allows significantly simpler formulations of a many results, and it does not merit a proper notation (well, one always has the alternative 0).

The null-ring (nullringen)

(1.4) The set of polynomials $\mathbb{Q}[x_1, \ldots, x_r]$ in r variables $x_1, \ldots, x_r$ with rational coefficients is a ring when equipped with the usual sum and product, as are the set of real polynomials $\mathbb{R}[x_1, \ldots, x_r]$ and the set of complex polynomials $\mathbb{C}[x_1, \ldots, x_r]$.

(1.5) The complex rational functions in a variable x form a field $\mathbb{C}(x)$. The elements are meromorphic functions in $\mathbb{C}$ expressible as the quotient $p(x)/q(x)$ of two polynomials p and q with q not being identically zero. One is not confined to just one variable; the field $\mathbb{C}(x_1, \ldots, x_r)$ of rational functions in the variables $x_1, \ldots, x_r$ consists of fractions $p(x_1, \ldots, x_r)/q(x_1, \ldots, x_r)$ where p and q are polynomials and where q is not the zero polynomial.

(1.6) For any set $X \subseteq \mathbb{C}^r$ one may consider the set of *polynomial functions* on X; that is, the functions on X that are restrictions of polynomials in r variables. They form a ring $A_{\mathbb{C}}(X)$ under point-wise addition and multiplication.

Polynomial functions (polynomiale funskjoner)

(1.7) Associated with any topological space X are the sets $C_{\mathbb{R}}(X)$ and $C_{\mathbb{C}}(X)$ of continuous functions on X assuming respectively real or complex values. Point-wise addition and multiplication make them (commutative) rings. When X has more structure than just a topology, there are further possibilities. Two instances are the ring of smooth functions on a smooth manifold, and the ring $\mathcal{O}(\Omega)$ of holomorphic functions in an open domain Ω of the complex plane.

Quadratic extensions (kvadratiske utvidelser)

(1.8) *Quadratic extensions:* An example of a class of rings, important in algebraic number theory, is the class of the *quadratic extensions* $\mathbb{Z}[\sqrt{n}]$ obtained by adjoining a square root to $\mathbb{Z}$; that is, $\mathbb{Z}[\sqrt{n}] = \{\, a + b\sqrt{n} \mid a, b \in \mathbb{Z} \,\}$, where n is any integer (positive or negative). These rings are contained in the field of complex numbers $\mathbb{C}$ and inherit their ring structure from $\mathbb{C}$; to verify that they are rings is suffices to see they are closed under addition—which is obvious—and multiplication, which ensues from the little calculation

$$(a + b\sqrt{n})(a' + b'\sqrt{n}) = (aa' + nbb') + (ab' + a'b)\sqrt{n},$$

the point being that $(aa' + nbb')$ and $(ab' + a'b)$ are integers when a, a', b, b' and n are. A few special case have their proper name; for instance, elements of $\mathbb{Z}[i]$ are called Gaussian integers.

☆

Homomorphisms

When studying mathematical objects endowed with certains structures—like rings for instance, which have an additive and a multiplicative structure—maps preserving the structures are fundamental tools. Working with topological spaces one uses continuous maps all the time, and linear algebra is really about linear maps between vector spaces. And of course, the theory of groups is inconceivable without group homomorphisms; that is, maps respecting the group laws. A new class of objects in mathematics is always accompanied by a new class of maps. This observation can be formalized and leads to the definition of *categories*.

Categories (kategorier)

Ring homomorphisms (ringhomomorfier)

(1.5) In our present context the relevant maps are the so-called *ring homomorphism*, which also will be referred to as *maps of rings* or *ring-maps*. These are maps $\phi\colon A \to B$ between two rings A and B preserving all the structures around; that is, the additive group structure, the multiplication and the unit element 1. In other words, they comply with the two rules

- $\phi(a + b) = \phi(a) + \phi(b)$;
- $\phi(ab) = \phi(a)\phi(b)$ and $\phi(1) = 1$.

The sum of two maps of rings is in general *not* a map of rings (it is additive, but does not respect the multiplication) neither is their product (it respects multiplication, but not

addition), but of course, the composition of two composable ring-maps is a ring-map. The rings (commutative with unit) together with their homomorphisms form a category denoted Rings.

Isomorphisms of rings (isomorfier av ringer)

(1.6) A homomorphism $\phi\colon A \to B$ is an *isomorphism* if there is a ring homomorphism $\psi\colon B \to A$ such that the two relations $\psi \circ \phi = \mathrm{id}_A$ and $\phi \circ \psi = \mathrm{id}_B$ hold true. One most often writes ϕ^{-1} for the inverse map, and it is common usage to call isomorphisms *invertible* maps. For ϕ to be invertible it suffices it be *bijective*. Multiplication will then automatically be respected since $\phi^{-1}(ab) = \phi^{-1}(a)\phi^{-1}(b)$ is equivalent to $ab = \phi(\phi^{-1}(a)\phi^{-1}(b))$, and the latter equality is a consequence of ϕ respecting multiplication. Applying ϕ^{-1} to $\phi(1_A) = 1_B$ one sees that $\phi^{-1}(1_B) = 1_A$, so the inverse map sends the unit element to the unit element as well. An analogous argument shows that ϕ^{-1} also is additive.

Examples

(1.9) So-called *evaluation maps* are omnipresent examples of ring homomorphisms. To illustrate this concept, we pick a point $a \in \mathbb{C}^n$. Sending a polynomial f to the value it assumes at a, gives a map $\mathbb{C}[x_1, \ldots, x_n] \to \mathbb{C}$, and by the very definition of the ring structure of the polynomial ring (addition and multiplication are point-wise operations) this is a map of rings.

Any ring A of functions—say with complex values—on any space X possesses analogue evaluation maps. The operations in A being defined point-wise the map $f \mapsto f(x)$ is a ring-map from A to $\mathbb{C}$ for any point $x \in X$.

(1.10) Another series of well-known examples of ring-maps are the maps $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ that send an integer a to its residue class $[a]$ modulo the integer n.

☆

Subrings and polynomial expressions

(1.7) We begin by recalling the notion of a polynomial expression. Assume given a ring A and sequence $a = (a_1, \ldots, a_r)$ of elements from A. For any multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$; that is, a sequence on non-negative integers, one has the *monomial expression*

Monomial expression (monomiale uttrykk)

$$a^\alpha = a_1^{\alpha_1} \cdot \ldots \cdot a_n^{\alpha_n}.$$

Polynomial expressions (polynomiale uttrykk)

These expressions show an exponential behavior in that $a^\alpha \cdot a^\beta = a^{\alpha+\beta}$. A *polynomial expression* in the a_i's is just a finite linear combination of such monomials. Frequently one wants to confine the coefficients to a specific subset S of A, and then one speaks about *polynomial expressions with coefficients in* S. They are thus elements of A shaped like

$$\sum_\alpha s_\alpha \cdot a^\alpha = \sum_\alpha s_\alpha \cdot a_1^{\alpha_1} \cdot \ldots \cdot a_n^{\alpha_n},$$

where the summation extends over all multi-indices, and where the non-zero coefficients are finite in number and confined to S.

A successive application of the distributive law and the exponential behaviour of monomials gives the classical formula for the product of two polynomial expressions:

$$(\sum_{\alpha} s_\alpha \cdot a^\alpha) \cdot (\sum_{\beta} t_\beta \cdot a^\beta) = \sum_{\gamma} (\sum_{\alpha+\beta=\gamma} s_\alpha t_\beta) \cdot a^\gamma. \tag{1.1}$$

Subrings (underringer)

(1.8) A *subring* B of A is a ring contained in A whose ring operations are induced from those of A. Phrased differently, it is an additive subgroup containing the unit element which is closed under multiplication; to be specific, it holds that $0 \in B$ and $1 \in B$, and for any two elements a and b belonging to B, both the sum $a + b$ and the product ab belong to B. The intersection of any family of subrings of A clearly is a subring.

EXAMPLE 1.11 The integers $\mathbb{Z}$ is a subring of the rationals $\mathbb{Q}$. ☆

(1.9) Given a ring A and a subring B and a set of elements $a_1, \ldots, a_r$ from A, one constructs a subring $B[a_1, \ldots, a_r]$ of A as the set of all polynomial expressions

$$\sum b_\alpha \cdot a_1^{\alpha_1} \cdot \ldots \cdot a_r^{\alpha_r}$$

where $\alpha = (\alpha_1, \ldots, \alpha_r)$ runs through the multi-indices and the b_α's are elements from B, only finitely many of which are different from zero. It is straightforward to check, using the classical formula (1.1) above, that this subset is closed under multiplication and hence is a subring of A (it is obviously closed under addition). It is called the *subring generated* by the a_i's over B, and is the smallest subring of A containing the ring B and all the elements a_i. Common usage is also to say that $B[a_1, \ldots, a_r]$ is obtained by adjoining the a_i's to B.

Subrings generated by elements (underringer generert av elementer)

This construction works fine even for infinitely many a_i's since each polynomial expression merely involves finitely many of them. Thus there is a subring $B[a_i | i \in I]$ for any subset $\{a_i\}_{i \in I}$ of A. It equals the intersection of all sub rings of A containing B all the a_i's.

Examples

(1.12) Let n be an integer. The ring $\mathbb{Z}[1/n] = \{ m/n^i \mid i \in \mathbb{N}_0, m \in \mathbb{Z} \}$ is a subring of $\mathbb{Q}$. The elements are the rational numbers whose denominator is a power of n. More generally, if S is any set of integers, one may form $\mathbb{Z}[n^{-1} | \, n \in S]$, which is the subring of $\mathbb{Q}$ consisting of the rational numbers whose denominator is a product of numbers from S.

Be aware that quite different sets S can give rise to the same subring. For instance, when $p_1, \ldots, p_r$ are the primes occurring in the prime factorization of the integer n, it holds true that $\mathbb{Z}[1/n] = \mathbb{Z}[p_1^{-1}, \ldots, p_r^{-1}]$.

(1.13) The subring $\mathbb{C}[t^2, t^3]$ of $\mathbb{C}[t]$ is a ubiquitous example in algebraic geometry; it is the coordinate ring of a so-called *cusp* and consists of all polynomials whose first derivative vanishes at the origin; or phrased differently, the polynomials without a linear term.

(1.14) The subring $\mathbb{C}[x, 1/x]$ of the rational function field $\mathbb{C}(x)$ consists of elements of the form $p(x^{-1}) + c + q(x)$ where p and q are polynomials vanishing at the origin and c a complex constant.

☆

The prime ring and the characteristic

(1.10) Every ring has a canonical subring called the *prime ring*. The unit element 1 in A generates an additive cyclic subgroup of A whose elements are just sums of 1 or -1 a certain number of times; that is, they are shaped like $n = 1 + \ldots + 1$ or $n = -1 - \ldots - 1$. This subgroup is obviously closed under multiplication and is hence a subring. It is called *the prime ring* of A.

The prime ring (primringen)

As is well known, a cyclic group is either finite and isomorphic to $\mathbb{Z}/n\mathbb{Z}$ for some positive integer n, or it is infinite and isomorphic to $\mathbb{Z}$. The prime ring is therefore either isomorphic to one of the rings $\mathbb{Z}/n\mathbb{Z}$ or to $\mathbb{Z}$. In the former case the integer n is called the *characteristic* of A, in the latter case one says that A is of *characteristic zero*. So, in any case, the characteristic of a ring A is a non-negative integer attached to A.

The characteristic of a ring (karakteristikken til en ring)

(1.11) Any field contained in A contains the prime ring. Hence, if A contains a field, the characteristic is either prime or zero. In case it is prime, the prime ring equals the field $\mathbb{F}_p$, and in case the characteristic is zero, the ring A contains $\mathbb{Q}$ as well. We say that $\mathbb{F}_p$ respectively $\mathbb{Q}$ is the *prime field* of A.

The prime field (primkroppen)

Algebras

Frequently when working in commutative algebra there are "coefficients" around; that is, one is working over a "ground ring". So the most natural objects to work with are perhaps not rings, but the so-called *algebras*.

(1.12) The notion of an algebra is a relative notion involving two rings A and B. To give a *B-algebra structure* on A is just to give a map of rings $\phi\colon B \to A$. One may then form products $\phi(b) \cdot a$ of elements a from A with elements of the form $\phi(b)$. The map ϕ, even though it is an essential part of the B-algebra structure of A, is often tacitly understood and suppressed from the notation; one simply writes $b \cdot a$ for $\phi(b) \cdot a$. Later on, when we have introduced modules, a B-algebra structure on a ring A will be the same as a B-module structure on A with the extra condition that multiplcation in A is B-linear.

Algebras (algebraer)

Example 1.15 Every ring has a canonical structure as a $\mathbb{Z}$-algebra (defined as in Paragraph 1.10 above). The class of algebras is therefore a strict extension of the class of rings. Since a ring is an algebra over any subring, over-rings give a large number of examples of algebras. ☆

*Or any morphological derivative thereof, like map of B-algebras *or* B-algebra-map *etc.*

Algebra homomorphisms (algebrahomomorfismer)

(1.13) Faithful to the principle that any new type of objects is accompanied by a corresponding new type of maps; one says that a map of rings $\phi\colon A \to A'$ between two B-algebras is an *B-algebra homomorphism** if it respects the action of B; in other words, it holds true that $\phi(b \cdot a) = b \cdot \phi(a)$ for all elements $a \in A$ and $b \in B$. Composition of two composable B-algebra homomorphisms is a B-algebra homomorphism so that the B-algebras form a category denoted Alg_B.

Finitely generated algebras (endeliggenererte algebraer)

Finite type algebras (algebraer av endelig type)

(1.14) One says that A is *finitely generated over B*, or is of *finite type* over B, if $A = B[a_1, \ldots, a_r]$ for elements $a_1, \ldots, a_r$ from A.

EXAMPLE 1.16 A note of warning might be appropriate, algebra structures can be deceptive. Every ring is of course an algebra over itself in a canonical way (the algebra structure is given by the identity map), but there can be other unorthodox ways A can be an A-algebra. A simple example to have in mind is the field $\mathbb{C}$ of complex numbers, which has an alternative algebra structure induced by complex conjugation. In this structure a complex number z acts on another complex number w as $\bar{z} \cdot w$.

The two structures are *not* isomorphic as $\mathbb{C}$-algebras although the underlying rings are the same. A good try for an isomorphism would be the identity map, but it does not respect the two algebra-structures*. Similar unorthodoxy will arise from any ring endomorphism $A \to A$. Examples of such are furnished by the *Frobenius homomorphisms* of rings of positive characteristic (see Exercsise 1.7 below). ☆

*Of course, it holds true that $\mathrm{id}_{\mathbb{C}}(zw)$ equals $z\,\mathrm{id}_{\mathbb{C}}\,w$ and not $\bar{z}\,\mathrm{id}_{\mathbb{C}}\,w$

Exercises

✲ (1.1) Assume that A is a finite ring. Show that the units are precisely the elements that are not zero divisors. Conclude that if A is an integral domain, it is a field.

✲ (1.2) Find all nilpotents and all zero divisors in $\mathbb{Z}/72\mathbb{Z}$. What are the units?

✲ (1.3) Generalize the previous exercise: Let n be a natural number. Determine nilpotents, zero divisors and units in $\mathbb{Z}/n\mathbb{Z}$.

(1.4) Show that the prime ring is the smallest subring of a ring; *i. e.* it is contained in all other subrings of the given ring.

(1.5) *The Binomial Theorem.* Convince yourself that the binomial theorem persists being true in any commutative ring; that is, check that your favourite proof still holds water.

(1.6) Show that the sum of two nilpotent elements is nilpotent. HINT: You can rely on the binomial theorem.

(1.7) *The Frobenius homomorphism.* Let A be a ring of positive prime characteristic p. Show that the relation

$$(a+b)^p = a^p + b^p$$

holds true for all $a, b \in A$. Hence the map $A \to A$ that sends a to the p^{th} power a^p is a ring homomorphism. It is called the *Frobenius homomorphism*. HINT: The binomial coefficients $\binom{p}{r}$ have p as factor when $1 < r < p$.

⁂ (1.8) Show that any intermediate ring $\mathbb{Z} \subseteq A \subseteq \mathbb{Q}$ is of the form $A = \mathbb{Z}[p^{-1} | p \in S]$ for some set S of primes.

(1.9) Let $\phi\colon A \to B$ be a map of rings. Show that ϕ induces a group homomorphism mapping A^* into B^*.

⁂ (1.10) *Units in imaginary quadratic extensions.* Let n be a natural number. Show that an element $x \in \mathbb{Z}[\sqrt{-n}]$ is a unit if and only if $|x| = 1$ (where $|x|$ denotes the ordinary absolute value of the complex number x), and use this to determine the units in $\mathbb{Z}[\sqrt{-n}]$.

⁂ (1.11) Assume that a is a nilpotent element of the ring A. Show that $1 + a$ is invertible. More precisely: If $a^n = 0$, the inverse is given as $(1+a)^{-1} = 1 - a + a^2 - \ldots + (-1)^{n-1}a^{n-1}$. Conclude that if u is a unit and a nilpotent, then $u + a$ is invertible.
HINT: Use the good old formula for the sum of a geometric series.

★

1.2 *Polynomials*

We are well acquainted with polynomials with real or complex coefficients; we met them already during the happy days at school. They were then introduced as functions depending on a real (or complex) variable whose values were given by a polynomial expressions. In this section we shall introduce polynomials with coefficients in any (commutative) ring A. The point of view will necessarily be formal and without reference to functions, and there is no reason to confine oneself to just one variable.

(1.15) In an earlier paragraph we met polynomial expressions in a set of ring elements. In the present situation where there is no surrounding ring, we must, as signalled above, proceed in a formal way. A *polynomial* in the variables $x_1, \ldots, x_r$ is defined as a formal sum

Polynomials (Polynomer)

$$f(x_1, \ldots, x_n) = \sum_{\alpha} a_\alpha x_1^{\alpha_1} \cdot \ldots x_n^{\alpha_n}, \tag{1.2}$$

where the summation extends over all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n)$ with the α_i's being non-negative integers, and where the coefficients a_α are elements from the ground ring A, only finitely many of which are non-zero. Do not speculate much* about what the term "formal sum" means, the essential point is that two such "formal sums" are equal exactly when corresponding coefficients agree.

*Or do Exercise 1.19 below where a more general construction of so-called monoidal algebras is described in a precise manner.

(1.16) The "pure" terms $a_\alpha x_1^{\alpha_1} \cdot \ldots x_n^{\alpha_n}$ occurring in (1.2) are called *monomials*, and the abbreviated notation $x^\alpha = x_1^{\alpha_1} \cdot \ldots x_n^{\alpha_n}$ is convenient and practical. The *degree* of a non-zero monomial $a_\alpha \cdot x^\alpha$ is the sum $\sum_i \alpha_i$ of the exponents, and the highest degree of a non-zero monomial term in a polynomial is the *degree* of the polynomial. Non-zero constants are of degree zero, but the zero polynomial is not attributed a well-defined degree—it is rather considered to be of any degree (it equals $0 \cdot x^\alpha$ for any α).

Monomials (monomer)

The degree of a polynomial (graden til et polynom)

Homogenous polynomials (Homogene polynomer)

A polynomial is said to be *homogenous* if all its monomial terms are of the same degree. For example, the polynomial $x^2y + z^3$ is homogeneous of degree three whereas $x^2y + z^2$ is not; it is still of degree three, but not homogeneous.

Homogenous components of a polynomials (homogene komponenter til et polynom)

Every polynomial may be expressed as a sum of homogeneous polynomials of different degrees—just recollect the homogenous terms with the same degree—and these are called the *homogenous components* of f. They are unambiguously associated with f.

(1.17) Adding two polynomials is simply done term by term, and neither is there any hocus-pocus about multiplying them. The good old pattern is followed where

$$\sum_\alpha a_\alpha x^\alpha \cdot \sum_\beta b_\beta x^\beta = \sum_\gamma \Big(\sum_{\alpha+\beta=\gamma} a_\alpha b_\beta \Big) x^\gamma. \tag{1.3}$$

In particular, the product of monomials comply to the exponential law $x^\alpha x^\beta = x^{\alpha+\beta}$; with this in mind, the content of formula (1.3) is that the product is bilinear over A.

Equipped with the operations just described the set $A[x_1, \dots, x_r]$ of polynomials in the variables $x_1, \dots, x_r$ becomes a ring. Of course, there are axioms to be verified; a tedious and uninteresting process without obstacles, so we voluntarily skip it (such an indolence being reserved for professors, students are urged to do the checking).

EXERCISE 1.12 Let $f = \sum_\alpha a_\alpha x^\alpha$ and $g = \sum_\alpha b_\alpha x^\alpha$ be two non-zero polynomials with coefficients from the ring A such that fg is non-zero. Show that $\deg fg \leqslant \deg f + \deg g$. Show that equality holds when A is an integral domain. Give examples where strict inequality holds. ★

(1.18) There is a notable difference between a polynomial and a polynomial function. Over finite rings, like $\mathbb{Z}/n\mathbb{Z}$ for instance, different polynomials can give rise to identical polynomial functions. Simple examples being polynomials in $\mathbb{F}_2[t]$; that is, polynomials in one variable over the field $\mathbb{F}_2$ with two elements. For instance, such polynomials without constant term and with an even number of non-zero terms will vanish identically as a function on $\mathbb{F}_2$. Over infinite fields however, the two notions coincide.

The universal mapping property

A universal mapping property (en universell avbildningsegenskap)

(1.19) The polynomial ring $A[x_1, \dots, x_r]$ has a so-called *universal mapping property*; one may *freely* assign values to the variables to obtain homomorphisms.

PROPOSITION 1.20 (THE UNIVERSAL MAPPING PROPERTY) *Let A be a ring. Assume given a sequence $b_1, \dots, b_r$ of elements from an A-algebra B. Then there is a uniquely determined algebra homomorphism $\phi\colon A[x_1, \dots, x_r] \to B$ such that $\phi(x_i) = b_i$ for $1 \leqslant i \leqslant r$.*

PROOF: A polynomial p is given as $p = \sum_\alpha a_\alpha x^\alpha$. Since the coefficients a_α are unambiguously determined by p, setting $\phi(p) = \sum_\alpha a_\alpha b_1^{\alpha_1} \cdot \ldots \cdot b_r^{\alpha_r}$ gives a well-defined map which is easily seen to be additive. Since a relation like the one in (1.3) is universally

valid in commutative rings, ϕ respects multiplication as well, and we have an algebra homomorphism. ❑

EXAMPLE 1.17 The universal mapping property is a rather special property most algebras do not have. For instance, the algebra $\mathbb{C}[t^2, t^3]$ from Example 1.13 on page 19 does not have it. That algebra has the generators t^2 and t^3, and the equality $(t^2)^3 = (t^3)^2$ imposes a constraint on the values a homomorphism ϕ can assume on the two generators: it must hold true that $\phi(t^2)^3$ coincides with $\phi(t^3)^2$ (note that there is no such thing as $\phi(t)$). ☆

Two further constructions

There are two further constructions closely related to the construction of the polynomial rings.

(1.21) One may consider polynomial expressions over A in an infinite number of variables $x_1, x_2, \ldots, x_n, \ldots$ although each polynomial merely involves finitely many of the variables. For every n the polynomial ring $A[x_1, \ldots, x_n]$ is obviously contained in $A[x_1, \ldots, x_{n+1}]$, and these polynomial rings thus form a nested sequence of rings. The polynomial ring in countably many variables $A[x_1, x_2, \ldots]$ is just the union of all these. It will also be denotede $A[x_i | i \in \mathbb{N}]$.

EXERCISE 1.13 Convince yourself that the universal mapping property holds even for polynomial rings in infinitely many variables. ★

(1.22) The second type of rings we have in mind, are the rings of *formal power series*. A formal power series is an expression as in (1.2)

Rings of formal power series (ringen av formelle potensrekker)

$$f(x_1, \ldots, x_n) = \sum_{\alpha} a_\alpha x^{\alpha_1} \cdot \ldots x_n^{\alpha_n},$$

except that the sum is not require to be finite (but the summation still extends over multi-indices with the α_i's being non-negative). Addition is done term by term, and the multiplication is defined by formula (1.3), which is legitimate since the expression for each coefficient involves only finitely many terms. The formal power series ring is denoted $A[\![x_1, \ldots, x_n]\!]$.

The case of power series in one single variabel with coefficients in a field k, merits a few comments. The units in the ring $k[\![x]\!]$ are precisly the series with a non-zero constant term; *i. e.* those shaped like $f(x) = a_0 + a_1 x + a_2 + \ldots$ with $a_0 \neq 0$. A potential inverse series $g(x) = b_0 + b_1 x + b_2 + \ldots$ must in addition to $a_0 b_0 = 1$ satisfy the relations

$$a_0 b_n + a_1 b_{n-1} + \ldots + a_n b_0 = 0 \tag{1.4}$$

for $n \geqslant 1$, simply because (1.4) expresses that the terms of degree $n \geqslant 1$ of fg vanish. But since a_0 is invertible, b_n is readily solved from (1.4) in terms of the a_i's and the b_i's for $i < n$, thus allowing the inverse g to be constructed by recursion.

Formal Laurent series (formelle Laurent rekker)

In courses about complex function theory it is taught that meromorphic functions have Laurent series near a pole, and there is a formal analogue of those. So a *formal Laurent series* with coefficients in k is a formal sum $F(x) = \sum_{i \geqslant -n} a_i x^{-i}$. It looks pretty much like a power series, but finitely many terms with negative exponents are allowed. The Laurent series form a ring $k((x))$. Sums of two Laurent series are done term-wise, and the coefficients of the product of $\sum_{i \geqslant -n} a_i x^i$ and $\sum_{i \geqslant -m} b_i x^i$ are defined by the formula

$$c_r = \sum_i a_i b_{r-i},$$

the sum extends *a priori* over all integers i, but in reality it is finite since $b_{r-i} = 0$ for large i and $a_i = 0$ for i small—some checking of axioms is of course necessary, but we leave that to the industrious students.

Every element $g(x)$ is thus on the form $x^n f(x)$ where $n \in \mathbb{Z}$ and $f(x)$ is an invertible power series, and it ensues that $k((x))$ is a field. Clearly $k[\![x]\!]$ is contained in $k((x))$, and every element in $k((x))$ is the quotient between two elements from $k[\![x]\!]$. We say that $k((x))$ is the *fraction field* of $k[\![x]\!]$.

Exercises

⁂ (1.14) *Units in polynomial rings.* Show that a polynomial $f(x) = \sum_i a_i x^i$ in $A[x]$ is invertible if and only if a_0 is invertible and all the other coefficients are nilpotent. HINT: Assume that $f(x) = 1 + a_1 x + \ldots + a_n x^n$ is invertible with inverse $f(x)^{-1} = 1 + b_1 + \ldots + b_m x^m$. Show that $a_n^{i+1} b_{m-i} = 0$ for $0 \leqslant i < n + m$. Conclude that a_n is nilpotent.

(1.15) Let A be a reduced ring. Show that group of units in the polynomial ring $A[x]$ equals A^*.

⁂ (1.16) Assume that k is a field and let t and u be variables. Show that there is no injective ring homomorphism from $k[t, 1/t]$ to $k[u]$.

(1.17) Let k be a (finite) field. Prove that there is infinitely many monic irreducible polynomials with coefficients in k. HINT: Mimic Euclid's proof of the prime numbers being infinitely many.

⁂ (1.18) *Long division.* Let A be a ring and $g(t)$ a polynomial in $A[t]$.

a) Show that the following version of long division by $g(t)$ works in $A[t]$; that is, for any polynomial $f(t) \in A[t]$ there are polynomials $q(t)$ and $r(t)$ with $\deg r < \deg b$ and an element $a \in A$ such that $af(t) = g(t)q(t) + r(t)$.

b) Assume that A is a domain. Conclude that the number of different zeros of a non-zero polynomial in $A[t]$ is less than the degree. In a later example (Example 2.11 on page 38) we shall be exhibit a counterexample when A is not domain.

(1.19) *Monoidal algebras.* In this exercise the definition of polynomial rings is made precise and generalized.

Let G be commutative monoid* written additively. As an abelian group $A[G]$ is the direct sum of copies of A indexed by G; that is, $A[G] = \bigoplus_{\alpha\in G} A$. The elements are sequences $p = (p_\alpha)_{\alpha\in G}$ with finite support, and addition is defined component-wise. Introduce a product on $A[G]$ by the formula

$$(p\cdot q)_\alpha = \sum_{\beta,\gamma\in G,\ \beta+\gamma=\alpha} p_\beta \cdot q_\gamma.$$

Let x^α denote the sequence all whose components are zero apart from the one in the slot with index α, which equals one.

a) Show that the x^α form an additive basis for $A[G]$.
b) Show that $x^\alpha \cdot x^\beta = x^{\alpha+\beta}$.
c) Show that $(\sum_\alpha p_\alpha x^\alpha)x^\beta = \sum_\alpha p_\alpha x^{\alpha+\beta}$. Verify that $A[G]$ is a ring.
d) Show that $A[\mathbb{N}_0^r] \simeq A[x_1,\dots,x_r]$.

(1.20) *The formal derivative.* Let $f(x) = \sum_{0\leqslant i\leqslant d} a_i x^i$ be a polynomial with coefficients in a field k. Copying the classical formula for the derivative of a polynomial, one defines the *formal derivative* $f'(x)$ of f as $f'(x) = \sum_{1\leqslant i\leqslant d} i\cdot a_i x^{i-1}$.

a) Show that derivation is a linear operation and that both the Leibnitz' rule and the chain rule hold true; that is $(f(x)g(x))' = f(x)g'(x) + f'(x)g(x)$ and $\big(f(g(x))\big)' = f'(g(x))g'(x)$.
b) Show that if k is of charactristic zero, then $f' = 0$ if and only if f is constant; *i. e.* f if and only if is of degree zero.
c) Show that if $f' = 0$ and f is not of degree zero, then k has positive characteristic, say p, and $f(x) = g(x^{p^r})$ for some r where g is a polynomial with $g' \neq 0$.

(1.21) Let $\{A_i\}_{i\in I}$ be a collection of subrings of the ring A. Prove that the intersection $\bigcap_{i\in I} A_i$ is a subring.

(1.22) Give examples of two subrings A_1 and A_2 of a ring A such that their union is not a subring. Assume that the collection $\{A_i\}_{i\in I}$ of subrings of A has the property that any two rings from it are contained in a third. Prove that in that case the union $\bigcup_{i\in I} A_i$ is a subring.

★

*A monoid *is a set endowed with an associative binary operation that has a neutral element. It is like a group but with elements lacking an inverse. The set $\mathbb{N}_0$ of non-negative integers and its Cartesian powers $\mathbb{N}_0^r$ are arch-examples of monoids.*

1.3 *Direct products and idempotents*

As an introductory motivation to this section, consider the disjoint union $X \cup Y$ of two topological spaces X and Y. Giving a continuous function on $X \cup Y$ is the same as giving one continuous function on X and one on Y. Therefore the ring of continuous functions $C_{\mathbb{R}}(X \cup Y)$ decomposes as the direct product $C_{\mathbb{R}}(X \cup Y) = C_{\mathbb{R}}(X) \times C_{\mathbb{R}}(Y)$, in which both addition and multiplication are given component-wise. This indicates

that in the interplay between geometry and rings, direct product of rings correspond to disconnected spaces.

Below we shall define the direct product of a collection of rings regardless of its cardinality and introduce the notion of *idempotent elements* (elements e such that $e^2 = e$). Multiplication by idempotents are projection operators (they are equal to their squares) and serve to decompose rings (and later on modules) into direct products.

The archetype of an idempotent function is the characteristic function e_X of a connected component, say X, of a topological space Z; that is, the function that assumes the value one on X and zero on the rest of Z. Since X is a connected component of Z, this function is continuous, and of course, $e_X^2 = e_X$. Moreover, the restriction $f|_X$ to X of any function f on Z equals $f \cdot e_X$, or put more precisely, $f \cdot e_X$ is the restriction $f|_X$ extended by zero to the entire space Z. Anyhow, in this way the set $C_{\mathbb{R}}(Z) \cdot e_X$ is a ring naturally identified with $C_{\mathbb{R}}(X)$ with the idempotent e_X corresponding to the unit element in $C_{\mathbb{R}}(X)$. The lesson learned is that idempotents are algebraic counterparts to the geometric notion of connected components (at least when the components are finite in number).

Direct products of rings

We start out by considering two rings A_1 and A_2. The Cartesian product $A = A_1 \times A_2$ consisting of the pairs (a_1, a_2) becomes a ring when equipped with the componentwise operations. The underlying additive group is the direct product of the underlying groups of the two rings, and the product is given as

$$(a_1, a_2) \cdot (a_1', a_2') = (a_1 \cdot a_1', a_2 \cdot a_2').$$

The unit element is the pair $(1, 1)$, and the two projections $\pi_i \colon A \to A_i$ are ring homomorphisms. Moreover, the direct product possesses two special elements $e_1 = (1, 0)$ and $e_2 = (0, 1)$, which satisfy $e_i^2 = e_i$ and $e_1 e_2 = 0$. The sets $e_1 A$ and $e_2 A$ equal respectively $A_1 \times \{0\}$ or $\{0\} \times A_2$, and are, with a liberal interpretation*, subrings of A isomorphic to respectively A_1 and A_2.

*Since the unit element $(1, 1)$ does not lie in either A_i, they are properly speaking not subrings even though they are closed under both addition and multiplication.

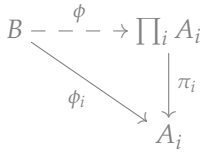
(1.23) To generalize what we just did for a pair of rings, let $\{A_i\}_{i \in I}$ be any collection of rings, which can be of any cardinality. In our context it will mostly be finite, but occasionally will be countable. The direct product $\prod_{i \in I} A_i$ has as underlying additive group the direct product of the underlying additive groups of the A_i's. The elements are "*tuples*" or *strings* $(a_i)_{i \in I}$ indexed* by I whose i-th component a_i belongs to A_i, and the addition of two such is performed component-wise. The same is true of the multiplication, also performed component for component; that is, it holds true that $(a_i) \cdot (b_i) = (a_i \cdot b_i)$. The ring axioms can be checked component-wise and thus come for free.

*The reference to the index set I will frequently be dropped and strings written as (a_i).

Interpreting tuples $a = (a_i)$ as maps $a\colon I \to \bigcup_{i\in I} A_i$, the ring operations of the direct product are just the point-wise operations. The unit element, for instance, is the "constant*" function that sends each index i to 1.

*Why the quotation marks?

(1.24) The projections $\pi_i\colon \prod_{i\in I} A_i \to A_i$ are ring homomorphisms (this is just another way of saying that the ring operations are defined component-wise) and enjoy the following universal property: Given any ring B and any collection $\phi_i\colon B \to A_i$ of ring homomorphisms, there is an unambiguously defined map of rings $\phi\colon B \to \prod_{i\in I} A_i$ such that $\phi_i = \pi_i \circ \phi$ for all $i \in I$. Indeed, this amounts to the map given by $\phi(x) = (\phi_i(x))_{i\in I}$ being a ring homomorphism.



Idempotents

(1.25) In any ring A an element e satisfying $e^2 = e$ is said to be *idempotent*, and if f is another idempotent, one says that f and e are *orthogonal* when $fe = 0$. The element $1 - e$ is always idempotent when e is and is orthogonal to e as shown by the little calculations

Idempotent elements (idempotente elementer)

Orthogonal idempotents (ortogonale idempotenter)

$$(1-e)^2 = 1 - 2e + e^2 = 1 - 2e + e = 1 - e,$$
$$e(1-e) = e - e^2 = e - e = 0.$$

The subset $Ae = \{\, ae \mid a \in A \,\}$ is a ring with e as a unit element. Indeed

$$ae \cdot be = abe^2 = abe,$$

so Ae is closed under multiplication and trivially it is closed under addition as well; finally

$$e \cdot ae = ae^2 = ae,$$

so that e serves as the unit element.

It is common usage to count the unit element and zero among the idempotents; they are called the *the trivial idempotents*.

The trivial idempotents (de trivielle idempotentene)

(1.26) We saw above that in the direct product $A_1 \times A_2$ there appears two natural defined idempotents. Conversely, let A a ring. To any family $\{e_i\}_{1\leqslant i\leqslant r}$ of mutually orthogonal idempotents that add up to 1, there corresponds a decomposition of A as a direct product:

PROPOSITION 1.27 *Let $e_1, \dots, e_r$ be pairwise orthogonal idempotents in a ring A and assume that $\sum_i e_i = 1$. Then each set Ae_i is a subring in the restricted sense, and the association $x \mapsto (xe_1, \dots, xe_r)$ gives an isomorphism of rings*

$$A \xrightarrow{\simeq} \prod_i Ae_i.$$

The projection onto Ae_i is realized as multiplication by e_i.

PROOF: To begin with, we verify that the map in the proposition, call it ϕ, is a ring homomorphisms. So let x and y be two elements from A. Clearly ϕ is additive, moreover, the e_i's being idempotents, we find

$$\phi(x)\phi(y) = (xe_i)_i \cdot (ye_i)_i = (xye_ie_i)_i = (xye_i)_i = \phi(xy),$$

and thus ϕ also respects the multiplication. The unit element 1 maps to the string $(e_i)_i$ which is the unit element in the product since each e_i serves as unit element in Ae_i.

Now, we have supposed that the e_i's add up to one; that is, $1 = \sum_i e_i$. Hence $x = \sum_i xe_i$, from which ensues that ϕ injective; indeed, that $\phi(x) = (xe_i)_i = 0$, means that each $xe_i = 0$.

Finally, let us check that ϕ is surjective. Given an element $(x_ie_i)_i$ in the product, we set $x = \sum_i x_ie_i$. Using that the e_i's are mutually orthogonal, we find $xe_j = \sum_i x_ie_ie_j = x_je_j$, and x maps to the given element $(x_ie_i)_i$. ❑

Exercises

⁂ (1.23) Determine the idempotents in $\mathbb{Z}/12\mathbb{Z}$ and in $\mathbb{Z}/36\mathbb{Z}$.

(1.24) What is the prime ring of $\mathbb{Q} \times \mathbb{F}_p$?

★

Lecture 2

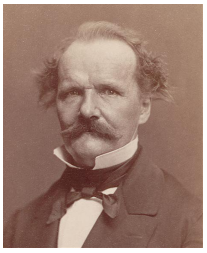
Ideals

Ideals were first defined by Richard Dedekind in 1876, but the name comes from the so called "ideal numbers" of Ernst Eduard Kummer which he introduced in a series of papers around 1847.

Working with rings of integers in algebraic number fields, the algebraists of the period realized that analogues of the Fundamental Theorem of Arithmetic do not always hold in such rings. Recall that the Fundamental Theorem asserts that any integer is a product $n = p_1 \cdot \ldots \cdot p_r$ of signed primes, and that the factors are unique up to order and sign—changing the order of the factors does not affect the product, and changing the sign of one factor can be compensated by simultaneously changing the sign of another. It is not too complicated to show that in a vast class of rings, including the rings of algebraic numbers above, any element can be expressed as a product of irreducible elements; that is, as a product of elements which may not be factored further (they can of course always be altered by a unit, but that is not considered an honest factorization). The point is however, that these factors are not always unique (apart from the innocuous ambiguities caused by unit factors and change of order).

Richard Dedekind
(1831–1916)
German mathematician

The classical example, which is omnipresent in text books, is the factorization $2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5})$ in the ring $\mathbb{Z}[i\sqrt{5}]$. The four involved numbers are all irreducible, and no two of them are related by units.



Ernst Eduard Kummer
(1810–1893)
German mathematician

The ideals came about to remedy this fault and, in fact, in certain rings called Dedekind rings, the situation can be salvaged; there is a factorization theorem for *ideals* replacing the Fundamental Theorem of Arithmetic. Hence the name *ideals*, they were "the ideal numbers"

Dedekind rings are however a very restricted class of rings, and today ideals play an infinitely wider role than just being "ideal numbers". In algebraic geometry for instance, they appear as the sets of polynomials in $k[x_1, \ldots, x_r]$ vanishing along a subset of k^r, and this is the clue to the coupling between algebra and geometry.

2.1 Ideals

Ideals (idealer)

(2.1) Let A be a ring. An additive subgroup $\mathfrak{a}$ of A is called an *ideal* if it is closed under multiplication by elements from A. That is, $\mathfrak{a}$ satisfies the two following requirements; the first merely being a rephrasing that $\mathfrak{a}$ is a subgroup.

- ❏ If $a \in \mathfrak{a}$ and $b \in \mathfrak{a}$, then $a + b \in A$, and $0 \in A$;
- ❏ If $a \in A$ and $b \in \mathfrak{a}$, then $ab \in \mathfrak{a}$.

Both the trivial additive subgroup (0) and the entire ring satisfy these requirements and are ideals, although special ideals. In many texts the ring itself when considered an ideal, is denoted by (1).

Proper ideals (ekte idealer)

(2.2) An ideal $\mathfrak{a}$ is said to be a *proper ideal* if it is not equal to the entire ring. This is equivalent to no member of $\mathfrak{a}$ being invertible. Indeed, if $a \in \mathfrak{a}$ is invertible, one has $b = ba^{-1}a \in \mathfrak{a}$ for any $b \in A$; and if $\mathfrak{a} = A$, of course, $1 \in \mathfrak{a}$. From this observation ensues the following characterization of fields in terms of ideals:

PROPOSITION 2.3 *A ring A is a field if and only if its only ideals are the zero ideal and A itself.*

PROOF: We just saw that an ideal $\mathfrak{a}$ equals A precisely when $\mathfrak{a} \cap A^* \neq \varnothing$. If A is a field, then $A^* = A\backslash\{0\}$, and any ideal, apart from the zero ideal, meets A^*. The other way round, any non-zero and proper ideal must contain a non-zero element, which cannot be invertible, and consequently A is a not a field. ❏

Examples

(2.1) The subset $n\mathbb{Z}$ of $\mathbb{Z}$ consisting of all multiples of the integer n is an ideal; a so-called *principal ideal*. The ideal $n\mathbb{Z}$ is frequently written (n) or $(n)\mathbb{Z}$.

(2.2) For any subset $S \subseteq \mathbb{C}^r$ the polynomials in $\mathbb{C}[x_1, \ldots, x_r]$ vanishing on S form an ideal.

☆

Operations on ideals—the lattice of ideals

(2.4) The set $\mathcal{I}(A)$ of ideals in the ring A has—in addition to being partially ordered under inclusion—a lot of structure. One may form *the intersection* $\bigcap_{i\in I} \mathfrak{a}_i$ of any family $\{\mathfrak{a}_i\}_{i\in I}$ of ideals. It is easily seen to be an ideal, and it is the largest ideal contained in all the $\mathfrak{a}_i$'s. Likewise, one has the notion of *the sum* of a family of ideals. It is the ideal consisting of all finite sums of elements from the $\mathfrak{a}_i$'s:

$$\sum_{i\in I} \mathfrak{a}_i = \{\, a_{j_1} + \ldots + a_{j_r} \mid a_{j_i} \in \mathfrak{a}_i,\ r \in \mathbb{N} \,\},$$

and it is the smallest ideal containing all the $\mathfrak{a}_i$'s. So $\mathcal{I}(A)$ is what one technically calls a *complete lattice*; every subset of $\mathcal{I}(A)$ has a greatest lower bound (the intersection) and a smallest upper bound (the sum). It is the *lattice of ideals* in A.

The lattice of ideals (Ideallattiset(???))

(2.5) A construct similar to the sum of a family of ideals is the ideal generated by a set of elements $\{a_i\}_{i\in I}$ from A. It will be denoted $(a_i|i \in I)$, or in case the set is finite, say equal to $\{a_1, \ldots, a_r\}$, the alternative notations $(a_1, \ldots, a_r)$ or $(a_1, \ldots, a_r)A$ are common usage. Its members are the finite linear combinations of the a_i's with coefficients from the ring A; that is, it holds that

$$(a_i|i \in I) = \{ \textstyle\sum_{i\in J} c_i a_i \mid c_i \in A, J \subseteq I \text{ finite} \}.$$

Generators (generatorer)

The elements a_i are called *generators*. Ideals which are generated by finitely many elements are naturally called *finitely generated*.

Finitely generated ideals (endeliggenererte idealer)

(2.6) An ideal generated by a single element is called a *principal ideal* and is denoted by (a) or by aA. It consists of all multiples of the generator; *i. e.* $(a) = \{ c \cdot a \mid c \in A \}$.

In some rings all ideals are principal as is the case for the integers $\mathbb{Z}$ and the polynomial ring $k[t]$ over a field. These rings, if also being domains, are called *Principal Ideal Domains*, frequently referred to by the acronym PID.

Principal ideals (hovedidealer)

Different elements can of course generate the same principal ideal, but they will, at least in domains, be closely related, as described in the next lemma.

Principal Ideal Domains (hovedidealområder)

LEMMA 2.7 *Two non-zero divisors a and b in a ring A generate the same principal ideal precisely when they are related by a unit; that is, when $a = ub$ where $u \in A^*$.*

PROOF: When $(a) \subseteq (b)$ there is a ring element u so that $a = ub$, and when $(b) \subseteq (a)$ it holds that $b = va$, Hence $a = vua$. And a not being a zero-divisor, we conclude that $vu = 1$. ❏

One often says that a and b are *associates* if one is a unit times the other; *i. e.* if $a = ub$ with $u \in A^*$. The lemma then says that two non-zero divisors a and b generate the same principal ideal if and only if they are associates. So in a domain the set of principal ideals is naturally identified with A modulo association (which easily is seen to be an equivalence relation).

Associates (assosierte elementer)

In the case when the ring A possesses zero-divisors, things are more complicated, see Example 2.10 on page 38 below.

(2.8) The *product* of two ideals $\mathfrak{a}$ and $\mathfrak{b}$ is the ideal generated by all products of one element from $\mathfrak{a}$ and one from $\mathfrak{b}$; that is, the product $\mathfrak{ab}$ is formed of all finite sums of such products:

$$\mathfrak{ab} = \{ a_1b_1 + \ldots + a_rb_r \mid a_i \in \mathfrak{a}, b_i \in \mathfrak{b}, r \in \mathbb{N} \}.$$

The product of ideals (produktet av idealer)

(2.9) The last operation we offer is the formation of *the transporter* between two ideals. Some texts call it the *quotient* of the two deals—however, that term should be reserved

The transporter (transportøren)

for another construction we shortly come to. So let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in A. We define the *transporter* $(\mathfrak{a} : \mathfrak{b})$ to be set of elements which on multiplication send $\mathfrak{b}$ into $\mathfrak{a}$; that is

$$(\mathfrak{a} : \mathfrak{b}) = \{\, x \in A \mid x\mathfrak{b} \subseteq \mathfrak{a} \,\}.$$

It is easily seen to be an ideal. In the particular case that $\mathfrak{a} = (0)$ and $\mathfrak{b}$ is a principal ideal, say $\mathfrak{b} = (a)$, the transporter $(0 : a)$ (an immediate simplification of the notational overloaded expression $((0) : (a))$) coincides with the *annihilator* of a; that is

The annihilator of an element (annihilatoren til et element)

$$(0 : a) = \operatorname{Ann} a = \{\, x \in A \mid xa = 0 \,\}.$$

Similarily, any ideal $\mathfrak{a}$ has an annihilator ideal which is defined as $\operatorname{Ann}\mathfrak{a} = (0 : \mathfrak{a}) = \{\, x \in A \mid xa = 0 \text{ for all } a \in \mathfrak{a} \,\}$.

Examples

(2.3) In $\mathbb{Z}$ it holds that $(100 : 10) = (10)$. More generally if a and b are elements from the ring A and b is not a zero divisor, one has $(ab : b) = (a)$. Indeed, $xb = yab$ is equivalent to $x = ya$ since cancellation by b is allowed b being a non-zero divisor. If b is a zero-divisor it anyhow holds that $(ab, b) = (a) + \operatorname{Ann} b$.

(2.4) In $\mathbb{Z}/40\mathbb{Z}$ one has $\operatorname{Ann} 2 = (20)$, that $\operatorname{Ann} 4 = (10)$ and that $\operatorname{Ann} 20 = (5)$.

(2.5) In the polynomial ring $\mathbb{C}[x, y]$ it holds that $\big((xy, y^2) : (x, y)\big) = (y)$. Clearly (y) is contained in $\big((xy, y^2) : (x, y)\big)$. For the converse inclusion assume that $fx = gxy + hy^2$ where f, g and h are polynomials in $\mathbb{C}[x, y]$. Since x divides the terms fx and gxy, it divides hy^2 as well, and by cancelling x, we infere that $f = gy + h'y$ with $h' \in \mathbb{C}[x, y]$; that is, $f \in (y)$.

☆

Functorially

A map of rings $\phi\colon A \to B$ induces two maps between the ideal lattices $\mathcal{I}(A)$ and $\mathcal{I}(B)$, one in a covariant and one in a contravariant way. One can move ideals forward with the help of ϕ, and the the usual inverse image construct gives a way to move ideals backwards along ϕ. The new ideals are in some texts respectively called *extensions* or *contractions* of the old.

Extensions of ideals (utvidlese av ideal)

Contractions of ideals (tilbaketrekning)

Pullback (tilbaketrekning)

(2.10) We begin with the contravariant way. The inverse image $\phi^{-1}(\mathfrak{b})$ of an ideal $\mathfrak{b}$ in B, also called the *pullback*, is evidently an ideal in A—indeed, $\phi(ab) = \phi(a)\phi(b)$ belongs to $\mathfrak{b}$ whenever $\phi(b)$ does—and this gives rise to a map $\phi^{-1}\colon \mathcal{I}(B) \to \mathcal{I}(A)$. In the frequently reccurring case when A is a subring of B, the reference to the inclusion map is most often suppressed, and one uses the natural notation $\mathfrak{b} \cap A$ for the "pullback" of the ideal $\mathfrak{b}$.

Obviously the inverse image preserves inclusions, and it takes intersections to intersections (pullbacks of sets respect intersections in general). Sums and products

of ideals however, are not generally preserved, but the inclusions in the upcoming proposition are easily verified. One has:

PROPOSITION 2.11 (PULLBACKS) *Let $\phi\colon A \to B$ be a ring homomorphism and let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in B. The pullback map $\phi^{-1}\colon \mathcal{I}(B) \to \mathcal{I}(A)$ preserves inclusions, and the following claims hold true:*

i) $\phi^{-1}(\mathfrak{a}) \cap \phi^{-1}(\mathfrak{b}) = \phi^{-1}(\mathfrak{a} \cap \mathfrak{b})$;
ii) $\phi^{-1}(\mathfrak{a}) + \phi^{-1}(\mathfrak{b}) \subseteq \phi^{-1}(\mathfrak{a} + \mathfrak{b})$;
iii) $\phi^{-1}(\mathfrak{a}) \cdot \phi^{-1}(\mathfrak{b}) \subseteq \phi^{-1}(\mathfrak{a} \cdot \mathfrak{b})$.

Equality does not hold in general in the two last statements, but notice that both inclusions will be equalities when ϕ is a surjective map. We postpone giving examples to the end of the paragraph (Examples 2.6 and 2.7 below).

(2.12) Next we come to the covariant construction. If $\mathfrak{a}$ is an ideal in A, the image $\phi(\mathfrak{a})$ is not necessarily an ideal in B unless ϕ is surjective. A stupid example can be the image of any non-zero ideal in $\mathbb{Z}$ under the inclusion $\mathbb{Z} \subseteq \mathbb{Q}$. The ideal generated by $\phi(\mathfrak{a})$ however is, and we shall usually denote this ideal by $\phi(\mathfrak{a})B$ or simply by $\mathfrak{a}B$; as mentioned above it is called the *extension* of $\mathfrak{a}$, but is also frequently referred to as the *pushout* of $\mathfrak{a}$. This induces a map $\mathcal{I}(A) \to \mathcal{I}(B)$. Inclusions are obviously preserved, and one leisurely verifies the other relations in the following proposition.

Extension of an ideal (utvidelse av et ideal)

Pushout of an ideal (pushout, frempuff, fremskudd)

PROPOSITION 2.13 (PUSHOUTS) *Let $\phi\colon A \to B$ be a map of rings and let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in A. Then the map $\mathfrak{a} \mapsto \mathfrak{a}B$ preserves inclusions. Moreover, the following hold true:*

i) $\phi(\mathfrak{a} \cdot \mathfrak{b})B = \phi(\mathfrak{a})B \cdot \phi(\mathfrak{b})B$;
ii) $\phi(\mathfrak{a} + \mathfrak{b})B = \phi(\mathfrak{a})B + \phi(\mathfrak{b})B$;
iii) $\phi(\mathfrak{a} \cap \mathfrak{b})B \subseteq \phi(\mathfrak{a})B \cap \phi(\mathfrak{b})B$.

The inclusion in the last statement may be strict (see Example 2.9 on page 38), but just like with the previous proposition, equality holds in the third statement whenever ϕ is surjective.

Examples

(2.6) A simple example of strict inclusion in statement *ii)* in Proposition 2.11 above is the diagonal map $\delta\colon A \to A \times A$ that sends a to (a, a). The two ideals $\mathfrak{b} = \{\, (0, a) \mid a \in A \,\}$ and $\mathfrak{b}' = \{\, (a, 0) \mid a \in A \,\}$ are both pulled back to the zero ideal, but since $\mathfrak{b} + \mathfrak{b}' = A \times A$, their sum is pulled back to the entire ring A.

(2.7) We intend to give an example of strict inclusion in *iii)* in Proposition 2.11. Consider the subring $k[xy]$ of the polynomial ring $k[x, y]$ and let ϕ be the inclusion. Let $\mathfrak{a}$ and $\mathfrak{b}$ be the two principal ideals (x) and (y) in $k[x, y]$. We contend that $\mathfrak{a} \cap k[xy] = \mathfrak{b} \cap k[xy] = (xy)$; indeed, clearly $(xy) \subseteq (x) \cap k[xy]$, and equality holds since an identity like

$xf(x,y) = g(xy)$ between polynomials forces g to be without a constant term. Similarily, $(y) \cap k[xy] = (xy)$. So $(\mathfrak{a} \cap k[xy])(\mathfrak{b} \cap k[xy]) = ((xy)^2)$, but $(x) \cap (y) = (xy)k[x,y]$, which intersects $k[xy]$ in the ideal (xy).

☆

Exercises

☼ (2.1) Let $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ be ideals in a ring A.

a) Show that the two relations $\mathfrak{a} \cdot \mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b}$ and $(\mathfrak{a} \cap \mathfrak{b})^2 \subseteq \mathfrak{a} \cdot \mathfrak{b}$ hold. Show by giving examples that there might be a strict inclusion in both cases.

b) Assume that $\mathfrak{a} + \mathfrak{b} = (1)$. Show that $\mathfrak{a} \cdot \mathfrak{b} = \mathfrak{a} \cap \mathfrak{b}$.

c) Show that $\mathfrak{a}(\mathfrak{b} + \mathfrak{c}) = \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{c}$. Show that $\mathfrak{a} \cap \mathfrak{b} + \mathfrak{a} \cap \mathfrak{c} \subseteq \mathfrak{a} \cap (\mathfrak{b} + \mathfrak{c})$, and by exhibiting an example, show that the inclusion can be strict.

(2.2) Let $\{\mathfrak{a}_i\}$ be a collection of ideals in the ring A. Show that for any ideal $\mathfrak{b}$ it holds true that $(\bigcap_{i\in I} \mathfrak{a}_i : \mathfrak{b}) = \bigcap_{i\in I}(\mathfrak{a}_i : \mathfrak{b})$ and that $(\mathfrak{b} : \sum_{i\in I} \mathfrak{a}_i) = \bigcap_{i\in I}(\mathfrak{b} : \mathfrak{a}_i)$.

(2.3) Show that any non-zero ideal in the ring $\mathbb{Z}$ of integers is principal, generated by any of the two members of smallest absolute value. Show that each non-zero ideal in the polynomial ring $k[x]$ over a field k is principal, generated by any member of smallest degree.

(2.4) Given two ideals (n) and (m) in the ring of integers $\mathbb{Z}$.

a) Show that $(n) \subseteq (m)$ if and only if $m|n$. Conclude that the partially ordered set $\mathcal{I}(\mathbb{Z})\backslash\{(0)\}$ of non-zero ideals in $\mathbb{Z}$ is lattice isomorphic to the the set of natural numbers ordered by *reverse* divisibility;

b) Describe the ideals (n,m) and $(n) \cap (m)$. Show that $(n) \cdot (m) = \big((n) \cap (m)\big) \cdot (n,m)$;

c) Describe $((144) : (24))$;

d) Describe the transporter $(n : m)$ in terms of the prime factorizations of the two integers n and m.

(2.5) Let $k[x,y]$ be the polynomial ring in the variables x and y over the field k, and let $\mathfrak{m}$ be the ideal generated by x and y; that is $\mathfrak{m} = (x,y)$. Let n denote a natural number.

a) Exhibit a set of generators for the power $\mathfrak{m}^n$.

b) Let μ and ν be two natural numbers. Show that $\mathfrak{m}^n \subseteq (x^\mu, y^\nu)$ for n sufficiently large. What is the smallest n for which this holds?

☼ (2.6) Let $A = \mathbb{Z}[\sqrt{2}, \sqrt{3}]$. Show that as an abelian group A is free of rank four and exhibit a basis. Show that the underlying abelian groups of the principal ideals $(\sqrt{2})$ and $(\sqrt{3})$ both are of rank four. Exhibit additive bases for both.

★

2.2 *Kernels and quotients*

In one way ideals play the same role in the category of rings as normal subgroups do in the category of groups. They are precisely the subobjects that appear as kernels of homomorphisms, and consequently, the ones that can be factored out.

Kernels of ring homomorphisms (kjernen til en ringavbildning)

(2.14) By definition the *kernel* of a ring homomorphism $\phi\colon A \to B$ is the kernel of ϕ considered a map between the underlying additive groups; that is, it equals the subset of elements mapping to zero, or written in symbols $\ker\phi = \{\, a \in A \mid \phi(a) = 0 \,\}$. If $a \in \ker\phi$ and $b \in A$, we find

$$\phi(a \cdot b) = \phi(a) \cdot \phi(b) = 0 \cdot \phi(b) = 0,$$

and we can conclude that $ab \in \ker\phi$. Hence the kernel $\ker\phi$ is an ideal.

Quotient rings (kvotientringer)

(2.15) To see that any ideal is a kernel, one introduces the concept of *quotient rings*. An ideal $\mathfrak{a}$ in A being an additive subgroup, there is a quotient group $A/\mathfrak{a}$ which consists of the *residue classes* $[a] = a + \mathfrak{a}$ of elements in A. The sum of two such, say $[a]$ and $[b]$, equals $[a + b]$. To put a ring structure on $A/\mathfrak{a}$ we simply define the product of two classes $[a]$ and $[b]$ as

$$[a] \cdot [b] = [a \cdot b] = a \cdot b + \mathfrak{a}.$$

Some checking is needed; the most urgent one being that the product only depends on the residue classes $[a]$ and $[b]$ and not on the choice of representatives a and b. This is encapsulated in the formula where x and y are arbitrary elements from $\mathfrak{a}$

$$(a + x) \cdot (b + y) + \mathfrak{a} = a \cdot b + a \cdot y + b \cdot x + x \cdot y + \mathfrak{a} = a \cdot b + \mathfrak{a}.$$

It is left to the students to verify that this product complies with the associative, commutative and distributive laws. Finally, by definition of the ring operations in $A/\mathfrak{a}$, the quotient map $\pi\colon A \to A/\mathfrak{a}$ that sends a to the residue class $[a]$, is a map of rings whose kernel equals the given ideal $\mathfrak{a}$.

**This examplifies what purpose the null-ring serves; it allows a general existence theorem (avoiding the hypothesis $\mathfrak{a} \neq A$).*

EXAMPLE 2.8 It is appropriate to mention what quotients by the two "extreme" ideals are. The quotient $A/\mathfrak{a}$ equals A if and only if $\mathfrak{a}$ is the zero-ideal, and it equals* the null-ring if and only if $\mathfrak{a} = A$. ☆

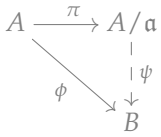
A universal property (en universell egenskap)

(2.16) The quotient ring $A/\mathfrak{a}$ together with the quotient map $\pi\colon A \to A/\mathfrak{a}$ enjoys a so-called *universal property*—the rather pretentious notion "solves a universal problem" is also common usage—which is a convenient way of characterizing many types of mathematical objects. The origin of the technique is found in category theory where objects not always have "elements" and one must rely on "arrows" to express properties.

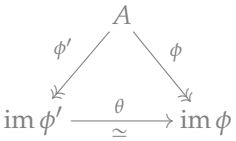
Any map of rings $\phi\colon A \to B$ that vanishes on $\mathfrak{a}$; that is, which satisfies $\mathfrak{a} \subseteq \ker\phi$, factors in a unique way through the quotient $A/\mathfrak{a}$. In other words, there is a unique ring-map $\psi\colon A/\mathfrak{a} \to B$ such that $\phi = \psi \circ \pi$. Indeed, since $\phi(\mathfrak{a}) = 0$, the map ϕ is

constant on every residue class $[a] = a + \mathfrak{a}$, and we put $\psi([a])$ equal to that constant value. This value is forced upon ψ, so ψ is unique, and it is a ring-map since ϕ is. We have proven:

PROPOSITION 2.17 (THE FACTORIZATION THEOREM) *Given an ideal $\mathfrak{a}$ in the ring A. A map of rings $A \to B$ vanishes on $\mathfrak{a}$ if and only if it factors through the quotient map $A \to A/\mathfrak{a}$. The factorization is unique.*



The statement may be illustrated by the first commutative diagram in the margin. The solid arrows are the given ones, and the dashed arrow is the one claimed to exist. If it happens that $\ker\phi = \mathfrak{a}$, the induced map ψ will be injective, and hence, if *a priori* surjective, an isomorphism. The images of all ring-maps with the same kernel are therefore isomorphic, in the strong sense that the isomorphisms fit into diagrams like the second one in the margin.



Ideals in quotients

(2.18) There is a natural one-to-one correspondence between ideals in $A/\mathfrak{a}$ and ideals in A containing the ideal $\mathfrak{a}$. Indeed, if $\mathfrak{b} \subseteq A$ is an ideal with $\mathfrak{a} \subseteq \mathfrak{b}$, the image $\pi(\mathfrak{b})$ equals the additive subgroup $\mathfrak{b}/\mathfrak{a} \subseteq A/\mathfrak{a}$ and since π is surjective, this is an ideal in $A/\mathfrak{a}$. Moreover, if $\mathfrak{c} \subseteq A/\mathfrak{a}$ is an ideal, the inverse image $\pi^{-1}(\mathfrak{c})$ is an ideal in A satisfying $\pi(\pi^{-1}(\mathfrak{c})) = \mathfrak{c}$ (again because π is surjective); or in other words, $\pi^{-1}(\mathfrak{c})$ contains $\mathfrak{a}$ and $\pi^{-1}(\mathfrak{c})/\mathfrak{a} = \mathfrak{c}$.

PROPOSITION 2.19 (IDEALS IN QUOTIENTS) *Let $\mathfrak{a}$ be an ideal in the ring A and let $\pi\colon A \to A/\mathfrak{a}$ the quotient map. The following three statements hold true:*

i) For every ideal $\mathfrak{b}$ in A it holds true that $\pi^{-1}(\pi(\mathfrak{b})) = \mathfrak{b} + \mathfrak{a}$. Each ideal $\mathfrak{c}$ in $A/\mathfrak{a}$ is of the form $\pi(\mathfrak{b}) = \mathfrak{b}/a$; indeed, $\mathfrak{c} = \pi(\pi^{-1}(\mathfrak{c}))$;

ii) The lattice of ideals in $A/\mathfrak{a}$ and the lattice of ideals A containing $\mathfrak{a}$ are isomorphic lattices, with $\mathfrak{c} \mapsto \pi(\mathfrak{c})$ and $\mathfrak{c} \mapsto \pi^{-1}(\mathfrak{c})$ as mutually inverse maps;

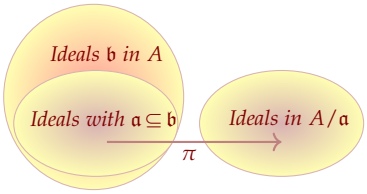
iii) An ideal is mapped to the zero ideal in $A/\mathfrak{a}$ if and only if it is contained in $\mathfrak{a}$.

PROOF: We already saw that $\pi(\mathfrak{b}) = \mathfrak{b}/\mathfrak{a}$ is an ideal and that $\mathfrak{c} = \pi(\pi^{-1}(\mathfrak{c}))$, so the last part of *i)* is clear. For the first claim, if $\pi(x) = \pi(b)$ for some element $b \in \mathfrak{b}$ it evidently holds true that $x - b \in \ker\pi = \mathfrak{a}$ and hence $x \in \mathfrak{b} + \mathfrak{a}$ so that $\pi^{-1}(\pi(\mathfrak{b})) \subseteq \mathfrak{b} + \mathfrak{a}$. The reverse inclusion follows immediately since $\pi(\mathfrak{b} + \mathfrak{a}) = \pi(\mathfrak{b})$, again because $\ker\pi = \mathfrak{a}$.

To show *ii)* observe that the equality $\pi^{-1}(\pi(\mathfrak{b})) = \mathfrak{b} + \mathfrak{a}$ entails that π^{-1} takes values in the sublattice of $\mathcal{I}(A)$ whose members contain $\mathfrak{a}$. It equally implies that $\pi^{-1}(\pi(\mathfrak{b})) = \mathfrak{b}$ whenever $\mathfrak{a} \subseteq \mathfrak{b}$, and we conclude that the two maps are mutual inverses. They both respect inclusions and are thus lattice isomorphisms.

The third claim *iii)* is just a rephrasing of $\mathfrak{a}$ being the kernel of π. ❑

Note that both maps π and π^{-1} respect intersections, sums and products of ideals. The picture below is an illustration of the situation:



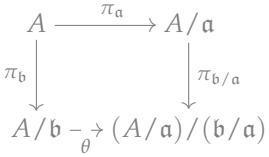
(2.20) The image in $A/\mathfrak{a}$ of an ideal $\mathfrak{b} \subseteq A$, which not necessarily contains $\mathfrak{a}$, is the ideal $(\mathfrak{b}+\mathfrak{a})/\mathfrak{a}$. This holds since obviously $\pi(\mathfrak{b}+\mathfrak{a}) = \pi(\mathfrak{b})$. Now, $\ker \pi|_{\mathfrak{b}} = \mathfrak{a} \cap \mathfrak{b}$ from which ensues the following isomorphism

$$\mathfrak{b}/\mathfrak{b} \cap \mathfrak{a} \simeq (\mathfrak{a}+\mathfrak{b})/\mathfrak{a}. \tag{2.1}$$

The two members of (2.1) are ideals in different rings, so we must be cautious about what isomorphic means (it does not mean equal even though it might seem so the map sending a class $[b]$ to the class $[b]$, but those classes are mod different ideals). The isomorphism is certainly an isomorphism of abelian groups, but it preserves a lot more structure. The two sides are what we later shall call A-modules: Elements from A operate by multiplication on both sides (this evidently holds for ideals in any quotient ring of A), and the isomorphism respects these operations.

Finally, we mention that when $\mathfrak{a}$ and $\mathfrak{b}$ are two ideals with $\mathfrak{a} \subseteq \mathfrak{b}$, there is a natural isomorphism

$$(A/\mathfrak{a})/(\mathfrak{b}/\mathfrak{a}) \simeq A/\mathfrak{b}. \tag{2.2}$$



Indeed, in the diagramme in the margin, the composition $\pi_{\mathfrak{b}/\mathfrak{a}} \circ \pi_{\mathfrak{a}}$ has the ideal $\mathfrak{b}$ as kernel, and therefore factors through $\pi_{\mathfrak{b}}$ by say θ. The map θ is surjective since the composition is and injective since the composition has $\mathfrak{b} = \pi_{\mathfrak{a}}^{-1}(\mathfrak{b}/\mathfrak{a})$ as kernel. The two formulas (2.1) and (2.2) are often referred to as the *Isomorphism Theorems*.

THEOREM 2.21 (THE ISOMORPHISM THEOREM) *Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals in A. Then the following two isomorphism relations hold, where in the second is assumed that $\mathfrak{a} \subseteq \mathfrak{b}$:*

i) $\mathfrak{b}/\mathfrak{b} \cap \mathfrak{a} \simeq (\mathfrak{a}+\mathfrak{b})/\mathfrak{a}$;

ii) $(A/\mathfrak{a})/(\mathfrak{b}/\mathfrak{a}) \simeq A/\mathfrak{b}$.

As remarked above, the isomorphisms are module isomorphisms; they are isomorphisms of abelian groups which respect multiplication by elements from A.

A convenient convention

(2.22) Ever recurring ingredients of a set-up in commutative algebra are rings shaped like quotients $k[X_1, \ldots, X_r]/\mathfrak{a}$ of a polynomial ring. When working with such rings,

it is very natural and suggestive to denote the class of a variable by the lower case variant of the uppercase letter used for the variables. To avoid repeating this formula *ad infinum* like a yogi's mantra, we adhere to the following convention: we say that the ring $A = k[x_1, \dots, x_r]$ has *constituting relations* $f_1(x_1, \dots, x_r) = \dots = f_s(x_1, \dots, x_r) = 0$, if $A = k[X_1, \dots, X_r]/\mathfrak{a}$ with $\mathfrak{a} = (f_1, \dots, f_s)$ and the upper case X_i's correspond to the lower case ones.

Constituting relations (konstituerende relasjoner)

A convenient way of defining ring maps with source A and target a k-algebra B, is to assign values $b_i \in B$ to the generators x_i. This can of course not be done freely, but when the constituting relations persist holding in B when the b_i's are substituted for the x_i's, there is a well-defined and unique ring map $A \to B$ such that $x_i \mapsto b_i$. This ensues from the Factorization Theorem (Theorem 2.17 on page 36); indeed, sending X_i to b_i defines a map $k[X_1, \dots, X_r] \to B$ which factors via A since it vanishes on the ideal $\mathfrak{a}$.

Examples

(2.9) This example aim at illustrating that strict inclusion in the last inequality of Proposition 2.13 on page 33 may occur. So let $A = k[X, Y, Z]$ and let $B = k[x, y, z]$ with constituting relation $zx = zy$, and consider the natural map $\phi\colon A \to B$ that sends upper case letters to their lower case versions. Let $\mathfrak{a}$ and $\mathfrak{b}$ be the principal ideals (X) and (Y) in A. We conted that $\phi(\mathfrak{a} \cap \mathfrak{b}) \subset \phi(\mathfrak{a}) \cap \phi(\mathfrak{b})$.

It holds that $\mathfrak{a} \cap \mathfrak{b} = (XY)$, so $(\mathfrak{a} \cap \mathfrak{b}) \cdot B = (xy)$. Since $zx = zy$, we see that and $zx \in (x) \cap (y) = \mathfrak{a}B \cap \mathfrak{b}B$, but $zx \notin (xy)B$; indeed, one way of seeing this is to observe that sending z to 1 and both x and y to t gives a well-defined ring map $k[x, y, z] \to k[t]$ (since the relation $zx = zy$ persists as $t = t$), which maps zx to t and xy to t^2. We deduce that $\phi(\mathfrak{a} \cap \mathfrak{b}) \subsetneq \phi(\mathfrak{a}) \cap \phi(\mathfrak{b})$.

(2.10) Let $A = k[x, y, z]$ with constituting relation $z = zxy$. In A the two principal ideals (z) and (xz) coincide, but there is no unit $u \neq 1$ in A so that $uz = xz$; hence z and xz are not associates even though $(z) = (xz)$.

The salient point is that $A^* = k^*$. One way of seeing this, is to observe that killing z gives a well-defined ring map $A \to k[x, y]$. It takes units to units, and the group of units in the polynomial ring equals k^*. So u would be a scalar. Setting z equal to 1 and $x = y^{-1}$ gives a ring map $A \to k[y, y^{-1}]$, and in the latter ring obviously x is not a scalar being equal to y^{-1}.

(2.11) Let $A = k[x_i | i \in \mathbb{N}]$ with constituting relations $x_i^2 = -1$ for $i \in \mathbb{N}$. Then the polynomial $t^2 + 1$ has infinitely many roots in A. The ring A is not an integral domain in that $x_i^2 = x_j^2$ fot all i and j; so that $(x_i + x_j)(x_i - x_j) = 0$.

☆

Exercises

(2.7) Let k be a field and $\phi\colon k \to A$ a ring homomorphism. Show that ϕ is injective unless A is the null ring.

(2.8) Let A be a ring and let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in A. Show that there is a natural equality $\mathfrak{b}(A/\mathfrak{a}) = (\mathfrak{a}+\mathfrak{b})/\mathfrak{a}$. Use the Isomorphism Theorem to show that there is a canonical isomorphism $A/\mathfrak{a}+\mathfrak{b} \simeq (A/\mathfrak{a})/\mathfrak{b}(A/\mathfrak{a})$, so that you can divide out by a sum of two ideals by successively dividing out by one at a time. Of course, the order doesn't matter: swapping the two ideals yields an isomorphism $A/\mathfrak{a}+\mathfrak{b} \simeq (A/\mathfrak{b})/\mathfrak{a}(A/\mathfrak{b})$.

★

2.3 *Prime ideals and maximal ideals*

Two classes of ideals are infinitely more important than others. We are speaking about the *prime ideals* and the *maximal ideals*. The prime ideals are defined in terms of multiplicative properties of the ring, and are generalizations of prime numbers. They played the role of the primes in Kummer and Dedekind's world of "ideal numbers". If your ambitions are high and you try to understand all ideals in a ring, you have to begin with understanding the prime ideals, and then accomplish the draconian task to explain how other ideals are built from the prime ideals.

Prime ideals (primidealer)

(2.23) An ideal $\mathfrak{a}$ in a ring A is a *prime ideal* if it is proper and satisfies the following requirement:

- If $ab \in \mathfrak{a}$, then either $a \in \mathfrak{a}$ or $b \in \mathfrak{a}$.

Maximal ideals (maksimale idealer)

Maximal ideals are defined in terms of inclusions. They are, as the name indicates, maximal among the *proper* ideals; that is, they are maximal elements in the partially ordered set $\mathcal{I}(A)\backslash\{A\}$. So an ideal $\mathfrak{a}$ is *maximal* if it is proper and satisfies the following requirement:

- If $\mathfrak{b}$ is an ideal and $\mathfrak{a} \subseteq \mathfrak{b}$, then either $\mathfrak{a} = \mathfrak{b}$ or $\mathfrak{b} = A$.

Notice that both prime ideals and maximal ideals are proper by definition.

(2.24) One has the following characterization of the two classes of ideals in terms of properties of quotients.

PROPOSITION 2.25 *An ideal $\mathfrak{a}$ in A is a prime ideal if and only if the quotient $A/\mathfrak{a}$ is an integral domain. The ideal $\mathfrak{a}$ is maximal if and only if $A/\mathfrak{a}$ is a field.*

PROOF: The quotient $A/\mathfrak{a}$ is an integral domain if and only if $[a][b] = 0$ implies that either $[a] = 0$ or $[b] = 0$; that is, if and only if $ab \in \mathfrak{a}$ implies that either $a \in \mathfrak{a}$ or $b \in \mathfrak{a}$, which proves the first assertion.

Bearing in mind the relation between ideals in $A/\mathfrak{a}$ and ideals in A containing $\mathfrak{a}$ (as in Proposition 2.19 on page 36), the second assertion is pretty obvious. There is no ideal strictly between $\mathfrak{a}$ and A if and only if $A/\mathfrak{a}$ has no non-trivial proper ideal; that is, if and only if $A/\mathfrak{a}$ is a field (Proposition 2.3 on page 30). ❏

Notice that the zero ideal (0) is a prime ideal if and only if A is an integral domain, and it is maximal if and only if A is a field. When $\mathfrak{m}$ is a maximal ideal, the field $A/\mathfrak{m}$ is called *the residue class field* of A at $\mathfrak{m}$ and now and then denoted by $k(\mathfrak{m})$.

Residue class fields (restklassekropper)

Since fields are integral domains, we see immediately that maximal ideals are prime. The converse does not hold as we shortly shall see examples of (Example 2.14 below).

PROPOSITION 2.26 *A maximal ideal $\mathfrak{m}$ is prime.*

(2.27) Not only for elements is it true that a product lies in a prime ideal only when one of the factors does, the same applies to products of ideals as well:

PROPOSITION 2.28 *Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in A such that $\mathfrak{ab}$ is contained in the prime ideal $\mathfrak{p}$. Then either $\mathfrak{a}$ or $\mathfrak{b}$ is contained in $\mathfrak{p}$.*

PROOF: If neither $\mathfrak{a}$ nor $\mathfrak{b}$ lies in $\mathfrak{p}$, one may find elements $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ not being members of $\mathfrak{p}$. Since $\mathfrak{ab}$ is contained in $\mathfrak{p}$, the product ab belongs to $\mathfrak{p}$, and since $\mathfrak{p}$ is prime, it either holds that $a \in \mathfrak{p}$ or that $b \in \mathfrak{p}$. Contradiction. ❏

The claim is not restricted to products of only two ideals. With an easy induction one proves that if a finite product $\mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r$ of ideals is contained in a prime ideal $\mathfrak{p}$, one of the factors $\mathfrak{a}_i$ lies in $\mathfrak{p}$.

Prime ideals in quotients

(2.29) In the correspondence between ideals in A and $A/\mathfrak{a}$ described in Proposition 2.19 on page 36 prime ideals correspond to prime ideals (containing $\mathfrak{a}$) and maximal ideals to maximal ideal (containing $\mathfrak{a}$). The last statement is clear since with the notation as in Proposition 2.19 the inverse image map π^{-1} is an isomorphisms of the lattice $\mathcal{I}(A/\mathfrak{a})$ with the sublattice of $\mathcal{I}(A)$ whose members contain $\mathfrak{a}$, and hence maximal elements correspond to maximal elements. The first ensues from the general truth that prime ideals pull back to prime ideals along ring maps ; indeed, assume that $\mathfrak{p}$ is a prime ideal in B and that $\phi\colon A \to B$ a ring map. That the product ab lies in the inverse image $\phi^{-1}(\mathfrak{p})$, means that $\phi(ab) \in \mathfrak{p}$, but $\phi(ab) = \phi(a)\phi(b)$, and hence either $\phi(a)$ lies in $\mathfrak{p}$ or $\phi(b)$ lies there; that is, either $a \in \phi^{-1}(\mathfrak{p})$ or $b \in \phi^{-1}(\mathfrak{p})$. We have thus established:

PROPOSITION 2.30 (PRIME AND MAXIMAL IDEALS IN QUOTIENTS) *Assume A is a ring and $\mathfrak{a}$ an ideal. The prime ideals in the quotient $A/\mathfrak{a}$ are precisely those of the form $\mathfrak{p}/\mathfrak{a}$ with $\mathfrak{p}$ a prime ideal in A containing $\mathfrak{a}$, and the maximal ideals are those shaped like $\mathfrak{m}/\alpha$ with $\mathfrak{m}$ a maximal ideal in A likewise containing $\mathfrak{a}$.*

Examples

(2.12) The archetype of maximal ideals are the kernels of evaluation maps. For instance, let $a = (a_1, \dots, a_r)$ be a point in k^r where k is any field, and consider the map $k[x_1, \dots, x_r] \to k$ sending a polynomial f to its value $f(a)$ at a. The kernel $\mathfrak{m}$ is a maximal ideal since $k[x_1, \dots, x_r]/\mathfrak{m}$ is the field k. The kernel may be described as $\mathfrak{m} = (x_1 - a_1, \dots, x_r - a_r)$. This is obvious when a is the origin, and introducing fresh coordinates $x_i' = x_i - a_i$, one reduces the general case to that case.

(2.13) The zero ideal in A is prime if and only if A is a domain, and it is maximal if and only if A is a field.

(2.14) There are plenty of prime ideals that are not maximal. Continuing the previous example the ideal $\mathfrak{p}$ generated by a proper subset of the variables is prime but not maximal; that is, after eventually renumbering the variables, $\mathfrak{p} = (x_1, \dots, x_s)$ with $s < r$ is prime but not maximal. This is best seen by considering the partial evaluation map $k[x_1, \dots, x_r] \to k[x_{s+1}, \dots, x_r]$ that sends a polynomial $f(x_1, \dots, x_r)$ to $f(0, .., 0, x_{s+1}, \dots, x_r)$, whose kernel is $\mathfrak{p}$. Since the polynomial ring $k[x_{s+1}, \dots, x_r]$ is a domain, it ensues that $\mathfrak{p}$ is prime, and $\mathfrak{p}$ is obviously not maximal as $s < r$. By a linear change of variable one also shows that the ideals $(x_1 - a_1, \dots, x_s - a_s)$ are all prime.

(2.15) Consider the ring of Gaussian integers $\mathbb{Z}[i]$. It is isomorphic to the quotient $\mathbb{Z}[t]/(t^2 + 1)$ of the polynomial ring $\mathbb{Z}[t]$, the isomorphism sends t to i. Let $p \in \mathbb{Z}$ be a prime number and consider the ideal $p\mathbb{Z}[i]$. Citing Exercise 2.8 on page 39 we infer that $\mathbb{Z}[t]/(p, t^2 + 1)$ on the one hand will be isomorphic to $\mathbb{Z}[i]/p\mathbb{Z}[i]$ and on the other to the quotient $\mathbb{F}_p[t]/(t^2 + 1)$, and we conclude that there is an isomorphism $\mathbb{Z}[i]/p\mathbb{Z}[i] \simeq \mathbb{F}_p[t]/(t^2 + 1)$ that swaps t and i.

☆

Prime avoidance and a pair of twin lemmas

A lemma about prime ideals that will be useful now and then, is the so-called Prime Avoidance Lemma. It asserts that an ideal contained in a finite union of prime ideals must lie entirely in one of them. The name stems from the equivalent statement that if an ideal $\mathfrak{a}$ is not contained in any of the prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$, it has an element not lying in any of the $\mathfrak{p}_i$'s.

(2.31) As a warm up, let us do the case of two prime ideals, in which case the statement is simply a statement about abelian groups: If a subgroup B of an abelian group is contained in the union of two others, A_1 and A_2, it is contained in one of them; indeed, assume not and pick elements $x_i \in A_i \cap B$ but $x_i \notin A_j$ for $\{i, j\} = \{1, 2\}$. Then $x_1 + x_2 \in B$ but $x_1 + x_2 \notin A_1 \cup A_2$, for were it in A_j, it would follow that $x_i \in A_j$, again with $\{i, j\} = \{1, 2\}$, and this is not the case.

The corresponding statement for three subgroups is faulty as shows the vector space $\mathbb{F}_2 \oplus \mathbb{F}_2$ with four elements. It is the union of the three one-dimensional subspaces it has; so for a general claim involving more than two ideals to be true, some multiplicative structure is required, but the case of two groups is reflected in the assertion in that two of the ideals need not be prime:

LEMMA 2.32 (PRIME AVOIDANCE LEMMA) *Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ be ideals in the ring A of which all but at most two are prime. If $\mathfrak{a}$ is an ideal contained in the union $\bigcup_i \mathfrak{a}_i$, then $\mathfrak{a}$ is contained in at least one of the $\mathfrak{a}_j$'s.*

PROOF: We shall assume that $\mathfrak{a}$ is not contained in any of the $\mathfrak{a}_i$'s and search for an element in $\mathfrak{a}$ not lying in the union $\bigcup_i \mathfrak{a}_i$; that is, not belonging to any of the $\mathfrak{a}_i$'s. The proof proceedes by induction on r, the case $r = 2$ already being settled. So we assume $r > 2$ and that the lemma holds true for $r - 1$. Hence $\mathfrak{a}$ is not contained in the union $\bigcup_{j\neq i} \mathfrak{a}_j$ for any i. We can therefore for each i pick an element x_i in $\mathfrak{a}$ not in $\bigcup_{j\neq i} \mathfrak{a}_j$, and we may safely assume that $x_i \in \mathfrak{a}_i$ (were it not, we would be through). Since $r \geqslant 2$ at least one of the $\mathfrak{a}_i$'s is a prime ideal, and we may as well assume that it is the case for $\mathfrak{a}_r$. With these assumptions, we contend that the element

$$x = x_1 \cdot \ldots \cdot x_{r-1} + x_r,$$

which clearly lies in $\mathfrak{a}$, does not belong to any $\mathfrak{a}_i$. If $i \leqslant r - 1$ this holds because x_i lies in $\mathfrak{p}_i$, but x_r does not. For $i = r$ we know that x_r lies in $\mathfrak{a}_r$, but $x_1 \cdot \ldots \cdot x_{r-1}$ does not since none of the factors lie there and $\mathfrak{a}_r$ is prime, so $x \notin \mathfrak{a}_r$ as well. ❑

Notice that the proof merely requires $\mathfrak{a}$ to be closed under addition and multiplication, so the ideal $\mathfrak{a}$ may be replaced with a "weak subring" of A; that is, a subset closed under addition and multiplication*. The second remark to make is that if A is an algebra over an infinite field, one may even skip the requirement that the ideals be prime (Exercise 2.12 below).

Lacking a unit element it is not genuine ring according to our conventions

(2.33) At several later occasions unions and intersections of prime ideals will play an important role, and we use the occasion to introduce some terminology.

Redundant (redundant)

A union $\bigcup_i S_i$ of sets is said to be *redundant* if one of the sets can be discarded without the union changing. This means that for some index ν it holds that $S_\nu \subseteq \bigcup_{i\neq r} S_i$.

Irredundant unions (irredundante unioner)

If the union is not redundant, naturally one calls it *irredundant*. For finite unions of prime ideals the Prime Avoidance Lemma entails that the union $\bigcup_i \mathfrak{p}_i$ is irredundant if and only there is no inclusion relation among the $\mathfrak{p}_i$'s. Indeed, if there is such a relation, the union is obviously redundant, and if say $\mathfrak{p}_\nu \subseteq \bigcup_{i\neq\nu} \mathfrak{p}_i$, the lemma gives that $\mathfrak{p}_\nu$ is contain in one of the other $\mathfrak{p}_i$'s.

Similarly, an intersection $\bigcap_i S_i$ is irredundant if one cannot discard one of sets without changing the intersection. For a finite intersection of prime ideals Proposition 2.28

on page 40 implies that the intersection being irredundant is equivalent to there being no inclusions among the prime ideals.

(2.34) Irredundant unions and intersections of prime ideals enjoy strong uniqueness properties; in fact, the prime ideals involved are determined by their intersection or their union, as expressed in the following twin lemmas.

LEMMA 2.35 *Let $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$ and $\{\mathfrak{q}_1, \ldots, \mathfrak{q}_s\}$ be two families of prime ideals having the same union; that is, $\mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_r = \mathfrak{q}_1 \cup \cdots \cup \mathfrak{q}_s$. Assume that there are no non-trivial inclusion relations in either family. Then the two families coincide.*

PROOF: For each index ν one has $\mathfrak{p}_\nu \subseteq \bigcup_j \mathfrak{q}_j$ and the Prime Avoidance Lemma gives that there is an index $\alpha(\nu)$ so that $\mathfrak{p}_\nu \subseteq \mathfrak{q}_{\alpha(\nu)}$. By symmetry, for each μ there is a $\beta(\mu)$ such that $\mathfrak{q}_\mu \subseteq \mathfrak{p}_{\beta(\mu)}$. Now

$$\mathfrak{p}_\nu \subseteq \mathfrak{q}_{\alpha(\nu)} \subseteq \mathfrak{p}_{\beta(\alpha(\nu))},$$

and since there are no non-trivial inclusion relations among the $\mathfrak{p}_i$'s, we infer that $\beta(\alpha(\nu)) = \nu$. In a symmetric manner one shows that $\alpha(\beta(\mu)) = \mu$; so α is a bijection from $\{1, \ldots, r\}$ to $\{1, \ldots, s\}$ with $\mathfrak{p}_\nu = \mathfrak{q}_{\alpha(\nu)}$, and we are through. ❑

LEMMA 2.36 *Let $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$ and $\{\mathfrak{q}_1, \ldots, \mathfrak{q}_s\}$ be two families of prime ideals having the same intersection; that is, $\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_r = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_s$. Assume that there are no non-trivial inclusion relations in either family. Then the two families coincide.*

PROOF: For each index ν one has $\mathfrak{p}_1 \ldots \mathfrak{p}_r \subseteq \bigcap \mathfrak{q}_j \subseteq \mathfrak{q}_\nu$ and therefore at least for one index, say $\alpha(\nu)$, the relation $\mathfrak{p}_{\alpha(\nu)} \subseteq \mathfrak{q}_v$ holds. By symmetry, for each μ there is a $\beta(\mu)$ such that $\mathfrak{q}_{\beta(\mu)} \subseteq \mathfrak{p}_\mu$. Now

$$\mathfrak{p}_{\alpha(\beta(\mu))} \subseteq \mathfrak{q}_{\beta(\mu)} \subseteq \mathfrak{p}_\mu,$$

and there being no non-trivial inclusion relations among the $\mathfrak{p}_i$'s we may conclude that $\alpha(\beta(\mu)) = \mu$. In a symmetric manner one shows that $\beta(\alpha(v)) = \nu$ and we can conclude that α is a bijection from $\{1, \ldots, r\}$ to $\{1, \ldots, s\}$ with with $\mathfrak{p}_{\alpha(v)} = \mathfrak{q}_v$ ❑

Exercises

⁂ (2.9) Let $\mathfrak{p}$ be a prime ideal in a ring A. Show that $\mathfrak{p}A[t]$ is prime.

⁂ (2.10) Prove that pullback of prime ideals are prime, but show by examples that pullbacks of maximal ideals need not be maximal. Show by giving examples that the extension of a prime ideal is not necessarily prime.

⁂ (2.11) Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in a ring A, furthermore let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be prime ideals in A. Show that if $\mathfrak{a} \backslash \mathfrak{b}$ is not contained in any of the $\mathfrak{p}_i$'s, then $\mathfrak{a}$ is not contained in the union $\bigcup_i \mathfrak{p}_i$.

⁂ (2.12) Assume that A is an algebra over an infinite field; show that the Prime Avoidance Lemma persists being true without any of the $\mathfrak{p}_i$'s being prime. HINT: Prove a "vector subspace avoiding lemma" over infinite fields.

2.4 *Primes and irreducibles*

There are two familiar characterizations of prime numbers. One says a number is prime precisely when the only factors are plus or minus one and the number itself, and the other asserts a number is prime precisely when it divides one of the factors when dividing a product. These two aspects of prime numbers generalize to separated notions, which are not equivalent in general.

Prime elements (primelementer)

(2.37) The first of the twin notions is that of *prime elements*. The definition is verbatim the same as the second characterizations above: A *prime element* in a ring A is an element a which is neither zero nor a unit and is such that if a divides a product, it divides one of the factors; in other words, a relation like $bc = ya$ for some y implies $c = xa$ or $b = xa$ for some x; or expressed in symbols, $a|bc$ implies that $a|b$ or $a|c$. The property is not restricted to products with two factors, a straightforward induction proves that if a is a prime element that divides the product $b_1 \cdots b_r$, it divides one of the factors.

The concept of a prime ideal is also inspired by that of prime numbers, and for principal ideals the two coincide; an element a being prime is equivalent to the principal ideal (a) being a prime ideal.

PROPOSITION 2.38 *A principal ideal (a) is a prime ideal exactly when a is a prime element.*

PROOF: Recall what a being prime means: If $a|bc$ then either $a|b$ or $a|c$. Translated into a statement about ideals the divisibility relation $x|y$ means that $y \in (x)$. Hence, $bc \in (a)$ is equivalent to $a|bc$, and $b \in (a)$ or $c \in (a)$ to respectively $a|b$ or $a|c$. ❑

Irreducible elements (irreduktible elementer)

(2.39) The other aspect of prime numbers is that they can not be further factored; that is, their sole factors are ± 1 and the prime itself. Irreducible polynomials in $k[x]$ share this quality except that they can be changed by non-vanishing constant factors (of course, $f = c^{-1} \cdot cf$ for any non-zero constant c). Generalizing this, one says that a non-zero element a from a ring A is *irreducible* if it is not a unit, and if a relation $a = bc$ implies that either b or c is a unit. This can be phrased in terms of a certain maximality condition for principal ideals.

PROPOSITION 2.40 *An element a in the ring A is irreducible if and only if (a) is maximal among the proper principal ideal.*

PROOF: A relation $a = bc$ is equivalent to an inclusion $(a) \subseteq (b)$, and when (a) enjoys the maximality property it ensues that either $(a) = (b)$ and c is a unit, or $(b) = A$ and b is a unit and a is irreducible. Assume then that a is irreducible and let (b) be a proper principal ideal containing (a), which means that $a = cb$. So c is a unit since b is not, the ideal (b) being proper, and we deduce that $(a) = (b)$. ❑

PROPOSITION 2.41 *Every prime element in a domain A is irreducible.*

PROOF: Assume that a is prime element in A and that $a = bc$. Since a is prime, it holds true that $b = xa$ or $c = xa$ for some $x \in A$, say $b = xa$. Substituting back yields $a = xca$, and cancelling a, which is legal since A is supposed to be a domain, we arrive at $1 = xc$, which shows that c is a unit. ❑

The converse of this proposition is not generally valid, in fact one is tempted to say that in most rings it does not hold. There are simple examples of irreducibles not being prime in quadratic extensions of $\mathbb{Z}$. We give one, the standard one you find in every text, in the ring $\mathbb{Z}[\sqrt{-5}]$ below (Example 2.18 on page 46). We will, however, shortly meet classes of rings where it is true (see Proposition 2.44).

(2.42) Rings in which all ideals are principal, the PID's, are among the easiest rings to understand. One of the particular properties they enjoy is that there is no distinction between maximal ideals and non-zero prime ideals.

PROPOSITION 2.43 *In a principal ideal domain A, any non-zero prime ideal is maximal.*

PROOF: A non-zero prime ideal is generated by a prime element a, and as any other prime element, a is irreducible. From Proposition 2.40 above ensues that (a) then is maximal among the proper principal ideals, but all ideals being principal, (a) is maximal. ❑

Neither is there any distinction between prime and irreducible elements:

PROPOSITION 2.44 *In a principal ideal domain A an element is prime if and only if it is irreducible.*

PROOF: An irreducible element a generates according to Proposition 2.40 above an ideal maximal among the proper principal ideals, but because all ideals are principal, (a) is a maximal ideal. Hence it is a prime ideal, and a is a prime element. ❑

Examples

(2.16) In the polynomial ring $k[x]$ over a field k, principal ideals $(f(x))$ with f irreducible, are maximal ideals. The quotient $k[x]/(f(x))$ is the field obtained from k by adjoining a root of f. If you wonder what that root is, it is just the residue class $[x]$ of the variable x. This illustrates the saying that what matters in modern mathematics is "what objects do, not what they are", or as Obi-Wan Kenobi in Star Wars teaches Luke Skywalker: "Do not equate ability with appearance".

(2.17) The quotient $\mathbb{R}[x]/(x^2+1)$ is isomorphic to $\mathbb{C}$ as one sees by mapping x to i. In a similar vein, if p is a prime number, the polynomial $\Phi_p(x) = x^{p-1} + x^{p-2} + \ldots + x + 1$ is irreducible over $\mathbb{Q}$, so that $\mathbb{Q}[x]/(\Phi_p(x))$ is a field. Sending x to a primitive p^{th}-root of unity ξ , gives an isomorphism with $\mathbb{Q}(\xi)$.

(2.18) Simple and concrete examples of irreducible elements that are not prime are easily found in the ring $\mathbb{Z}[i\sqrt{5}]$ where among others the relation

$$2 \cdot 3 = (1 + i\sqrt{5}) \cdot (1 - i\sqrt{5}) \tag{2.3}$$

holds. For instance, it follows that 2 is not a prime element since it neither divides $(1 + i\sqrt{5})$ nor $(1 - i\sqrt{5})$: indeed, squares of absolute values of members of $\mathbb{Z}[i\sqrt{5}]$ are natural numbers so that if a relation $2 = x \cdot (1 + i\sqrt{5})$ held, we would find

$$4 = |2|^2 = |x|^2|(1 + i\sqrt{5})|^2 = |x|^2 6,$$

with $|x| \in \mathbb{N}$, which absurd. The element 2 is however irreducible. Indeed, a factorization $2 = zw$ yields $2 = |z||w|$, which entails that either $|z| = 1$ or $|w| = 1$, and in view of the units in $\mathbb{Z}[i\sqrt{5}]$ precisely being the elements of norm one (Exercise 1.10 on page 21), either z or w would be a unit. Of course, the three other numbers appearing in (2.3) are irreducible as well, and Exercise 2.17 below asks you to check this. For a generic example of irreducible elements not being prime, see Exercises 2.53 and 2.54 on page 62.

Euclidean function (Euklidsk funksjon)

(2.19) *The ring $\mathbb{Z}[i]$ of Gaussian integers is a* PID*:* The absolute value works as a so-called *Euclidean function* on the ring $\mathbb{Z}[i]$, which means there is division algorithm valid in $\mathbb{Z}[i]$ similar to the Euclidean algorithm for integers. For any two given Gaussian integers a and b with $b \neq 0$, one may find two others, the quotient q and the remainder r, so that $a = qb + r$, and most importantly, the remainder satisfies $|r| < |b|$.

To establish this, observe that geometrically the Gaussian integers form the integral lattice in the complex plane; that is, the set of points with both coordinates integers. Given two Gaussian integers a and b with $b \neq 0$, the distance from ab^{-1} to the nearest point in the integral lattice is obviously less then half the diagonal of a lattice square; that is, there is an element $q \in \mathbb{Z}[i]$ so that $|ab^{-1} - q| \leqslant \sqrt{2}/2 < 1$. Putting $r = b(ab^{-1} - q)$, we have $a = qb + r$ with $|r| < |b|$.

Now, any ideal in $\mathbb{Z}[i]$ is generated by a shortest non-zero member a_0. Indeed, if $a \in \mathfrak{a}$ divide a by a_0 to obtain $a = qa_0 + r$ with $|r| < |a_0|$. But $r = a - qa_0$ lies in $\mathfrak{a}$ and since a_0 is the shortest non-zero member of $\mathfrak{a}$, it ensues that $r = 0$.

☆

Exercises

✲ (2.13) Let $p_1, \ldots, p_r$ be prime numbers and let $A = \mathbb{Z}/(p_1 \cdot \ldots \cdot p_r)$. Show that the prime ideals in A are precisely the principal ideals (p_i). Prove that A/p_iA is the field

$\mathbb{F}_{p_i}$ with p_i elements. How many elements does (p_i) have? An how many are there in the principal ideal $(p_1 \cdot \ldots \cdot \widehat{p_i} \cdot \ldots \cdot p_r)$ (the "hat" indicating that p_i is not part of the product).

(2.14) Find an explicit isomorphism between $\mathbb{R}[x]/(x^2+x+1)$ and $\mathbb{C}$.

✲ (2.15) *Primes in the Gaussian integers.* The aim of this exercise is to analyse the primes in the ring of Gaussian integers $\mathbb{Z}[i]$. Let p be a prime number.

a) Assume that p is odd. Show that there is an exact sequence of multiplicative groups

$$1 \longrightarrow \mu_2 \longrightarrow \mathbb{F}_p^* \xrightarrow{\phi} \mathbb{F}_p^* \xrightarrow{\psi} \mu_2 \longrightarrow 1$$

where $\mu_2 = \{\pm 1\}$, and the maps ϕ and ψ are given as $\phi(x) = x^2$ and $\psi(x) = x^{(p-1)/2}$;

b) Conclude that -1 has a square root in $\mathbb{F}_p$ is and only if $p \equiv 1 \mod 4$;

c) Show that x^2+1 is irreducible over the field $\mathbb{F}_p$ if and only if p is odd and $p \not\equiv 1 \mod 4$;

d) Show that the ring $\mathbb{F}_p[x]/(x^2+1)$ is isomorphic to the field $\mathbb{F}_p(\sqrt{-1})$ if p is odd and $p \not\equiv 1 \mod 4$ and isomorphic to $\mathbb{F}_p \times \mathbb{F}_p$ if $p \equiv 1 \mod 4$. Show that if $p = 2$, it is isomorphic to $\mathbb{F}_2[x]/(x^2)$;

e) Consider the ring of Gaussian integer as an extension $\mathbb{Z} \subseteq \mathbb{Z}[i]$. Discuss the possible shapes of the quotient $\mathbb{Z}[i]/p\mathbb{Z}[i]$ where $p \in \mathbb{Z}$ is a prime. When is $p\mathbb{Z}[i]$ a prime ideal?

Euclidean function
(euklidsk funksjon)

(2.16) Into which of the fields $\mathbb{F}_3$, $\mathbb{F}_5$ and $\mathbb{F}_7$ is there a map of rings from $\mathbb{Z}[i]$? If there is one, describe the kernel.

(2.17) Referring to Example 2.18 show that the three other involved numbers 3, $1+i\sqrt{5}$ and $1-i\sqrt{5}$ are irreducible.

✲ (2.18) *Euclidian functions and* PID*'s.* Let W be a well-ordered set (for instance $\mathbb{N}_0$). A *Euclidean function* with values in W on a ring A is mapping $\delta\colon A \to W$ such that for any pair a and b of elements from A there are elements q and r in A with $a = bq + r$ and $\delta(r) < \delta(b)$. Show that a domain A which possesses a Euclidian function, is a PID.
HINT: Minimize δ over non-zero elements in ideals.

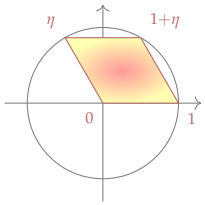
(2.19) Prove that $\mathbb{Z}[\sqrt{-2}]$ is a principal ideal domain by showing it has a division algorithm with the absolute value as a Euclidean function. HINT: The convex hull of $i\sqrt{2}$, 1 and 0 is a rectangle whose longest diagonal has length $\sqrt{3}$.

Gotthold Eisenstein
(1823–1852)
German mathematician

✲ (2.20) *The Eisenstein integers.* Gotthold Eisenstein is among the young geniuses who died early. He succumbed to tuberculosis in 1852 at the age of 29. The numbers in the ring $\mathbb{Z}[\eta]$ with η the cube root of unity $\eta = e^{2\pi i/3} = (-1+i\sqrt{3})/2$ are named after him.

a) Verify that $\mathbb{Z}[\eta] = \{\, n + m\eta \mid n, m \in \mathbb{Z} \,\}$ is a subring of $\mathbb{C}$;
HINT: $\eta^2 + \eta + 1 = 0$

b) Determine the units in $\mathbb{Z}[\eta]$;
c) Prove that $\mathbb{Z}[\eta]$ has a division algorithm with the absolute value as a Euclidean function. Conclude that $\mathbb{Z}[\eta]$ is a PID with every ideal generated by a shortest member. HINT: Compute the diameter of the convex set spanned by η, $1+\eta$, 1 and 0:
d) How many different generators does an ideal have?



2.5 *Existence theorems*

A useful technique for showing that ideals (and in later chapters submodules) of various kinds exist relies on the so-called Zorn's lemma. The lemma is a general result about existence of maximal elements in partially ordered sets, sets which for us mostly will be subsets of the lattices $\mathcal{I}(A)$ of ideals in a ring A ordered by inclusion, or later on subsets of the lattice of submodules of a module. The lemma turns out to be utterly useful when studying rings, and in the sequel it will be crucial at several occasions.

Max August Zorn (1906–1993)
German mathematician

Zorn's lemma

Zorn's lemma is one of quite a few theorems that for some reason keep being called lemmas. It is usually attributed to Max Zorn, but as often happens, its history can be traced further back; Felix Hausdorff published versions of it some ten years before Zorn. Anyhow, "Zorn's lemma" is a good name (so good that an experimental and non-narrative film made by Hollis Frampton in 1970 was called "Zorns lemma").

Maximal elements
(maksimale elementer)

(2.45) We begin with introducing some terminology. A *maximal element* x in the partially ordered set Σ is one for which there is no strictly larger element; that is, if $y \geqslant x$ then $y = x$. One should not confuse "maximal elements" with "largest elements" the latter being elements larger than all other elements in Σ. A partially ordered set can have several maximal elements whereas a largest element, if there is one, is unique. There is of course, analogous notions of *minimal elements* and *least* elements.

A partially ordered set is said to be *linearly ordered* or *totally ordered* if any two of its elements can be compared. Phrased differently, for any pair x, y of elements either $x \leqslant y$ or $y \leqslant x$ should hold. *A chain* in Σ is a linearly ordered subset of Σ. The chain is *bounded above* if for some element $x \in \Sigma$ it holds true that $y \leqslant x$ for all elements y in the chain, and then of course, x is called an *upper bound* for the chain. Similarly, the chain is said to be *bounded below* when having a *lower bound* in Σ; that is, an element $x \in \Sigma$ satisfying $x \leqslant y$ for all members y of the chain.

Linearly ordered sets (lineært ordnede mengder)
Totally ordered sets (totalt ordnede mengder)
Chains (kjeder)

We are now prepared to formulate Zorn's lemma, however we shall not prove it, only mention that it is equivalent to the axiom of choice (If you are interested in reading more about this, consult [[**?**]])

Bounded above (opptil begrenset)
Upper bound (øvre skranke)
Bounded below (nedtil begrenset)
Lower bound (nedre skranke)

THEOREM 2.46 (ZORN'S LEMMA) *Let Σ be a partially ordered set in which every chain is bounded above. Then Σ possesses a maximal element.*

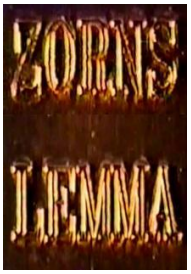
(2.47) A chain C in Σ is called *saturated* or *maximal* if it is not properly contained in any larger chain; that is, if C' is another chain with $C \subseteq C'$, then $C = C'$. A chain is saturated precisely when it impossible to insert any new element in-between two members of C. As an illustration of the mechanism of Zorn's lemma, let us prove the following

Saturated or maximal chains (mettede eller maksimale kjeder)

PROPOSITION 2.48 *Let C be a chain in the partially ordered set Σ. Then there is a saturated chain containing C.*

PROOF: The set of chains S in Σ containing C is partially ordered by inclusion, and we intend to apply Zorn's lemma to that set.

If $\mathcal{C}$ is a chain of chains (!!) the union $\bigcup_{C \in \mathcal{C}} C$ is anew a chain: indeed, suppose that x and y belong to the union so that there are chains C_x and C_y with $x \in C_x$ and $y \in C_y$. By assumption $\mathcal{C}$ is a chain, so either $C_x \subseteq C_y$ or $C_y \subseteq C_x$ holds. In either case x and y lie in a common chain and are therefor comparable. Every chain of chains is thus bounded above, and by Zorn's lemma, there is a maximal chain (*i. e.* saturated) chain in S. ❑



A fundamental existence result

A frequent application of Zorn's lemma in commutative algebra is to prove existence of ideals that are maximal subjected to a given condition, and in surprisingly many situations such maximizing ideals turn out to be prime .

Along these lines this section establishes a basic existence result with several important applications, one being that every ring has at least one maximal ideal. Our interest will be in ideals maximal among those containing a fixed ideal and being disjoint from a fixed set S. These maximizing ideals turn out to be prime when S is multiplicatively closed; that is, if the product of any two elements from S lie in S.

THEOREM 2.49 (THE FUNDAMENTAL EXISTENCE THEOREM FOR IDEALS) *A ring A, an ideal $\mathfrak{a}$ in A and a subset S not meeting $\mathfrak{a}$ are given. Then there exists an ideal $\mathfrak{b}$ maximal subjected to the two following conditions*

i) $S \cap \mathfrak{b} = \varnothing$;

ii) $\mathfrak{a} \subseteq \mathfrak{b}$.

If S is multiplicatively closed, the ideal $\mathfrak{b}$ will be a prime ideal.

PROOF: Consider the set Σ of ideals in A satisfying the two requirements in the theorem. It is non-empty because $\mathfrak{a}$ is supposed not to meet S and is a member of Σ. Obviously, the union of the ideals belonging to a chain in Σ will lie in Σ, and thus will be an upper bound for the chain. Zorn's lemma applies, and we may conclude that there is a maximal element in Σ.

Assume then that the set S is closed under multiplication, and let a and b be elements in A such that $ab \in \mathfrak{b}$. If neither belongs to $\mathfrak{b}$, the ideals $\mathfrak{b} + (a)$ and $\mathfrak{b} + (b)$ both meet S, being strictly larger than $\mathfrak{b}$. Hence we can find elements $x + \alpha a$ and $y + \beta b$ in S with $x, y \in \mathfrak{b}$ and $\alpha, \beta \in A$, and multiplying out, we find

$$(x + \alpha a)(y + \beta b) = xy + \alpha a y + \beta b x + \alpha\beta ab.$$

The left side belongs to S as S is supposed to be multiplicatively closed, and since x, y and ab all lie in $\mathfrak{b}$, the right side belongs to $\mathfrak{b}$, which contradicts the fact that $S \cap \mathfrak{b} = \varnothing$. ❑

THEOREM 2.50 (EXISTENCE OF MAXIMAL IDEALS) *Let A be a ring different from the null-ring. Every proper ideal $\mathfrak{a}$ in a ring A is contained in a maximal ideal. In particular, there is at least one maximal ideal in every ring, except the null-ring.*

PROOF: We apply the proposition with S merely consisting of the unit element, that is $S = \{1\}$. The maximizing ideal is proper and not contained in any other proper ideal. Hence it is maximal. To prove the second statement, apply the first to the zero ideal. ❑

The radical of an ideal

Prime factors frequently occur with higher multiplicities in a factorization of an integer, and it is of course interesting to get hold of the primes involved. In the transcription of Kummer and Dedekind into the language of ideals, this leads to the notion of the *radical* of a given ideal.

The radical of an ideal (radikalet til et ideal)

(2.51) The *radical* $\sqrt{\mathfrak{a}}$ of a given ideal $\mathfrak{a}$ in A consists of the elements a power of which lies in $\mathfrak{a}$; that is,

$$\sqrt{\mathfrak{a}} = \{\, a \in A \mid a^n \in \mathfrak{a} \text{ for some } \ n \in \mathbb{N} \,\}.$$

The nil radical (nilradikalet)

The elements of $\sqrt{\mathfrak{a}}$ are also characterized as the elements in A whose residue classes in $A/\mathfrak{a}$ are nilpotent. Along the same line, taking $\mathfrak{a}$ to be the zero ideal, we see that $\sqrt{(0)}$ is the set of nilpotent elements in A; it is called the *nil radical* of A.

(2.52) The first thing to establish is that the radical $\sqrt{\mathfrak{a}}$ in fact is an ideal.

LEMMA 2.53 *Let $\mathfrak{a}$ be an ideal in the ring A. Then the radical $\sqrt{\mathfrak{a}}$ is an ideal.*

PROOF: The radical is obviously closed under multiplication by ring elements, and we merely have to check it is closed under addition. So assume that a and b are two elements in the ring such that $a^n \in \mathfrak{a}$ and $b^m \in \mathfrak{a}$. The binomial theorem gives

$$(a+b)^N = \sum_{0 \leqslant i \leqslant N} \binom{N}{i} a^{N-i} b^i.$$

Choosing $N = n + m - 1$, we see that when $i < m$ it holds that $N - i \geqslant n$, so either a^{N-i} or b^i lies in $\mathfrak{a}$. Every term of the sum therefore lies in $\mathfrak{a}$, and by that the sum itself. ❑

Specializing $\mathfrak{a}$ to be the zero ideal yields the following.

COROLLARY 2.54 *The set of nilpotent elements in A form an ideal.*

(2.55) An ideal $\mathfrak{a}$ in A is said to be *radical* if it equals its own radical; *i. e.* if it holds true that $\sqrt{\mathfrak{a}} = \mathfrak{a}$. One easily verifies that the radical of an ideal is a radical ideal so that the equality $\sqrt{(\sqrt{\mathfrak{a}})} = \sqrt{\mathfrak{a}}$ holds true. In a similar manner as prime ideals and maximal ideals radical ideals may be characterized in terms of quotients:

Radical ideals (radikale idealer)

PROPOSITION 2.56 *An ideal $\mathfrak{a}$ in the ring A is radical if and only if the quotient $A/\mathfrak{a}$ is reduced.*

PROOF: The residue class $[a]$ in $A/\mathfrak{a}$ of an element a is nilpotent precisely when a power a^n lies in $\mathfrak{a}$, so that A is reduced precisely when $\sqrt{\mathfrak{a}} = \mathfrak{a}$. ❑

(2.57) The radical $\sqrt{\mathfrak{a}}$ must be contained in any prime ideal containing $\mathfrak{a}$ because if $a^n \in \mathfrak{a}$ and $\mathfrak{a} \subseteq \mathfrak{p}$ with $\mathfrak{p}$ prime, it holds that $a \in \mathfrak{p}$, so $\mathfrak{a}$ lies within the intersection of the prime ideals containing it. The converse inclusion also holds and hinges on the basic existence result above.

PROPOSITION 2.58 (THE RADICAL AS INTERSECTION OF PRIMES) *Let A be a ring and assume that $\mathfrak{a}$ is a proper ideal in A. The radical $\sqrt{\mathfrak{a}}$ equals the intersection of the prime ideals containing $\mathfrak{a}$; that is,*

$$\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{a} \subseteq \mathfrak{p},\ \mathfrak{p} \text{ prime}} \mathfrak{p}.$$

PROOF: We already observed that $\sqrt{\mathfrak{a}} \subseteq \bigcap_{\mathfrak{a} \subseteq \mathfrak{p}} \mathfrak{p}$, so assume that a is an element not lying in the radical $\sqrt{\mathfrak{a}}$. We shall apply The Basic Existence Theorem (Theorem 2.49 on page 49) with S being the set $\{ a^n \mid n \in \mathbb{N} \}$ of powers of a (which obviously is closed under multiplication). Since $a \notin \sqrt{\mathfrak{a}}$, it holds true that $S \cap \mathfrak{a} = \varnothing$, and by the theorem we conclude that there is prime ideal $\mathfrak{a}$ containing $\mathfrak{a}$ disjoint from S; that is, $a \notin \mathfrak{a}$. ❑

(2.59) The special case that $\mathfrak{a} = (0)$ merits to be pointed out:

COROLLARY 2.60 *The set of nilpotent elements in A equals the intersection of all prime ideals in A; that is,*

$$\sqrt{(0)} = \bigcap_{\mathfrak{p} \text{ prime}} \mathfrak{p}.$$

Of course the larger of two ideals, one containing the other, is not needed in an intersection, and one might be tempted to discard from the intersection in Proposition 2.58 the prime ideals not being minimal among those containing $\mathfrak{a}$ and thus write

$$\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{p} \text{ minimal}} \mathfrak{p}, \tag{2.4}$$

where the intersection extends over all prime ideals minimal over $\mathfrak{a}$. Such a representation is certainly valid, but the argument is more complicated than indicated since *a priori* there could be infinitely descending chains of distinct prime ideals. However, if $\{\mathfrak{p}_i\}_{i\in I}$ is a chain of prime ideals, the intersection $\bigcap_{i\in I} \mathfrak{p}_i$ is a prime ideal (you are asked to check this in Exercise 2.21 on the next page), and so by Zorn's lemma, every prime containing $\mathfrak{a}$ contains a prime ideal minimal among those containing $\mathfrak{a}$; and this is exactly what we need to have a representation as in (2.4).

(2.61) The operation of forming the radical commutes with forming finite intersections; one has:

LEMMA 2.62 *For every finite collection $\{\mathfrak{a}_i\}$ of ideals in A the equality $\bigcap_i \sqrt{\mathfrak{a}_i} = \sqrt{\bigcap_i \mathfrak{a}_i}$ holds true.*

PROOF: When an element a from A belongs to each of the radicals $\sqrt{\mathfrak{a}_i}$, there are integers n_i so that $a^{n_i} \in \mathfrak{a}_i$. With $n = \max n_i$ (here we use that the n_i's are finite in number), it then holds true that $a^n \in \mathfrak{a}_i$ for each i, and thus $a \in \sqrt{\bigcap_i \mathfrak{a}_i}$. This shows that one has the inclusion $\bigcap_i \sqrt{\mathfrak{a}_i} \subseteq \sqrt{\bigcap_i \mathfrak{a}_i}$. The other inclusion is straightforward. ❑

Examples

(2.20) Even if a power of every element in $\sqrt{\mathfrak{a}}$ lies in $\mathfrak{a}$, no power of $\sqrt{\mathfrak{a}}$ will in general be contained in $\mathfrak{a}$; a simple, but typical example, being the ideal $\mathfrak{a} = (x_1, x_2^2, x_3^3, \ldots)$ generated by the powers x_i^i in the polynomial ring $k[x_1, x_2, x_3, \ldots]$ in countably many variables. The radical of $\mathfrak{a}$ equals the maximal ideal $\mathfrak{m} = (x_i | i \in \mathbb{N}_0)$ generated by the variables, but no power of $\mathfrak{m}$ is contained in $\mathfrak{a}$. Indeed, the exponent needed to force a power of x_i to lie in $\mathfrak{a}$, tends to infinity with i.

(2.21) The operation of forming radicals does not necessarily respect infinite intersections. For instance, if p is a prime number, one has $\sqrt{p^r\mathbb{Z}} = p\mathbb{Z}$ and therefore $\bigcap_r \sqrt{p^r\mathbb{Z}} = p\mathbb{Z}$. But evidently it holds true that $\bigcap_r p^r\mathbb{Z} = 0$.

☆

Exercises

⁂ (2.21) *Unions and intersections of chains of primes.* Let $\{\mathfrak{p}_i\}_{i\in I}$ be a chain of prime ideals in the ring A. Show that both the union $\bigcup_{i\in I}\mathfrak{p}_i$ and the intersection $\bigcap_{i\in I}\mathfrak{p}_i$ are prime ideals. Show that every prime ideal containing a given ideal $\mathfrak{a}$ contains a prime ideal minimal over $\mathfrak{a}$. Show that any prime ideal contained in $\mathfrak{a}$, is contained in a prime maximal among those contained in $\mathfrak{a}$.

⁂ (2.22) Assume that $\mathfrak{p}\subset\mathfrak{q}$ are two distinct prime ideals in the ring A. Show that there are prime ideals $\mathfrak{p}'$ and $\mathfrak{q}'$ with $\mathfrak{p}\subseteq\mathfrak{p}'\subset\mathfrak{q}'\subseteq\mathfrak{q}$ and so that there is no prime ideal lying strictly between $\mathfrak{p}'$ and $\mathfrak{q}'$.

⁂ (2.23) *Saturated multiplicative sets and zero divisors.* A multiplicatively closed set S in the ring A is said to be *saturated* if with x it contains every factor of x; that is, if $x\in S$ and $x=yz$, then $y\in S$ (and by symmetry $z\in S$).

Saturated multiplicative sets (mettede multiplikative mengder)

a) Show that S is a saturated multiplicative set if and only if the complement $A\backslash S$ is a union of prime ideals.

b) Show that the set of non-zero divisors in A form a saturated multiplicative set.

c) Conclude that the set of zero divisors in A is a union of prime ideals.

(2.24) Show that the group of units A^* is contained in any saturated multiplicative set.

⁂ (2.25) The prime ideals that appear as maximizing in the proof of Proposition 2.58 are of special kind. Let $S=\{a^n\}$ be the set of powers of an element from the ring A, and let $\mathfrak{p}$ be maximal among the ideals not meeting S. Show that the class $[a]$ in $A/\mathfrak{p}$ is contained in every non-zero prime ideal of $A/\mathfrak{p}$.

(2.26) Let $\mathfrak{a}$ be a finitely generated ideal. Show that a sufficiently high power of $\mathfrak{a}$ is contained in the radical $\sqrt{\mathfrak{a}}$.

⁂ (2.27) Let A be a PID.

a) Show that every ascending chain of ideals in A is eventually constant.

b) Show that up to association, there are only finitely many irreducible elements dividing a given a.

★

2.6 *Local rings*

Rings having only a single maximal ideal are called *local rings*. They occupy a central place in the theory being of a simpler kind of rings, and a frequently applied strategy of proof is to reduce an issue to a statement about local rings. In the analogy with rings of functions, the local rings correspond to rings of germs* of functions near a point—hence the name—and the maximal ideal consists of the germs vanishing at the point. There is also the notion of a *semi-local ring*, which is a ring with merely finitely many maximal ideals.

Local rings (lokale ringer)

**Recall that a germ of functions at a point is a class of function coinciding on neighbourhoods of the point*

Semi-local rings (semilokale ringer)

(2.63) In a local ring A the complement of the maximal ideal $\mathfrak{m}$ coincides with the group of units; that is, every element $a \in A$ not lying in $\mathfrak{m}$ is invertible. Indeed, if it were not, the principal ideal (a) would be a proper ideal, and by existence of maximal ideals (Theorem 2.50 on page 50) it would be contained in a maximal ideal, obviously different from $\mathfrak{m}$, which is incompatible with $\mathfrak{m}$ being the sole maximal ideal in A. This proves that the first statement in the following proposition implies the second.

PROPOSITION 2.64 *Let A be a ring and $\mathfrak{m}$ a proper ideal in A. The following three statements are equivalent.*

i) A is a local ring with maximal ideal $\mathfrak{m}$;
ii) The group of units and the complement of $\mathfrak{m}$ coincide; that is, $A^* = A \setminus \mathfrak{m}$;
iii) The ideal $\mathfrak{m}$ is maximal and consists of elements a such that $1 + a$ is invertible.

PROOF: To see that the last statement ensues from the second, observe that if a is a member of $\mathfrak{m}$, then $1 + a$ is not in $\mathfrak{m}$ and hence is invertible.

Finally, assume that $\mathfrak{m}$ be maximal and that all elements be shaped like $1 + a$ with $a \in \mathfrak{m}$ are invertible. Let x be an element not in $\mathfrak{m}$. Since $\mathfrak{m}$ is maximal, it holds true that $\mathfrak{m} + (x) = A$; hence $x = 1 + a$ for some $a \in \mathfrak{m}$, and x is invertible. This shows that *iii)* implies *i)*. ❑

The assumption in the last statement that $\mathfrak{m}$ be maximal, is necessary; for an example see Exercise 2.30 below.

Jacobson radical (Jacobson-radikalet)

(2.65) The argument in the previous paragraph partially goes through in a slightly more general staging involving the so-called *Jacobson radical* $J(A)$ of a ring A—the intersection of all the maximal ideals in A—that is, $J(A) = \bigcap_{\mathfrak{m} \subseteq A \text{ maximal}} \mathfrak{m}$.

PROPOSITION 2.66 *Let A be a ring. The Jacobson radical of A consists of the ring elements a so that $1 + xa$ is invertible for all $x \in A$.*

PROOF: Fix an element a in A. Firstly, assume that all elements of shape $1 + xa$ are invertible. If there is a maximal ideal $\mathfrak{m}$ so that $a \notin \mathfrak{m}$, it holds true that $\mathfrak{m} + (a) = A$, and there is a relation $1 = y + ax$ with $y \in \mathfrak{m}$. It ensues that $1 - xa$ lies in $\mathfrak{m}$, but on the other hand, $1 - ax$ is invertible by assumption; and we have the contradiction* that $\mathfrak{m} = A$.

*Remember, maximal ideals are proper ideals

Assume then that a lies in all maximal ideals. If $(1 + ax)$ were a proper ideal, it would by the Fundamental Existence Theorem (Theorem 2.49 on page 49) be contained in a maximal ideal $\mathfrak{n}$. Since $a \in \mathfrak{n}$, it would follow that $1 \in \mathfrak{n}$, contradicting that $\mathfrak{n}$ is proper. Hence the principal ideal $(1 + ax)$ is not proper, and $1 + ax$ is invertible. ❑

The category of local rings

Local homomorphisms (lokal homomorfier)

(2.67) Assume that A and B are two local rings whose maximal ideals are $\mathfrak{m}_A$ and $\mathfrak{m}_B$ respectively. A map of rings $\phi\colon A \to B$ is said to be a *local homomorphism*, or a

map of local rings, if it holds true that $\phi(\mathfrak{m}_A) \subseteq \mathfrak{m}_B$. Equivalently, one may request that $\phi^{-1}(\mathfrak{m}_B) = \mathfrak{m}_A$ (the inclusion $\phi^{-1}(\mathfrak{m}_B) \subseteq \mathfrak{m}_A$ always holds since $\phi^{-1}(\mathfrak{m}_B)$ is a proper ideal). The field $k = A/\mathfrak{m}_A$ is called *restklasse kroppen* of A, often abbreviated to the *residue field* of A. Together with these homomorphisms the local rings form a category LocRings. Ring maps between local rings that are not local abound, a stupid example being the inclusion of a local domain in a field; *e.g.* the inclusion of the ring $\mathbb{Z}_{(p)}$ in $\mathbb{Q}$ (cfr. Example 2.23 below)

The residue class field
(the residue class field)

Examples

(2.22) The set of rational functions over a field k that may be expressed as $P(x)/Q(x)$ with $P(x)$ and $Q(x)$ polynomials and $Q(0) \neq 0$, is a local ring whose maximal ideal equals the set of the functions vanishing at the origin. The evaluation map given by $P(x)/Q(x) \mapsto P(0)/Q(0)$ identifies the residue field with the ground field k.

(2.23) Let p be a prime number and let $\mathbb{Z}_{(p)}$ be the ring of rational numbers expressible as n/m where the denominator m is relatively prime to p. Then $\mathbb{Z}_{(p)}$ is a local ring whose maximal ideal is generated by p. Even more is true, the only ideals in $\mathbb{Z}_{(p)}$ are the principal ideals (p^v); indeed, every rational number lying in $\mathbb{Z}_{(p)}$ may be written as $p^v n/m$ with $v \geqslant 0$ and neither n nor m having p as factor; so if $\mathfrak{a}$ is an ideal, $\mathfrak{a} = (p^v)$ with v the least power of p dividing an element from $\mathfrak{a}$. And among these ideals (p) contains all the others. The residue class field of $\mathbb{Z}_p$ is the finite field $\mathbb{F}_p$ with p elements.

(2.24) In a polynomial ring $\mathbb{C}[x_1, \ldots, x_r]$ for all points $a \in \mathbb{C}^r$ the ideal of polynomials vanishing at a is a maximal ideal. It follows that the Jacobson radical of $\mathbb{C}[x_1, \ldots, x_r]$ equals (0).

(2.25) Assume that p and q are two prime numbers. Let A be the ring of rational numbers with denominator relatively prime to pq. That is $A = \{\, n/m \mid n, m \in \mathbb{Z}, (m, pq) = 1 \,\}$. The principal ideals by (p) and (q) are the only two maximal ideals in A, and $J(A) = (p) \cap (q) = (pq)$.

☆

Exercises

⁂ (2.28) Show that a ring has just one prime ideal if and only if its elements are either invertible or nilpotent. Prove that this is the case if and only if $A/\sqrt{(0)}$ is a field.

(2.29) Let k be a field. Show that power series ring $k[\![t]\!]$ is a local ring with maximal ideal $(t)k[\![t]\!]$.

(2.30) Let A be the subring of $\mathbb{Q}$ whose elements are the rational numbers a expressible as $a = m/n$ where n does have neither 2 nor 3 as factor. Show that A has two maximal ideals (2) and (3) whose intersection equals (6). What are the two residue fields? Show that $1 + a$ is invertible in A for all members $a \in (6)$.

⁂ (2.31) Let $p_1, \ldots, p_r$ be distinct prime numbers and let A be the subset of $\mathbb{Q}$ whose

members can be written as m/n with n relatively prime to p_i for $1 \leqslant i \leqslant r$. Show that A is a semi-local ring. Describe the maximal ideals and the residue fields. What is the Jacobson radical?

(2.32) Let $k_1, \ldots, k_r$ be fields. Show that the product ring $\prod_i k_i$ is a semi-local ring. What are the maximal ideals?

(2.33) Let $f(x)$ be any polynomial in $k[x]$ where k is a field. Show that $k[x]/(f(x))$ is semi-local.

⁂ (2.34) Let A be a principal ideal domain with infinitely many maximal ideals. Show that $J(A) = (0)$.

⁂ (2.35) Show that the polynomial ring $\mathbb{C}[x_1, \ldots, x_n]$ has a vanishing Jacobson radical. ★

2.7 *Direct products and the Chinese Remainder Theorem*

Ideals in a direct product

Let $A = \prod_{1 \leqslant i \leqslant r} A_i$ be a direct product of rings A_i. There is a simple description of all the ideals in A in terms of ideals in the A_i's. One produces an ideal $\mathfrak{a}$ in A from a sequence of ideals $\mathfrak{a}_i$ in the A_i's simply by putting $\mathfrak{a} = \prod_i \mathfrak{a}_i$. And, indeed, all ideals in A are of this shape. To see this, let $\{e_i\}_{1 \leqslant i \leqslant r}$ be the orthogonal idempotents corresponding to the decomposition of A as a direct product. Then $A_i = e_i A$ and each $e_i \mathfrak{a}$ is an ideal in A_i contained in $\mathfrak{a}$, and because $\sum_i e_i = 1$, it holds true that $\mathfrak{a} = \sum_i e_i \mathfrak{a}$.

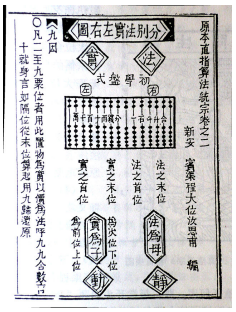
PROPOSITION 2.68 *The ideals of $A = \prod_{1 \leqslant i \leqslant r} A_i$ are all of the form $\prod_{1 \leqslant i \leqslant r} \mathfrak{a}_i$ where each $\mathfrak{a}_i$ is an ideal in A_i. It holds true that $A/\mathfrak{a} \simeq \prod_{1 \leqslant i \leqslant r} A_i/\mathfrak{a}_i$. The ideal $\mathfrak{p}$ is a prime ideal if and only if $\mathfrak{p}_i = A_i$ for all but one index i_0 and $\mathfrak{p}_{i_0}$ is a prime ideal.*

PROOF: The first claim is already dealt with. To the second: the projections $A \to A_i$, coinciding with multiplication by e_i, send an ideal $\mathfrak{a}$ to $\mathfrak{a}_i = e_i \mathfrak{a}$, and hence they induce maps $A/\mathfrak{a} \to A_i/\mathfrak{a}_i$. In their turn these give rise to a ring-map $A/\mathfrak{a} \to \prod_i A_i/\mathfrak{a}_i$, which is surjective: $\sum e_i a_i$ maps to $(e_i a_i)$ since the idempotents are orthogonal. If x is an element in A such that $e_i x \in \mathfrak{a}_i$ for each i, the element x belongs to $\mathfrak{a}$ since $x = \sum_i e_i x$ and $\mathfrak{a} = \sum_i \mathfrak{a}_i$, and the map is injective.

What remains to be verified is the statement about the prime ideals: it follows since the principal idempotents in $\prod_i A_i/\mathfrak{a}_i$ are orthogonal, and so when at least two of them are non-trivial, the product $\prod_i A_i/\mathfrak{a}_i$ is not an integral domain. ❑

(2.69) It is appropriate to give a comment about the zero ring at this stage. In Proposition 1.27 the idempotents e_i's are not required to be different from zero, but if $e_i = 0$, of course $e_i A$ is the zero ring, and does not contribute in a significant way to the product (it holds true that $0 \times A \simeq A$). This is particularly pertinent for the formulation of Proposition 2.68; it might happen that $\mathfrak{a}_i = A_i$ so that $A/\mathfrak{a}_i$ is the zero ring.

EXAMPLE 2.26 The description of the ideals in Proposition 2.68 is not valid for infinite products. For instance, the ideals in an infinite product $\prod_{i\in I} k_i$ of fields are described by so-called *filters* and *ultrafilters* on the index set I. One simple example not among those described, is the set $\mathfrak{a}$ of strings (a_i) with $a_i \neq 0$ only for finitely many i (which by the way equals the direct sum of the k_i's). This is easily checked to be an ideal; it is certainly not prime, but is contained in at least one maximal ideal as every proper ideal is. Ideals containing $\mathfrak{a}$ can not be of the form described in the theorem since $\mathfrak{a}$ contains each k_i. ☆



Some old Chinese mathematics.

The Chinese Remainder Theorem

A classical result which at least goes back to third century AD, is the so called Chinese Remainder Theorem. It seems that the first written account of this result appears in the book *Sunzi Suanjing* by a Chinese mathematician "Master Sun"— hence the Chinese theorem. A more informative name would be the Theorem of Simultaneous Congruences: As long as two integers, n_1 and n_2, called the *moduli*, are relatively prime, two congruences $x \equiv y_1 \mod n_1$ and $x \equiv y_2 \mod n_2$ have a common solution.

(2.70) This can be generalized to any number of congruences as long as the moduli are pairwise relatively prime, and there is a formulation for general rings with the moduli replaced by ideals. The appropriate condition on the ideals that replaces the moduli being relatively prime, is as follows: two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are said to be *comaximal* if $\mathfrak{a} + \mathfrak{b} = A$, equivalently, if one can write $1 = a + b$ with $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$.

Comaximal ideals (komaksimale idealer)

(2.71) Given a finite collection $\{\mathfrak{a}_i\}_{1\leqslant i\leqslant r}$ of ideals in the ring A. There is an obvious map

$$A \to \prod_i A/\mathfrak{a}_i$$

sending a ring-element a to the tuple whose i-th component is the residue class of a modulo $\mathfrak{a}_i$. Its kernel consists of the elements in A lying in all the $\mathfrak{a}_i$'s, and hence there is induced an injective map

$$\psi\colon A/\mathfrak{a}_1 \cap \ldots \cap \mathfrak{a}_r \hookrightarrow \prod_i A/\mathfrak{a}_i.$$

The Chinese statement is that, under certain circumstances, this map is an isomorphism.

THEOREM 2.72 (THE CHINESE REMAINDER THEOREM) *Let A be a ring and assume we are given a finite collection of pairwise comaximal ideals $\{\mathfrak{a}_i\}_{1\leqslant i\leqslant r}$. Then the canonical reduction maps induce an isomorphism $A/\mathfrak{a}_1 \cap \ldots \cap \mathfrak{a}_r \simeq \prod_{1\leqslant i\leqslant r} A/\mathfrak{a}_i$.*

PROOF: It suffices to find elements a_i in A which are congruent one modulo $\mathfrak{a}_i$ and congruent zero modulo all the other ideals in the collection. Indeed, the sum $\sum_i y_i a_i$,

with the y_i's being arbitrary ring-elements, will then have the same residue class as y_i modulo $\mathfrak{a}_i$.

For each pair of indices i and j with $i \neq j$ we may write $1 = c_{ij} + c_{ji}$ with $c_{ij} \in \mathfrak{a}_j$ and $c_{ji} \in \mathfrak{a}_i$. Then c_{ij} is congruent one modulo $\mathfrak{a}_i$ and congruent zero modulo $\mathfrak{a}_j$. Hence the product $a_i = \prod_{j \neq i} c_{ij}$ is congruent one modulo $\mathfrak{a}_i$ and congruent zero modulo $\mathfrak{a}_j$ for $j \neq i$; and we are done. ❑

Exercises

(2.36) Let $\mathfrak{a}$ and $\mathfrak{b}$ be two comaximal ideals such that $\mathfrak{a} \cap \mathfrak{b} = 0$. If $a + b = 1$ with $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$, show that a and b are idempotents.

(2.37) Show that two ideals whose radicals are comaximal, are comaximal.

(2.38) Show that $28y - 27z$ solves the simultaneous congruences $x \equiv y \mod 9$ and $x \equiv z \mod 4$.

(2.39) Let A be a semi-local ring. Show that $A/J(A)$ is a product of fields.

✲ (2.40) Assume that $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ are pairwise comaximal ideals in the ring A.

a) Show that $\mathfrak{a}_1$ and $\mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r$ are comaximal;
b) Show that one has $\mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r = \mathfrak{a}_1 \cap \ldots \cap \mathfrak{a}_r$;
c) For each i with $1 \leqslant i \leqslant r$ let $\mathfrak{b}_i = \prod_{j \neq i} \mathfrak{a}_j$. Prove that the $\mathfrak{b}_1, \ldots, \mathfrak{b}_r$ are comaximal ideals; *i. e.* that $\mathfrak{b}_1 + \ldots + \mathfrak{b}_r = A$.

✲ (2.41) Determine integers representing the idempotents in $\mathbb{Z}/30\mathbb{Z}$ and $\mathbb{Z}/105\mathbb{Z}$.

(2.42) Prove that a reduced ring decomposes as $A \simeq A_1 \times \ldots \times A_r$ where each A_i is an algebra over a finite field.

✲ (2.43) *Locally nilpotent ideals.* One says that an ideal $\mathfrak{n}$ in a ring A is *locally nilpotent* if each element in A is nilpotent. Show that for each ideal $\mathfrak{a}$ in A it holds that $\sqrt{\mathfrak{a} + \mathfrak{n}} = \sqrt{\mathfrak{a}}$ whenever $\mathfrak{n}$ is locally nilpotent. Let $A \to B$ be a surjection of commutative rings whose kernel is locally nilpotent. Show that the map Spec $B \to$ Spec A is a homeomorphism.

(2.44) *Lifting of idempotents.* Let $A \to B$ be a surjective map of (not necessarily commutative*) rings whose kernel $\mathfrak{a}$ is locally nilpotent; that is, every element of $\mathfrak{a}$ is nilpotent. Let e be an idempotent in B. The aim of the exercise is to show that there is an idempotent ϵ in A mapping to e. Choose any element x in A that maps to e and let $y = 1 - x$.

*For once, we work with non-commutative rings; it will useful to lift idempotent endomorphisms of modules over commutative rings

a) Show that $xy \in \mathfrak{a}$.
b) Let n be such that $(xy)^n = 0$ and define the two elements $\epsilon = \sum_{i>n} \binom{2n}{i} x^i y^{2n-i}$ and $\gamma = \sum_{i \leqslant n} \binom{2n}{i} x^i y^{2n-i}$. Show that $1 = \epsilon + \gamma$ and that $\epsilon\gamma = 0$.
c) Conclude that ϵ is an idempotent in A that maps to e.

★

2.8 Graded rings and homogenous ideals

Recall that any polynomial can be written as the sum of its homogenous components. Several techniques, useful when working with polynomials, involve such a decomposition; just to mention two powerful tools: induction on the degree of the lowest or the highest term. A class of rings sharing some of these nice properties polynomials have, are the so-called *graded rings* whose elements possess a decomposition mimicking the one of polynomials into homogeneous terms.

Even more forceful techniques are available to handle graded rings which satisfy appropriate finiteness conditions. For instance, when all the homogenous components R_ν are finite dimensional vector spaces over some field $k \subseteq R_0$, the so called *Hilbert function* $h_R(\nu) = \dim_k R_\nu$ is a very strong invariant of R.

A the present stage of the course we merely scratch the surface of the theory of graded rings, but they will reappear later at several occasions.

Graded rings (graderte ringer)

(2.73) A *graded ring* R is a ring together with a decomposition of the underlying abelian group as a direct sum

$$R = \bigoplus_{\nu \in \mathbb{Z}} R_\nu \tag{2.5}$$

of additive subgroups R_ν subjected to the rule that $R_\nu \cdot R_\mu \subseteq R_{\nu+\mu}$ for any pair of indices ν, μ.

Homogenous elements (homogene elementer)

(2.74) Elements from the subgroup R_ν are said to be *homogenous of degree* ν. Notice that the zero element 0 lies in every one of the subgroups R_ν, and one can not attribute a well-defined degree to it, but it will rather be considered to be homogeneous of any degree. From a decomposition as in (2.5) ensues that each non-zero element a in R can be expressed as a sum $a = \sum_\nu a_\nu$ whose terms a_v are homogenous of degree ν merely finitely many of which are different from zero. The a_ν's are uniquely determined by a and go under the name of the *homogenous components* of a.

Homogenous components (homogene komponenter)

Notice that $R_0 \cdot R_0 \subseteq R_0$, so R_0 is a subring of R. Similarly, for every ν it holds true that $R_0 \cdot R_\nu \subseteq R_\nu$, and the ring R_0 of elements homogeneous of degree zero acts on the group of those homogeneous of degree ν. In particular, if $k \subseteq R_0$ is a field, the additive subgroups R_ν will all be vector spaces over k.

Homogenous ideals (homogene idealer)

(2.75) If $\mathfrak{a}$ is an ideal in the graded ring R, we denote by $\mathfrak{a}_\nu$ the subgroup $\mathfrak{a}_\nu = \mathfrak{a} \cap R_\nu$ consisting of the homogenous elements of degree ν that lie in $\mathfrak{a}$. One says that $\mathfrak{a}$ is a *homogenous ideal* whenever $\mathfrak{a} = \sum_\nu \mathfrak{a}_\nu$; in other words: if a belongs to $\mathfrak{a}$ then all the homogeneous components of a belong to $\mathfrak{a}$ as well. Since homogenous components are unambiguously defined (or if you prefer, because $\mathfrak{a}_\nu \cap \mathfrak{a}_\mu = (0)$ whenever $\nu \neq \mu$), the sum is a direct sum, and we are entitled to write $\mathfrak{a} = \bigoplus_\nu \mathfrak{a}_\nu$.

PROPOSITION 2.76 *Let $\mathfrak{a}$ be an ideal in the graded ring R. The following three statements are equivalent.*

i) The ideal $\mathfrak{a}$ is homogenous;
ii) All homogenous components of elements in $\mathfrak{a}$ belong to $\mathfrak{a}$;
iii) The ideal $\mathfrak{a}$ may be generated by homogenous elements.

PROOF: That the first two statements are equivalent, is just a rephrasing of how homogenous ideals were defined. Let us then prove that *i)* and *iii)* are equivalent; so assume that $\mathfrak{a}$ is a homogenous ideal. The homogeneous components of all members of any set of generators of $\mathfrak{a}$ then belong to $\mathfrak{a}$, and obviously they generate $\mathfrak{a}$ (the original generators are sums of them).

The other implication is also straightforward. Let $\{a_i\}_{i\in I}$ be a set of homogeneous generators for $\mathfrak{a}$, and say a_i is of degree d_i. Any element $a \in \mathfrak{a}$ can then be expressed as $f = \sum_i f_i \cdot a_i$ with $f_i \in R$, and expanding the sum into a sum of homogenous term we find

$$f = \sum_i f_i \cdot a_i = \sum_i \big(\sum_\nu f_{i,\nu} \cdot a_i\big) = \sum_d \big(\sum_{\nu+d_i=d} f_{i,\nu} \cdot a_i\big),$$

where $f_{i,\nu}$ denotes the homogeneous component of f_i of degree ν, and where we in the last sum have recollected all terms $f_{i,\nu} \cdot a_i$ of the same degree d. Hence $\sum_{\nu+d_i=d} f_{i,\nu} \cdot a_i$ is the homogeneous component of f of degree d, and it belongs to $\mathfrak{a}$ since the a_i's lie there. ❑

(2.77) A rich source of graded rings are the quotients of polynomial rings by homogenous ideals, or more generally the quotient of any graded ring by a homogenous ideal.

PROPOSITION 2.78 *Let R be a graded ring and $\mathfrak{a} \subseteq R$ a homogeneous ideal. Then the quotient $R/\mathfrak{a}$ is a graded ring whose homogeneous components are given as $(R/\mathfrak{a})_\nu = R_\nu/\mathfrak{a}_\nu$.*

PROOF: This follows without great effort from the direct sum decompositions $R = \bigoplus_\nu R_\nu$ and $\mathfrak{a} = \bigoplus_\nu \mathfrak{a}_\nu$. Notice first that as $R_\nu \cap \mathfrak{a} = \mathfrak{a}_\nu$, there are natural inclusions $R_\nu/\mathfrak{a}_\nu \subseteq R/\mathfrak{a}$. Hence any class $[a] \in R/\mathfrak{a}$ with a decomposing as $a = \sum_i a_i$ in homogeneous terms decomposes as $[a] = \sum_i [a_i]$, where we at will can consider the $[a_i]$'s to be elements in $R/\mathfrak{a}$ or in $R_i/\mathfrak{a}_i$. Moreover, the classes $[a_i]$ are unique because if $\sum_i a_i$ and $\sum_i b_i$ were two such decompositions inducing the same element in $R/\mathfrak{a}$, it would hold true that $\sum_i (a_i - b_i) \in \mathfrak{a}$. The ideal $\mathfrak{a}$ being homogeneous and each term $a_i - b_i$ being homogeneous of degree i, it would follow that $a_i - b_i \in \mathfrak{a}$, and hence $[a_i] = [b_i]$. ❑

EXAMPLE 2.27 *A weighted grading:* There is a way of giving polynomial rings another grading than the traditional one, which sometimes turns out to be useful. We shall illustrate this in the of case two variables $R = k[x, y]$. The idea is to give each variables x and y a *weight*, that is putting $\deg x = \alpha$ and $\deg y = \beta$ where α and β can be any pair of integers. The degree of the monomial $x^i y^j$ is then defined as $\deg x^i y^j = i\alpha + j\beta$. This

defines a graded structure on the polynomial ring with

$$R_\nu = \bigoplus_{i\alpha+j\beta=\nu} k \cdot x^i y^j.$$

Since already R is the direct sum $R = \bigoplus_{i,j} k \cdot x^i y^j$, one arrives at the direct sum $R = \bigoplus_\nu R_\nu$ by just recollecting terms $k \cdot x^i y^j$ with the same degree; that is, the polynomials in R_ν satisfy $i\alpha + j\beta = \nu$. Some call these polynomials *isobaric*. ☆

EXAMPLE 2.28 A natural and useful condition on a graded ring is that $R_n = (0)$ for $n < 0$; that is, the degrees of any non-zero element is non-negative (it opens up for induction arguments). However, several graded rings occurring naturally are not like that. One example is the subring R of $k(x_1, \dots, x_n)$ consisting of rational functions shaped like f/g^ν where g is a fixed homogenous polynomial, and $f \in k[x_1, \dots, x_n]$ and $\nu \in \mathbb{N}_0$. Putting $\deg f/g^\nu = \deg f - \nu \cdot \deg g$ makes R a graded ring (check that!), and then $\deg 1/g^\nu = -\nu \cdot \deg g$. ☆

Exercises

(2.45) Generalize Example 2.27 above to polynomials in any number of variables by giving each variable x_i a weight α_i.

✲ (2.46) With reference to the Example 2.27 above , show that the subring R_0 of elements of degree zero in the case $\alpha = 1, \beta = -1$ is isomorphic to the polynomial ring over k in one variable. Describe R_i for all i.

✲ (2.47) *Homogeneous prime ideals.* Let $\mathfrak{p}$ be a homogenous ideal in a graded ring R. Show that $\mathfrak{p}$ is a prime ideal if and only of $x \cdot y \in \mathfrak{p}$ implies that either $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$ for *homogenous* elements x and y.

(2.48) *Monomial ideals.* An ideal $\mathfrak{a}$ in the polynomial ring $k[x_1, \dots, x_r]$ is said to be a *monomial* if it holds true that a polynomial f belongs to $\mathfrak{a}$ if and only if every monomial occurring in f lies there. Show that this is equivalent to $\mathfrak{a}$ being generated by monomials.

Monomial ideals (monomiale idealer)

✲ (2.49) Assume that k is an infinite field. The multiplicative group k^* acts on the polynomial ring $k[x_1, \dots, x_r]$ in a natural way; the result of the action of $\alpha \in k^*$ on the polynomial $f(x_1, \dots, x_r)$ being $f^\alpha(x_1, \dots, x_r) = f(\alpha x_1, \dots, \alpha x_r)$.

a) Show that the polynomial f is homogeneous of degree d if and only if $f^\alpha = \alpha^d \cdot f$ for all α.

b) Show that an ideal $\mathfrak{a}$ is homogeneous if and only if $\mathfrak{a}$ is invariant under this action.

(2.50) Assume that k is an algebraically closed field and that $f(x, y)$ is a homogeneous polynomial in $k[x, y]$. Show that $f(x, y)$ splits as a product of linear factors.

HINT: If f is of degree d, it holds that $f(x,y) = y^d f(x/y, 1)$. Consider $f(x/y, 1)$ as a polynomial in $t = x/y$.

✲ (2.51) *Homogenization of polynomials.* Let k be a field and let $f \in k[x_1, \ldots, x_r]$ be any non-zero polynomial. Let d be the degree of f. Define a fresh polynomial $f^H \in k[x_0, \ldots, x_r]$ in one more variable by putting $f^H(x_0, \ldots, x_r) = x_0^d f(x_1/x_0, \ldots, x_r/x_0)$. Show that f^H is *homogenous* of degree d. Show the quality $f^H(1, x_1, \ldots, x_r) = f(x_1, \ldots, x_r)$.

Homogenization (homogenisering)

✲ (2.52) *Dehomogenization of polynomials.* The homogenization process described in the previous exercise has a natural reverse process called *dehomogenization* It is not canonical, but depends on the choice of a variables, which will be x_0 in this exercise. When $g \in k[x_0, \ldots, x_r]$ is homogenous of degree d, one puts $g^D(x_1, \ldots, x_r) = g(1, x_1, \ldots, x_r)$. Show that $g = x_0^s (g^D)^H$ for some non-negative integer $s \leqslant d$. Give examples to see that s actually can have any value between 0 and d.

Dehomogenization (avhomogenisering)

✲ (2.53) Assume that $A = \bigoplus_{i \geqslant 0} A_i$ is a graded integral domain with A_0 being a field. Show that any element homogenous of degree one is irreducible. Conclude that $k[x,y,z,w]/(xy - zw)$ is not a UFD. HINT: Work with components of highest degree.

(2.54) Let $A = \mathbb{Z}[x,y,z,w]/(xy - wz)$ show that the class of x is irreducible but not prime.

★

2.9 *The prime spectrum and the Zariski topology*

Prime spectrum of a ring (Primspekteret til en ring)

Every ring has a geometric incarnation called the *prime spectrum*. It is denoted Spec A if the ring is A, and its points are the prime ideals in A. The spectrum carries a topology called the *Zariski topology* after Oscar Zariski. The topological space Spec A depends functorially on the ring A; a ring map $\phi\colon A \to B$ induces a map $\tilde{\phi}\colon \operatorname{Spec} B \to \operatorname{Spec} A$ simply by sending a prime ϕ in B to the inverse image $\phi^{-1}(\mathfrak{a})$ (which is a prime ideal in A).

The spectra of rings are the building blocks for the schemes as constructed in the ingenious scheme theory of Alexander Grothendieck; they form an infinitely larger ocean with as yet huge unexplored regions, where the spectra merely constitute the shore.

In happy marriages the spouses exert a strong mutual influence, so also in the relationship between algebra and algebraic geometry. Several geometric features of Spec A are paramount to understanding algebraic properties of the ring A, and vice versa. Both modern number theory and modern arithmetic are inconceivable without the geometric language and the geometric intuition of spectra and schemes. However, in these matters, we shall only superficially scratch the surface; giving the basic definitions and a few examples.

There is another geometric construct antecedent of the schemes by about a century. Basically it goes back to René Descartes's idea of using coordinates and equations to describe geometric objects. We have all experienced parts of the menagerie of plane curves and surfaces in the space. In general, subsets of $\mathbb{C}^r$ (or subsets of k^r for any algebraically closed field k) being the common zeros of a set of polynomials, are called *closed algebraic sets*. When they satisfy certain additional conditions, they are called *varieties*, and varieties are the main objects of interest for many algebraic geometers.

Varieties (varieteter)

Prime spectra

(2.79) In order to make it a genuine geometric object, the prime spectrum will be endowed with a topology, which is called the *Zariski topology*, named after one of the fathers of modern algebraic geometry, Oscar Zariski. This topology is best defined by giving the closed subsets. With any ideal $\mathfrak{a}$ in A is associated a closed subset denoted $V(\mathfrak{a})$ whose members are the prime ideals containing $\mathfrak{a}$; that is, one has

The Zariski topology (Zariski topologien)

$$V(\mathfrak{a}) = \{\, \mathfrak{p} \subseteq A \mid \mathfrak{p} \text{ a prime ideal } \mathfrak{p} \supseteq \mathfrak{a} \,\}.$$

Oscar Zariski (1899–1986) Russian-born American mathematician

There are some axioms to be verified. First of all, $V(0) = \operatorname{Spec} A$ and $V(A) = \varnothing$ (recall that prime ideals by definition are proper ideals), so the empty set and the entire space are both closed. The two other axioms for a topology require that the union of finitely many closed subsets is closed (it suffices to check it for the union of two) and that the intersection of any family of closed sets is closed. To the former, observe that $V(\mathfrak{a}) \cup V(\mathfrak{b}) \subseteq V(\mathfrak{a}\mathfrak{b})$ since both $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a}$ and $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{b}$ hold. The inclusion $V(\mathfrak{a}\mathfrak{b}) \subseteq V(\mathfrak{a}) \cup V(\mathfrak{b})$ follows since if $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{p}$, either $\mathfrak{a} \subseteq \mathfrak{p}$ or $\mathfrak{b} \subseteq \mathfrak{p}$ according to Proposition 2.28 on page 40. Hence $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a}\mathfrak{b})$. To verify the latter axiom, notice the trivial fact that $\sum_i \mathfrak{a}_i$ lies in $\mathfrak{a}$ if and only if each summand $\mathfrak{a}_i$'s lies in $\mathfrak{a}$. Summing up, we have shown most of the following proposition.

PROPOSITION 2.80 (THE ZARISKI TOPOLOGY IS A TOPOLOGY) *Let A be a ring.*

i) $V(0) = \operatorname{Spec} A$ *and* $V(1) = \varnothing$*;*
ii) For any ideals $\mathfrak{a}$ *and* $\mathfrak{b}$ *in* A *it holds true that* $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a}\mathfrak{b})$*;*
iii) For any family $\{\mathfrak{a}_i\}_{i\in I}$ *of ideals one has* $V(\sum_{i\in I} \mathfrak{a}_i) = \bigcap_{i\in I} V(\mathfrak{a}_i)$*;*
iv) If $\mathfrak{a} \subseteq \mathfrak{b}$*, then* $V(\mathfrak{b}) \subseteq V(\mathfrak{a})$
v) $V(\mathfrak{a}) = V(\sqrt{\mathfrak{a}})$*.*

PROOF: What remains to be observed are the two last assertions. It is trivial that $\mathfrak{a} \subseteq \mathfrak{b}$ implies that $V(\mathfrak{b}) \subseteq V(\mathfrak{a})$, and the last assertion ensues from the radical $\sqrt{\mathfrak{a}}$ being contained in any prime ideal containing $\mathfrak{a}$. ❑

(2.81) The Zariski topology has certain peculiar features never met when working with mundane topologies like the ones of manifolds. For instance, there are lots of points

in Spec A that are not closed; so in particular the prime spectra tend to be seriously non-Hausdorff. One has:

LEMMA 2.82 *The closed points in* Spec A *are the maximal ideals.*

PROOF: We saw that every proper ideal is contained in a maximal ideal (Theorem 2.50 on page 2.50); hence $V(\mathfrak{a})$ will always have a maximal ideal as member. So if $\{\mathfrak{p}\}$ is closed; that is, equal to $V(\mathfrak{a})$ for some $\mathfrak{a}$, the prime ideal $\mathfrak{p}$ must be maximal.

On the other hand whenever $\mathfrak{m}$ is maximal, obviously $V(\mathfrak{m}) = \{\mathfrak{m}\}$ since no prime ideal strictly contains $\mathfrak{m}$. ❑

Examples

We do not intend to dive deeply into a study of prime spectra, but only to give a faint idea of what might happen, let us figure out which of the topological spaces with only two points can be a prime spectrum.

There are three non-homeomorphic topologies on a two-point set; the discrete topology with all points being closed (and hence all being open as well), the trivial topology whose sole closed sets are the empty set and the entire space, and finally, we have the so-called *Sierpiński space*, a two-point space with just one of the points being closed (and consequently the other being open). And two of these occur as Zariski topologies.

Sierpiński space (Sierpiński-rommet)

(2.29) The direct product of two fields $A = k \times k'$ has merely the two prime ideals $(0) \times k'$ and $k \times (0)$ which both are maximal. Hence Spec $k \times k'$ consists of two points and is equipped with the discrete topology.

(2.30) The ring $\mathbb{Z}_{(p)}$ of rational numbers expressible as fractions with a denominator prime to p has just two prime ideals, namely (0) and the principal ideal (p) (Example 2.23 on page 55). Hence $\{(0)\}$ is an open set being the complement of the closed point (p). Hence Spec A is the Sierpiński space.

(2.31) Finally, the trivial topology having no closed point, can not be the Zariski topology of any non-empty prime spectrum: in every ring different from the null-ring there are maximal ideals, and the spectrum of the null-ring is empty.

(2.32) *The spectrum of a polynomial ring:* As a counterweight to the peculiarity of the previous examples, let us consider a more mainstream situation, namely the spectrum Spec $k[t]$ of the polynomial ring over an algebraically closed field k (let it be $\mathbb{C}$, if you want). Ideals in $k[t]$ are all principal and are prime when generated by irreducible polynomials. But k being algebraically closed, the only irreducible polynomials are the linear ones, and so all non-zero prime ideals are maximal and of the shape $(t - a)$ for $a \in k$. Thus the closed points of Spec $k[t]$ are in a one-to-one-correspondence with k. Additionally, Spec $k[t]$ contains one point (0) (the zero-ideal is prime). It is neither open

nor closed, and its closure is the entire spectrum Spec $k[t]$, and it goes under the name of the *generic* point. This looks familiar when $k = \mathbb{C}$, we just get $\mathbb{C}$ adjoined one generic point, but be aware that the topology is far from being the usual one. The only closed sets are the finite unions of closed points. This topology is frequently called *the finite complement topology* since the non-empty open sets are precisely those with a finite and closed complement (see also Exercise 2.56 below).

The finite complement topology (endelig-komplement-topologien)

☆

Functoriallity

(2.83) The spectrum Spec A depends functorially on the ring A; it is a functor from the category of (commutative) rings to the category of topological spaces. To justify this assertion, we have got to tell how maps between rings are affected. If $\phi\colon A \to B$ is a map of rings, pulling back ideals along ϕ takes prime ideals to prime ideals; indeed, that $ab \in \phi^{-1}(\mathfrak{p})$ means that $\phi(a)\phi(b) \in \mathfrak{p}$, and so either $\phi(a) \in \mathfrak{p}$ or $\phi(b) \in \mathfrak{p}$ whenever $\mathfrak{p}$ is prime; hence $\phi^{-1}(\mathfrak{p})$ is prime. This allows the definition of the map $\widetilde{\phi}\colon$ Spec $B \to$ Spec A simply as the inverse image map, the important observation being that $\widetilde{\phi}$ is continuous:

PROPOSITION 2.84 *The map $\tilde{\phi}$ is continuous.*

PROOF: Let $\mathfrak{a} \subseteq A$ be an ideal. The one has $\widetilde{\phi}^{-1}(V(\mathfrak{a})) = V(\mathfrak{a}B)$; indeed, tautologically it holds true that $\mathfrak{a} \subseteq \phi^{-1}\mathfrak{a}$ if and only if $\phi(\mathfrak{a}) \subseteq \mathfrak{a}$. ❑

It is clear that when ϕ and ψ are composable maps between rings, it holds true that $\widetilde{\phi \circ \psi} = \widetilde{\psi} \circ \widetilde{\phi}$, and it is totally trivial that $\widetilde{\mathrm{id}_A} = \mathrm{id}_{\mathrm{Spec}\ A}$. So sending A's to Spec A and ϕ's to $\widetilde{\phi}$ indeed yields a functor.

Inverse images

(2.85) A byproduct of the proof above is that the inverse image under $\widetilde{\phi}$ of the closed set $V(\mathfrak{a})$ is homeomorphic to Spec $B/\mathfrak{a}B$. Indeed, the prime ideals in $B/\mathfrak{a}B$ are in a one-to-one correspondence with the prime ideals in B containing $\mathfrak{a}B$ (Theorem 2.19 on page 36), and these are, as we saw in the proof, precisely the points in Spec B mapping to points in $V(\mathfrak{a})$. Moreover, the whole lattice of ideals $\mathcal{I}(B/\mathfrak{a}B)$ is isomorphic to the lattice of ideals in B containing $\mathfrak{a}B$. This takes care of the topology; closed sets correspond to closed sets, and we have established the following:

PROPOSITION 2.86 *Let $\widetilde{\phi}\colon$ Spec $B \to$ Spec A be induced by $\phi\colon A \to B$. Then the inverse image $\widetilde{\phi}^{-1}(V(\mathfrak{a}))$ is homeomorphic to* Spec $B/\mathfrak{a}B$. *In particular, for any point $\mathfrak{a} \in$ Spec A the fibre over $\mathfrak{a}$ is naturally homeomorphic with* Spec $B/\mathfrak{a}B$.

Exercises

(2.55) Let A and B be two rings. Show that Spec $(A \times B)$ is the disjoint union of Spec A and Spec B.

(2.56) *Finite complement topology.* Let Y be an infinite set and let η be a point not in Y (whatever you want but a point in Y). The union $X = \{\eta\} \cup Y$ has a topology* whose closed sets, apart from $\{\eta\}$, $\varnothing$ and X itself, are the finite subsets of Y.

*The name finite complement topology; a slightly misleading name since Y is not open, but has a finite complement. The non-empty opens in the induced topology on Y, however, are those with a finite complement.

a) Show that this is a topology.
b) Let k be a field and let Y be the set of monic irreducible polynomials with coefficients from k. Show that X is homeomorphic to Spec $k[t]$.
c) Show that if k and k' are two fields of the same cardinality, then the spectra Spec $k[t]$ and Spec $k'[t]$ are homeomorphic.

(2.57) Let $c\colon \mathbb{C} \to \mathbb{C}$ for a moment denote complex conjugation. Describe the action of $\tilde{c}$ on the spectrum Spec $\mathbb{C}[t]$. What are the fixed points? Describe the map Spec $\mathbb{C}[t] \to$ Spec $\mathbb{R}[t]$ induced by the inclusion $\mathbb{R}[t] \subseteq \mathbb{C}[t]$.

✲ (2.58) *Distinguished open sets.* The Zariski topology has a particular basis of open sets called the *distinguished open sets*. For each element $f \in A$ there is one such open set $D(f)$ whose members are the prime ideals not containing f; that is, $D(f) = \{\, \mathfrak{p} \mid f \notin \mathfrak{p} \,\}$.

The distinguished open sets (særskilte åpne mengder)

a) Show that $D(f)$ is open.
b) Show that the distinguished open sets form a basis for the Zariski topology on Spec A.
c) Let $\mathfrak{a}$ be an ideal in A and let $\{D(f_i)\}$ be a family of distinguished open sets. Show that they form a cover of Spec $A \backslash V(\mathfrak{a})$ if the f_i's generate $\mathfrak{a}$.
d) Show that Spec A has the compactness property: any covering by distinguished open sets can be reduced to a finite covering. HINT: Spec A is the complement of $V(1)$.

★

Lecture 3

Unique factorization domains

When working with integers the Fundamental Theorem of Arithmetic is a most valuable tool used all the time, consciously or unconsciously. In a general ring however, a corresponding theorem does not hold, and one has to do without it. The birth of algebraic number theory, and by that, the beginning of commutative algebra, came as a response to this "defect". Luckily, in certain nice rings the Fundamental Theorem persists. These rings are called *unique factorization domains* or *factorial rings*, and are the objects we shall study in this chapter. Out of the inherent human laziness springs the acronym UFD, which is in widespread use.

Unique factorization domains (entydig-faktoriseringsområder)

Factorial rings (faktorielle ringer)

UFD's (UFD)

3.1 Being a unique factorization domain

(3.1) To be precise, a UFD is a domain where every non-zero element which is not a unit, can be expressed in an essentially unique way as a product

$$a = p_1 \cdot \ldots \cdot p_r \tag{3.1}$$

of irreducible elements p_i. The qualifyer "essentially unique" must be understood in the large sense; the order of the factors can of course be changed at will, and replacing a factor p_i with up_i, where u is a unit, can be compensated by multiplying another factor by the inverse u^{-1}. So "essentially unique" means that the factors are unique up to order and multiplication by units. Or in the terminology introduced in Paragraph 2.6 on page 31, the factors are unique up to order and association.

The definition has two separate conditions—a stipulation of existence and a uniqueness requirement—and the two are of a quite different flavour. Before proceeding with the theory, we shall take a closer look at each separately.

Existence

(3.2) The first conditions stipulates that any element can be expressed as a finite product of irreducibles. This is in essence a finiteness condition on A, which is fulfilled *e.g.* in

the so-called Noetherian (important rings to be introduced later). It does certainly not hold in general rings; an example can be the ring of entire functions in the complex plane $\mathbb{C}$. The irreducibles in this ring are of the form $u(z)(z-a)$ where $u(z)$ is a unit (*i. e.* a non-vanishing entire function) and $a \in \mathbb{C}$ is any point, so any entire function with infinitely many zeros—like our good old friend $\sin z$—can not be expressed a finite product of irreducibles.

One can always attempt a recursive attack in the search for a factorization. Any ring element a which is not irreducible, is a product $a = a_1 a_2$ of two non-units. These being irreducible, makes us happy—we have a factorization—but if one or both are not, they are in their turn products of non-units. If all the fresh factors are irreducible, we are again happy; if not, some split into products of non-units. Continuing like this we establish a recursive process which, if terminating, yields a finite factorization of a into irreducibles.

In general the process may go on for ever—like it will for *e.g.* $\sin z$—but in many cases there are limiting condition making it end. For instance, in the case of the ring of integers $\mathbb{Z}$, the number of steps is limited by the absolute value $|a|$, and in the case of polynomials in $k[x]$ by the degree of a. There is a general finiteness condition that guarantees the process to stop as expressed in the following lemma. It comes in the disguise of a condition of the partial ordered set of principal ideals and is an anticipation of the notion of Noetherian rings.

LEMMA 3.3 *Let A be a domain such that any non-empty collection of principal ideals has a maximal element. Then every element in A can be expressed as a finite product of irreducibles.*

PROOF: Let Σ be the set of principal ideals (a) where a runs through all counterexamples to the lemma; that is, all elements not expressible as a finite product of irreducibles. If the lemma were false, the set Σ would be non-empty, and by assumption it would have a maximal member, say (b). By construction b can not be irreducible and may be factored as $b = b_1 b_2$ with neither b_1 nor b_2 being a unit; hence (b) is strictly contained in both (b_2) and (b_2). On the other hand, b is not a finite product of irreducibles, and *a fortiori* the same holds for either b_1 or b_2. Therefore either (b_1) or (b_2) belongs to Σ, which is impossible since both strictly contain the maximal member (b). ❑

EXERCISE 3.1 Apply Zorn's lemma to prove that any principal ideal domain satisfies the condition in Lemma 3.3. HINT: Any ascending chain of principal ideals must terminate. ★

Uniqueness

(3.4) The second condition a UFD must abide to is the uniqueness requirement, which is of a more algebraic nature. It generally holds true that prime elements are irreducible, and the uniqueness requirement essentially boils down to the converse holding true; *i. e.*

that irreducible elements be prime. In fact, in any domain the uniqueness requirement holds automatically for finite factorizations into *prime* elements:

LEMMA 3.5 *Let A be a domain. Assume that $\{p_i\}_{1\leqslant i\leqslant r}$ and $\{q_i\}_{1\leqslant i\leqslant s}$ are two collections of prime elements from A whose products agree; that is, it holds true that*

$$p_1 \cdots p_r = q_1 \cdots q_s.$$

Then the p_i's and the q_i's coincide up to order and unit factors.

PROOF: The proof goes by induction on r. Since p_1 is prime, it divides one of the q_i's, and after renumbering the q_i's and adjusting q_1 by a unit, we may assume that $p_1 = q_1$. Cancelling p_1 gives $p_2 \cdots p_r = q_2 \cdots q_s$, and induction finishes the proof. ❑

(3.6) Since irreducibles are prime in both the rings of integers $\mathbb{Z}$ and of polynomials $k[x]$ over a field k (both are principal ideal domains), we immediately conclude that $\mathbb{Z}$ and $k[x]$ are factorial rings.

Examples

The main examples of factorial rings are principal ideal domains and polynomial rings over those—as we shortly shall see—but producing other examples demands some technology not available to us for the moment. So we confine ourselves to give some classical examples of non-factorial rings, one from number theory and two from algebraic geometry.

(3.1) Our first example, the ring $\mathbb{Z}[i\sqrt{5}]$, is ubiquitous in number theory texts, and we already met it on page 46. In $\mathbb{Z}[i\sqrt{5}]$ the number 6, for instance, has two distinct factorizations in irreducibles (see Example2.18 on page 46):

$$6 = 2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5}).$$

(3.2) One of the standard example from algebraic geometry, which geometers would call the coordinate ring of "the cone over a quadratic surface in projective 3-space", is the quotient ring $A = k[X, Y, Z, W]/(XY - ZW)$. Indicating the classes in A of the variables by lower case versions of their name, we have $A = k[x, y, z, w]$ with constituting relation

$$xy = zw. \tag{3.2}$$

In Example 3.4 on page 79 we saw that A is a domain, and the polynomial $XY - ZW$ being homogenous, A is a graded ring. It is not too challenging to see that the class of any non-zero linear form is irreducible in A (see Exercise 2.53 on page 62). Hence the relation (3.2) shows that A is not factorial.

This also gives an easy example of the intersection of two principal ideals being non-principal; *i. e.* their intersection is distinct from their product, in that $(x) \cap (z) = (xy, xz)$.

(3.3) *Elliptic curves I:* Another famous example from algebraic geometry is the ring $A = k[X, Y]/(Y^2 - X(X-a)(X-b))$ where a and b are elements in a field k whose characteristic does not equal two. In Exercise 3.2 below you are asked to show that A is an integral domain. The relation

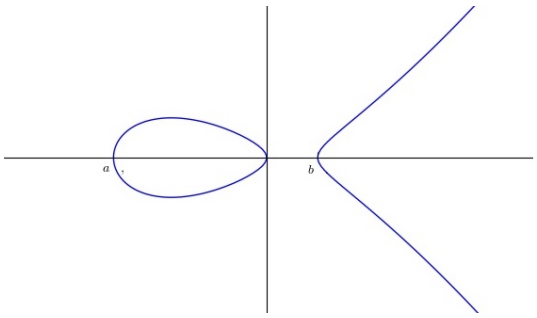
$$y^2 = x(x-a)(x-b), \tag{3.3}$$

where x and y denotes the classes of X and Y in A, holds in A and gives two different decompositions of an element into irreducibles; of course one must verify that the involved linear factors are irreducible (see Exercise 3.6 on page 81 below). Plane curves given by equations like

$$y^2 - x(x-a)(x-b)$$

with $a, b \in k$ are called *elliptic curves* when a and b are different and non-zero; to be precise one should say affine elliptic curves on Weierstrass form*. Elliptic curves have always been at the centre-stage of algebraic geometry and are closely related to the so-called *elliptic functions*—in fact, they were the very starting point for the development of modern algebraic geometry.

*Complete elliptic curves come in quite a lot of disguises, but can all be brought on a normal form called the Weierstrass form. If the characteristic is different from two or three, their affine incarnations are as described here.



Above we have included the sketch of the real points of such a curve with a and b real. Pictures can be beautiful and instructive, but should be taken with a grain of salt. If the ground field for instance, is the algebraic closure of the field $\mathbb{F}_3$ with three elements, or for that matter, the closure of the rational function field $\mathbb{Q}(t)$, the picture is of no relevance.

☆

Karl Theodor Wilhelm Weierstrass (1815–1897)

German mathematician

Exercises

(3.2) The aim of the exercise is to prove that for any field k the cubic polynomial $y^2 - x(x-a)(x-b)$ is irreducible in the polynomial ring $k[x, y]$. Were it not, it would

have a linear factor, and one could write

$$y^2 - x(x-a)(x-b) = (\alpha x + \beta y + \gamma)Q(x,y) \tag{3.4}$$

with α, β and γ constants from k and not both α and β being zero.

a) If $\beta \neq 0$, substitute $y = -\beta^{-1}(\alpha x + \gamma)$ in (3.4) to obtain the impossible polynomial identity
$$\beta^{-2}(\alpha x + \gamma)^2 - x(x-a)(x-b) = 0.$$

b) If $\beta = 0$, substitute $x = -\alpha^{-1}\gamma$ to make the right side of (3.4) vanish. Conclude that the left side will be a monic quadratic polynomial in y which is identically zero; which is absurd!

(3.3) Let the k-algebra $A = k[x,y,z]$ have the constituting relation

$$y^2 z - x(x-az)(x-bz)$$

where a are b are scalars.

a) Show that A is a graded domain HINT: Cast a glance at Exercise 3.2.

b) By evoking Exercise 2.53 on page 62, prove that y and $x - zc$ for any $c \in k$ are irreducible elements in A.

★

Irreducibles and primes in a UFD

(3.7) As we saw (Proposition 2.41 on page 45), in a domain prime elements are always irreducible, and as we are going to see, in factorial domains the converse holds as well. Among domains satisfying an appropriate finiteness condition that guarantees existence of a factorization—as for instance the Noetherian domains we shall come to—this even characterizes the factorial domains (Lemma 3.5 above).

PROPOSITION 3.8 *For members of a* UFD *being prime is equivalent to being irreducible.*

PROOF: We merely need to see that irreducibles are prime. So assume that a is irreducible, and that $a|xy$. Let $x = p_1 \cdots p_s$ and $y = q_1 \cdots q_t$ be decompositions into irreducibles. Then of course

$$xy = p_1 \cdots p_s \cdot q_1 \cdots q_r$$

is a factorization into irreducibles as well. On the other hand, xya^{-1} is an element in A and has a factorization $xya^{-1} = r_1 \cdots r_m$ into irreducibles; hence

$$a \cdot r_1 \cdots r_m = p_1 \cdots p_s \cdot q_1 \cdots q_r$$

are two equal products of irreducibles. The ring we work in being a UFD, irreducible factors coincide up to order and units, and this means that, up to a unit, a is either one of the p_i's or one of the q_i's. Phrased differently, a divides either x or y. ❑

3.2 *Common divisors and multiples*

The greatest common divisor (største felles divisor)

(3.9) In a UFD any two elements a and b have a *greatest common divisor*, which we shall denote by $\gcd(a,b)$. Recall that this is an element d such that $x|a$ and $x|b$ implies that $x|d$. Expressed in terms of ideals, d is an element whose principal ideal (d) is least among the principal ideals containing both principal ideals (a) and (b). The greatest common divisor of is only determined up to an invertible factor; however, the *principal ideal* (d) is unambiguously defined.

The notion of a greatest common divisor of two elements is meaningful in any domain, but in most domains not every pair has one. However, as mentioned above, in a factorial ring, any two elements do. The UFD's share this property with several other classes of rings, for instance, the rings of holomorphic functions in open domains in the complex plane $\mathbb{C}$. These rings are not UFD's, but have the property that any *finitely generated* ideal is principal.

The least common multiple (minste felles multiplum)

(3.10) Two elements may also have a *least common multiple*: an element m in A so that the principal ideal (m) is greatest among the principal ideals contained in (a) and (b); or phrased in terms of divisibility, it holds that $a|m$ and $b|m$, and for any other member x of A it ensues from $a|x$ and $b|x$ that $m|x$. We shall denote the least common multiple of a and b by $\operatorname{lcm}(a,b)$ Again, merely the principal ideal (m) is unambiguously defined.

PROPOSITION 3.11 *In a* UDF *any two elements have a greatest common divisor and a least common multiple.*

PROOF: Let a and b be the elements. Proceed to write down factorizations of a and b into irreducibles, say $a = p_1 \cdots p_r$ and $b = q_1 \cdots q_s$, and pick up the "common factors": Reordering the factors, we find a non-negative integer t so that $(p_i) = (q_i)$ for $i \leqslant t$ and $(p_i) \neq (q_i)$ for $t > i$. Then $d = p_1 \cdots p_t$ is a greatest common divisor. It might of course happen that no (p_i) equals any (q_j), in which case $t = 0$, and the greatest common divisor equals one.

To lay hands on a least common multiple of a and b mimic what we just did, or verify that $a \cdot b / \gcd(a,b)$ is a least common multiple of a and b. ❑

Exercises

(3.4) Show that two elements a and b from a domain A have a least common multiple if and only if the intersection $(a) \cap (b)$ is a principal ideal.

(3.5) Let a and b be two elements having a $\gcd(a,b)$ and a $\operatorname{lcm}(a,b)$. Show that $\gcd(a,b) \cdot \operatorname{lcm}(a,b) = ab$ up to a unit.

(3.6) Prove that in a ring where all finitely generated ideals are principal, all pairs of elements have a gcd and a lcm.

★

3.3 *A criterion of Kaplansky's*

(3.12) The author is especially fond of the formulation* in the following criterion found in Irving Kaplansky's book [**?**] and whose proof is an elegant application of the Basic Existence Theorem.

*The slogan is: "A domain A is factorial if and only if every prime contains a prime".

PROPOSITION 3.13 *A domain A is a* UFD *if and only if every non-zero prime ideal contains a prime element.*

Irving Kaplansky (1917–2006)
Canadian mathematician

PROOF: The implication one way is clear: Let $\mathfrak{p}$ be a non-zero prime ideal and consider any of its non-zero elements. It factors as a product of primes, and one of the factors must lie in $\mathfrak{p}$.

To prove the other implication it suffices, in view of Lemma 3.5 on page 69, to show that any non-zero member of A is either a unit or a finite product of prime elements, so let Σ be the set of elements in A that can be expressed as a finite product of prime elements. It is certainly multiplicative closed, and A having at least one maximal ideal it is not empty (maximal ideals are prime and contain prime elements by assumption). We contend that Σ coincides with the set of non-zero non-units of A.

Assume this is not true; that is, there is a member x of A, neither zero nor a unit, which does not lie in Σ. Then $(x) \cap \Sigma = \varnothing$; indeed, if there was an expression $xy = p_1 \cdots p_r$ with the p_i's being prime elements, we could chose one with r minimal, and this would force all the p_i's to divide x. Hence y would be a unit, and consequently $x \in \Sigma$. By the Basic Existence Theorem (Theorem 2.49 on page 49), there is a prime ideal in A maximal subjected to not meeting Σ and containing x. That prime ideal is not the zero ideal and by assumption therefore has a prime element as member, which is a contradiction since all primes lie in Σ. Hence Σ fills up the entire set of non-zero non-units in A. ❑

An immediate corollary is the following (which also can be proved in several other and more elementary ways):

COROLLARY 3.14 *Every* PID *is a* UFD.

PROOF: Prime ideals are generated by prime elements. ❑

There is an important corollary of Kaplansky's criterion valid for domains for which every non-zero ideal contains a prime ideal minimal among the non-zero prime ideals.

3.4 *Gauss' lemma and polynomials over factorial rings*

Fraction field (kvotientkropp)

Every domain A is contained in a canonically determined field called the *fraction field* of A. The elements are fractions of shape a/b with a and b elements from A and, of course, with $b \neq 0$. The arithmetic of these fractions is governed by the usual rules for rational

fractions. For the moment we have not shown that such fields exist, but shall assume it. Later on they will be constructed as particular cases of a general "localization process".

Several of the domains we have met so far are *a priori* contained in a field, like $\mathbb{Z}$ and $\mathbb{Z}[\sqrt{d}]$ which are subrings of $\mathbb{C}$, and a polynomial ring $k[x_1, \ldots, x_r]$ over a field k is contained in the rational function field $k(x_1, \ldots, x_r)$. So *a priori* these rings have a fraction field.

(3.15) The objective of the current section is to establish the important result that polynomial rings over UFD's are UDF's; a result that hinges on the key concepts of a primitive polynomial and the content of a polynomial, and a key lemma, the so-called Gauss' lemma.

Gauss' lemma was initially an approach to comparing factorizations of a polynomial in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$, but has a much broader horizon nowadays. As an illustration of the general mechanism consider, for instance, the simple polynomial $12x + 57$, which has the factorization $3 \cdot (4x + 19)$. When viewed as a polynomial in $\mathbb{Z}[x]$, it is not irreducible—neither 3 nor $(4x + 19)$ is a unit—but considered an element in $\mathbb{Q}[x]$ it is. Indeed, 3 becomes invertible in $\mathbb{Q}[x]$.

Content and primitive polynomials

(3.16) Proceeding along the lines of the example in the previous paragraph, imagine a general polynomial $f(x) = a_0 + a_1 x + \ldots + a_r x^r$ with coefficients from a factorial ring A. Extracting the greatest common divisor c_f of the coefficients a_i one may write $f = c_f \cdot f^\sharp$ where $f^\sharp \in A[x]$ is a polynomial whose coefficients have no common divisor. This naturally leads to the concept of a primitive polynomial: a polynomial over a UFD is called *primitive* if the greatest common divisor of the coefficients equals 1.

Primitive polynomials (primitive polynomer)

Content of a polynomial (inneholdet til et polynom)

With a splitting $f = c_f \cdot f^\sharp$ as above, where $f^\sharp$ primitive and $c_f \in A$, it is widespread usage to call the element c_f the *content* of f. It is not unambiguously determined by f; being a greatest common divisor, it is merely unique up to an invertible factor, and of course, $f^\sharp$ suffers the same ambiguity. Strictly speaking one should consider c_f an element in the quotient group K^*/A^* or equivalently, as the ideal (c_f) in the set of principal ideals.

(3.17) The notion of content may be extended to polynomials from $K[x]$ as well: Any such polynomial may be expressed as $f = c_f \cdot f^\sharp$ with $c_f \in K^*$ and $f^\sharp$ a primitive polynomial in $A[x]$; just multiply f by an element $d \in A$ so that $d \cdot f$ has coefficients in A (for instance, use the least common multiple of the coefficients), and put $c_f = d^{-1} \cdot c_{df}$ and $f^\sharp = (df)^\sharp$.

Both c_f and $f^\sharp$ will automatically be unique up to units in A. Indeed, if a relation $cg = c'g'$ with g, g' primitive polynomials in $A[x]$ and $c, c' \in K^*$ holds true, for some $a \in A$ one has $acg = ac'g'$ with $ac, ac' \in A$. The content of polynomials in $A[x]$ being unique up to a unit, it ensues that $ac = uac'$ for some unit $u \in A^*$. Cancelling a shows that $c = uc'$.

Notice that f has coefficients in A if and only if the content c_f belongs to A, and that f is primitive if and only if c_f is a unit in A.

Gauss's lemma

(3.18) This is one of many basic results to be found in Gauss' immortal *Disquisitiones Arithmeticae*. He wrote it in 1798 when he was 21 years old, and it was published in 1801. The *Disquisitiones Arithmeticae* is one of the most influential mathematical publications ever written; certainly among the all time top ten.

Johann Carl Friedrich Gauss (1777–1855)
German mathematician

LEMMA 3.19 (GAUSS'S LEMMA) *Assume that A is a* UFD. *Let f and g be primitive polynomials in $A[x]$. Then the product fg is primitive.*

PROOF: Write $f = \sum_{0\leqslant i\leqslant n} a_i x^i$ and $g = \sum_{0\leqslant i\leqslant m} b_i x^i$. Let d be a non-unit in A. The polynomial f being primitive, there is a least i_0 so that d does not divide a_{i_0} and ditto a least j_0 so that d does not divide b_{j_0}. Consider the coefficient of $x^{i_0+j_0}$ in the product fg; that is, the sum

$$\sum_{i+j=i_0+j_0} a_i b_j.$$

If $i \neq i_0$ and $j \neq j_0$, either $i < i_0$ or $j < j_0$; in the former case $d|a_i$ and in the latter $d|b_j$, so in both cases $d|a_i b_j$. Hence all terms of the sum are divisible by d except $a_{i_0}b_{j_0}$, and consequently the sum is not divisible by d. ❑

(3.20) The next lemma is an equivalent version of Gauss' lemma formulated in terms of the content of polynomials. It clearly implies Gauss' lemma, just apply it to two primitive polynomials, but as noted, the lemmas turn out to be equivalent.

LEMMA 3.21 *Assume that A is a* UFD *with fraction field K. For non-zero polynomials f and g in $K[x]$ it holds true that $c_{fg} = c_f c_g$ up to a unit factor.*

PROOF: One has $f = c_f \cdot f^\sharp$ and $g = c_g \cdot g^\sharp$ which gives $fg = c_f c_g f^\sharp g^\sharp$, but according to Gauss' lemma $f^\sharp g^\sharp$ is a primitive polynomial, and from the unicity of the content we deduce that $c_{fg} = c_f c_g$ up to units. ❑

(3.22) To facilitate future reference (and hopefully to make things clearer for the students) we sum up what we so far have done in this section:

PROPOSITION 3.23 *Let A a* UFD *with fraction field K. With very non-zero polynomial $f \in K[x]$ is associate an element $c_f \in K^*$ called the content of f, unique up to a unit factor. It holds true that $f = c_f \cdot f^\sharp$ where $f^\sharp$ is a primitive polynomial in $A[x]$, and this characterizes c_f and $f^\sharp$ up to unit factors.*

a) The content depends multiplicatively on f; that is, $c_{fg} = c_f c_g$ up to units;
b) f lies in $A[x]$ if and only if $c_f \in A$;
c) f is primitive if and only if $c_f \in A^*$.

Factoring polynomials over A and over K

We return to the issue mentioned in the beginning of the section: how does the factorizations of polynomials in $A[x]$ relate to its factorization in $K[x]$? And our approach will be in a setting with A a UFD and K its field of fractions.

(3.24) Let f be a primitive polynomial in $A[x]$ and assume that f splits as a product $f = gh$ in $K[x]$. It follows that $f = c_g c_h g^\sharp h^\sharp$, and from the content being unique up to units it ensues that $c_g c_h$ being an associate to c_f is a unit in A. Incorporating $c_g c_h$ in either $g^\sharp$ or $h^\sharp$, we arrive at a factorization $f = g'h'$ with $f, g' \in A[x]$. Notice that the new factors g' and h' are obtained from f and g merely by multiplying by elements from A, hence the degrees are preserved. We thus have established the following lemma:

LEMMA 3.25 *Assume that A is a UFD with quotient field K. A primitive polynomial $f \in A[x]$ that splits as product in $K[x]$, splits as a product in $A[x]$ with factors of the same degree. In particular, if f is irreducible in $A[x]$, it remains irreducible in $K[x]$.*

(3.26) We then have come to one of our main objectives in this chapter:

THEOREM 3.27 *If A is a UFD, then the polynomial ring $A[x]$ is a UFD. The irreducible elements in $A[x]$ are the irreducible elements in A and the primitive, irreducible polynomials in $A[x]$.*

Induction on the number of variables and the fact that $k[x]$ is a UFD immediately give the corollary that polynomial rings over fields are UFD's. It is also worthwhile mentioning that the same applies to polynomials rings over the integers.

COROLLARY 3.28 *If A is a UFD, the polynomial ring $A[x_1, \dots, x_r]$ is a UFD. In particular the rings $\mathbb{Z}[x_1, \dots, x_r]$ and $k[x_1, \dots, x_r]$ where k is a field, are UFD's.*

PROOF OF 3.27: Invoking Kaplansky's criterion (Proposition 3.13 above) it suffices to see that any prime ideal $\mathfrak{p}$ in $A[x]$ contains a prime element. If $\mathfrak{p} \cap A$ is non-zero, this follows from A being a UFD. If not, let $f \in \mathfrak{p}$ be a primitive polynomial of minimal degree. We contend that f is prime; so assume that $f|gh$. Certainly f will be irreducible (it is of minimal degree in $\mathfrak{p}$ and $\mathfrak{p} \cap A = 0$), it persists being irreducible in $K[x]$ and is therefore prime in $K[x]$ (the ring $K[x]$ is factorial). Consequently f divides either g or h, say $g = pf$ with $p \in K[x]$. But p is forced to lie in $A[x]$; indeed, $g = c_p p^\sharp f$, and c_p equals c_g up to a unit in A, hence belongs to A. ❑

Ideals in polynomial rings over PID's.

Gauss' lemma helps us better understand the prime ideals in the polynomial rings $A[x]$ over a PID A—this includes the polynomial rings $k[x, y]$ in two variables over a field k and the ring $\mathbb{Z}[x]$—and hopefully it will make you better appreciate David Mumford's drawing of Spec $\mathbb{Z}[x]$ in his famous red book **[?]**, a copy is shown below.

(3.29) The description of Spec $A[x]$ we are about to give, requires the principal ideal domain A to have infinitely many maximal ideals. When Spec A is finite, some of the

principal ideals $(g(x))$ might be maximal as illustrated in Exercise 3.8 below which treats the case when A has just one maximal ideal.

PROPOSITION 3.30 *Let A be a* PID *with infinitely many maximal ideals. Then the non-zero prime ideals $\mathfrak{p}$ in $A[x]$ are of two kinds.*

i) If $\mathfrak{p}$ is maximal, then $\mathfrak{p} = (g(x), p)$ where $p \in A$ is a prime element and $g(x)$ is a polynomial in $A[x]$ which is irreducible mod p.
ii) If $\mathfrak{p}$ is not maximal, it is principal and generated either by an irreducible and primitive polynomial or by a prime element in A.

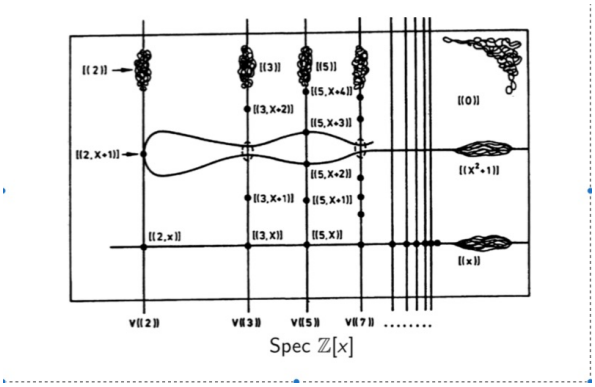
PROOF: The trick is to consider the intersection $\mathfrak{q} = \mathfrak{p} \cap A$ and separately treat the two cases according to $\mathfrak{q}$ being zero or not. Let K be the fraction field of A.

Assume first that $\mathfrak{q} = 0$. Then $\mathfrak{p}K[x]$ is a proper ideal: Assume there is a relation $1 = \sum_i a_i f_i$ with $a_i \in \mathfrak{p}$ and $f_i \in K[x]$ and multiply through by a common denominator d of all coefficients of the f_i's, and thus obtain $d \in \mathfrak{p}$. Now $d \in A$ and $\mathfrak{p} \cap A = 0$, and we have a contradiction.

It ensues that $\mathfrak{p}K[x]$ is principal, say generated by f. Replacing f by its primitive avatar $f^\sharp$, we may assume that f belongs to $A[x]$ and is primitive. We contend that $\mathfrak{p} = (f(x))$. Indeed, if $g \in \mathfrak{p}$ we may write $g = hf$ were h *a priori* belongs to $K[x]$. However, since f is primitive, we find that $c_h = c_h c_f = c_g$ lies in A, and hence h lies in $A[x]$.

Now, an ideal $(g(x))$ generated by an irreducible polynomial $g(x)$ is not a maximal: if it were, it would hold that $(g(x), p) = A[x]$ for all primes $p \in A$; in other words, $g(x)$ would be a unit in the polynomial ring $A/pA[x]$ for all primes p. But the leading coefficient of $g(x)$ has only finitely many prime factors (up to units), and the leading term survives in the reduction of $g(x)$ modulo any prime not among those. Our ring A is assumed to have infinitely many primes, so the reduction of $g(x)$ mod most of the primes in A (in fact, infinitely many) is not a unit. Contradiction.

Next, if $\mathfrak{q} \neq 0$, it holds that $\mathfrak{q} = (p)$ for some prime element $p \in A$. Consider $A[x]/pA[x] = k[x]$ were k is the residue class field $k = A/pA$. The ideal $\mathfrak{p}/(p)$ is either zero, in which case $\mathfrak{p} = pA[x]$, or it is generated by an irreducible polynomial $a(x)$, and any lift $g(x)$ of $a(x)$ to $A[x]$ will then generate $\mathfrak{p}$ together with p; that is, $\mathfrak{p} = (p, g(x))$. ❑



(3.31) Specializing A to be the polynomial ring $k[y]$ with k being an algebraically closed field, Proposition 3.30 yields the following description of the maximal ideals in $k[x,y]$, which is a precursor to the fabulous and all important Nullstellensatz of David Hilbert:

THEOREM 3.32 (NULLSTELLENSATZ IN DIMENSION TWO) *Let k denote a field that is algebraically closed. Then every maximal ideal $\mathfrak{m}$ in the polynomial ring $k[x,y]$ is of the form $\mathfrak{m} = (x-a, y-b)$ with a, b in k.*

PROOF: Indeed, the only irreducible polynomials in $k[x]$ are the linear ones. ❑

Exercises

(3.7) Assume that $p(x)$ is a monic polynomial in $\mathbb{Z}[x]$ which factors as $p(x) = r(x)s(x)$ in $\mathbb{Q}[x]$. Show that $r(x)$ and $s(x)$ both lie in $\mathbb{Z}[x]$ and are monic. HINT: Multiply through by the least common multiple of the coefficients' denominators and appeal to Gauss's lemma

⁂ (3.8) *Polynomial rings over DVR's.* Let A be a local PID (such rings are called *discrete valuation rings* abbreviated with the initialism DVR; they will be treated extensively in a later chapter) and let π be a generator for the maximal ideal; for instance, localizations like $\mathbb{Z}_{(p)}$ or $\mathbb{C}[t]_{(t-a)}$ are shaped like that.

a) Show that the principal ideal $\mathfrak{a} = (\pi x - 1)$ in the polynomial ring $A[x]$ is maximal. HINT: Let K be the fraction field of A and show that $\mathfrak{a}$ equals the kernel of the map $A[t] \to K$ that sends $f(x)$ to $f(1/\pi)$.

b) Show that any maximal ideal $\mathfrak{m}$ in $A[x]$ either is shaped like $(g(x), \pi)$ with $g(x)$ a polynomial in $A[x]$ which is irreducible modulo π, or like $(g(x))$ with $g(x)$ being an irreducible polynomial in $A[x]$ that is invertible modulo π. HINT: Adapt part of the proof of Proposition 3.30.

Separable polynomials (separable polynomer)

(3.9) *Separable polynomials and derivatives.* Let A be UFD with fraction field K. Polynomials without multiple roots in any field extension of K are said to be *separable*. Show that a polynomial $p(t)$ is not separable; that is, it has a multiple root in some field extension K' of K, if and only if there is a polynomial $q(t) \in A[t]$ of positive degree

which divides both $p(t)$ and its derivative $p'(t)$. HINT: Let α be the multiple root. Show that $p(\alpha) = p'(\alpha) = 0$. Consider the minimal polynomial of α over K and make it primitive.

★

Factoring homogenous polynomials

(3.33) Factoring a polynomial into a product of irreducibles, or for that matter, showing a polynomial is irreducible, is often an unpleasant task. Knowing that the polynomial is homogenous might sometimes be helpful as one then *a priori* knows that every irreducible factor will be homogeneous; indeed, one has the following proposition.

PROPOSITION 3.34 *Let R be a graded factorial domain satisfying $R_n = 0$ for $n < 0$. Then the irreducible factors of a homogeneous element are homogeneous.*

PROOF: Let a_i be the irreducible factors of a homogeneous element a and develop each a_i as a sum $a_i = \sum_\nu a_{i,\nu}$ of homogeneous components. Denote by a_{i,ν_i} the non-zero component of a_i of lowest degree. Since the degree of every element is non-negative, it holds true that

$$a = \prod_i a_i = \prod_i a_{i,\nu_i} + \text{ terms of higher degree,}$$

and $\prod_i a_{i,\nu_i}$ is non-zero as R is assumed to be a domain. But now, homogeneous components are unambiguously defined, and a is homogenous. Hence the sum of the high degree terms vanishes, and we have expressed a as a product of homogeneous elements

$$a = \prod_i a_{i,\nu_i}.$$

By induction on the degree, each a_{i,ν_i} has only homogeneous irreducible components, and the same applies therefore to a. ❑

EXAMPLE 3.4 The polynomial $f = xy - wz$ is irreducible. If f were the product of two linear terms, each variable would occur in precisely one of them since no term of f is a square, hence cross-terms like xw or xz would appear in f. ☆

EXAMPLE 3.5 The polynomials $y^p - x^q$ are irreducible when p and q are relatively prime. To see this give $k[x,y]$ the grading for which $\deg x = p$ and $\deg y = q$. An irreducible polynomial without constant term, unless it equals $\alpha \cdot x$ or $\beta \cdot y$, necessarily contains non-zero terms both of the form $\alpha \cdot y^n$ and of the form $\beta \cdot x^m$ for natural numbers n and m and scalars α and β. If it additionally is homogeneous, it holds true that $nq = mp$ and hence $n = ap$ and $m = bq$ for some non-negative integer a. It follows that if f is an irreducible factor of $y^p - x^q$ it is of degree at most pq, so either it reduces to x or y, or its degree is pq; the former case does obviously not occur, so f has degree pq, which is the same as the degree of $y^p - x^q$, and the two are equal up to a scalar. ☆

Exercises

✲ (3.10) Show that if $n \geqslant 3$, it holds that $q(x) = \sum_{1 \leqslant i \leqslant n} x_i^2$ is irreducible in $k[x_1, \ldots, x_n]$ unless that characteristic of k equals two.

(3.11) Let p_1, p_2 and p_3 be pairwise coprime integers. Consider the polynomial $f = x_1^{p_1} + x_2^{p_2} + x_3^{p_3}$ in the polynomial ring $k[x_1, x_2, x_3]$ over the field k. Show that f is irreducible.

This includes all connected Lie groups. For students initiated in Lie–theory, it is a nice exercise to verify this (What is the derivative of the n-power map?).

(3.12) Let R be a graded ring satisfying $R_n = 0$ for $n < 0$. Show that only homogeneous units in R are those in R_0^*

✲ (3.13) Let G be a group without finite quotients* and A a factorial ring. Assume that G acts on A in away that the units are invariant. Show the irreducible factors of any polynomial semi-invariant under G are semi-invariant. (Recall that f is called semi-invariant if there is a group homomorphism $\chi_f \colon G \to A^*$ so that $f^g = \chi_f(g) f$.)

★

3.5 *Example: Quadratic extensions and the norm*

An indispensable tool in algebraic number theory is the *norm*, but also in algebraic geometry it certainly plays a significant role. After having treated so-called integral extensions we shall expound on the norm and its stablemate the *trace*—but being indispensable in certain important examples we give this *ad hoc* treatment exclusively for the class of quadratic extensions.

We have already have come across the norm at occasions, *e.g.* when showing that 2 and 3 were irreducible elements in the quadratic extension $\mathbb{Z}[i\sqrt{5}]$, but in the guise of the square of the ordinary absolute value of complex numbers.

We have seen several examples where adjoining a square root to a ring is a central feature; the quadratic extensions $\mathbb{Z}[\sqrt{n}]$ where n is an integer are shaped like that, as is the coordinate ring of an affine elliptic curve: it equals $k[x, y]$ with constituting relation $y^2 = x(x-a)(x-b)$ so obtained form $k[x]$ by adjoining $\sqrt{x(x-a)(x-b)}$.

Of course, one is not confined to use cubic polynomial, but can adjoin $\sqrt{p(x)}$ where $p(x)$ is any polynomial (well, it must have certain good properties as being without multiple roots). The rings one obtains in this way are coordinate rings of the so-called affine *hyperellipitic curves* (they will be more closely discussed in Exercise 3.20).

These ring extensions have many features in common, and here we shall explore some. So we set the staging by assuming that A is a domain, and pick an element d from A which is not a square; furthermore we let $B = A[t]/(t^2 - d)$. Denoting the class of t by $\sqrt{d}$, one may write $B = A[\sqrt{d}]$ and think about B as A with the square root $\sqrt{d}$ adjoined.

Every $b \in B$ can be written as $b = x + y\sqrt{d}$ with x and y unique elements from a. indeed, no polynomial in $A[t]$ of degree one lies in $(t^2 - d)$, just consider the top term.

The multiplication in B is given by the formula

$$(x + \sqrt{dy})(x' + \sqrt{dy'}) = (xx' + yy') + (xy' + x'y)\sqrt{d}. \tag{3.5}$$

The extension B comes equipped with a conjugation $\sigma\colon B \to B$ map defined by $\sigma(x + y\sqrt{b}) = x - y\sqrt{b}$. The multiplication formula above immediately gives the relation $\sigma(ab) = \sigma(a)\sigma(b)$, and as evidently the conjugation is additive, it is a ring homomorphism. It is obviously an involution (applied twice it gives the identity), and if the characteristic of A is not two (in characteristic two σ degenerates into he identity), the invariant elements are precisely the members of A; indeed $x + y\sqrt{} = x - y\sqrt{d}$ if and only if $y = 0$ since we are in characteristic different from two. We have shown:

LEMMA 3.35 *The map σ is an involutive ring homomorphisms whose ring of invariants equals A.*

Next we introduce the norm $N(b)$ of elements from B as the product $N(b) = b\sigma(b)$.

PROPOSITION 3.36 *Let A, B and d be as above.*

i) The norm is multiplicative; i. e. $N(ab) = N(a)N(b)$;
ii) $N(x + y\sqrt{d}) = x^2 - dy^2$;
iii) An element $b \in B$ is invertible in B if and only if the norm $N(b)$ is invertible in A.

PROOF: The statement *i)* follows as the conjugation is a ring map: $N(ab) = ab\sigma(ab) = ab\sigma(a)\sigma(b) = a\sigma(a)b\sigma(b) = N(a)N(b)$. Next *ii)* ensues directly from the multiplication formula (3.5). Finally we prove *iii)*: if b is invertible it holds that $N(b)N(b^{-1}) = N(bb^{-1}) = N(1) = 1$; and if $N(b)$ is invertible, $N(b)^{-1}\sigma(b)$ serves as the inverse of b. ❑

EXAMPLE 3.6 *Elliptic curves II:* Let k be field and a and b two elements from k. Let A be the coordinate ring of an affine elliptic curve on Weierstrass form; that is, $A = k[x, y]$ with constituting relation $y^2 = x(x - a)(x - b)$. We let $p(x) = x(x - a)(x - b)$. The aim of this example is to complete Example 3.3 on page 70 by showing that A is not factorial. It remains to see that the factors in (3.3) are irreducible.

We consider the extension $k[x] \subseteq k[x, y] = A$ where $y = \sqrt{p}$. Elements are of the form $f + yg$ with $f, g \in k[x]$, whose norm equals $N(f + yg) = f^2 - pg^2$.

A salient points is that the norm $N(f + yg)$ is of degree as least three when $g \neq 0$; indeed, f^2 is of even degree and the degree pg^2 is odd, so dominating terms can not cancel. Hence $N(f + yg)$ is a scalar if and only if $g = 0$ and f is constant; and citing Proposition 3.36 above we conclude that the non-zero constants are the only units in A. Moreover, the norm $N(f + yg)$ is never of degree one.

With these observation up in our sleeve, we can check that any linear expression $z = \alpha y + \beta(x - \gamma)$ in x and y is irreducible (where α, β and γ lie in k). Assume there is

a factorization $z = uv$ and apply the norm to obtain

$$N(u)N(v) = \beta^2(x-\gamma)^2 - \alpha^2 p.$$

If $\alpha \neq 0$, the degree of the right hand side is three, and one of $N(u)$ and $N(v)$ must be of degree three and the other constant (degree one is forbidden) ; or if $\alpha = 0$, the right hand side is of degree two, so that either $N(u)$ or $N(v)$ must be of degree two and the other constant. In both cases, either u or v is invertible, and that is the end of the affair. ☆

EXAMPLE 3.7 *Units in real quadratic extensions:* Contrary to the imaginary quadratic extensions $\mathbb{Z}[i\sqrt{n}]$ that have a finite unit group, the real quadratic extensions $\mathbb{Z}[\sqrt{n}]$ (in both cases n is a natural number that is not a square) have infinitely many units, which constitute a group isomorphic to $\mathbb{Z} \times \mu_2$. We shall not prove this, but indicate why it holds true.

So let $a = x + y\sqrt{n}$ with $x, y \in \mathbb{Z}$ be an element in $\mathbb{Z}[\sqrt{n}]$. Its norm is given as $N(a) = x^2 - my^2$ and according to statement *iii)* of Proposition 3.36 above, a is a unit if and only if $N(a)$ is a unit; that is if and only if x and y satisfies the equation

$$x^2 - ny^2 = \pm 1.$$

Pell's equations (Pell's ligning)

This equation is called *Pell's equations*; iits history and can be traced far back and is loaded with anecdotes. It seems that the Indian mathematician Brahmagupta treated it extensively as early as in the year 628. It is not very deep, but requires a certain amount of work, to see that

$$x^2 - \sqrt{n}y^2 = 1$$

has a solution for any natural number n.

Let us take for granted there is a nontrivial unit v in $\mathbb{Z}[\sqrt{n}]$, and from that deduce there are infinitely many. We begin with showing there is a smallest unit larger than one. Indeed, if $u = x + y\sqrt{n}$, it holds that $u^{-1} = \pm 1(x - y\sqrt{n})$ (the sign is chosen according to $N(u) = 1$ or $N(u) = -1$) so that $u - u^{-1} = 2y\sqrt{n}$ or $u - u^{-1} = 2x$. In both cases $u - u^{-1}$ will be bounded away from zero since x and y are integers; so no sequence of units can approach 1 from above, and hence there must be a smallest one larger than one. Denote that smallest unit by u_0; it is called the *fundamental unit*, and we contend that every other unit is up to sign a power of u_0: Let u be any non-trivial unit. By exchanging u with $-u$ or u^{-1} or $-u^{-1}$ if necessary, we may assume that $u > 1$, and thus $u \geqslant u_0$. Let r be the smallest natural number so that $u_0^r \geqslant u$. Then $1 \leqslant u_0^r u_{-1} < u_0$; and in view of the minimality of u_0 we conclude that $u_0^r u^{-1} = 1$.

Apart from the existence of nontrivial units, we have thus shown that $\mathbb{Z}[\sqrt{n}]^* \simeq \mu_2 \times \mathbb{Z}$. ☆

Exercises

⁂ (3.14) Consider the ring $\mathbb{Z}[i\sqrt{2k}]$ with $k \geqslant 3$ an odd natural number. Show that $\mathfrak{p} = (2, i\sqrt{2k})$ is not a principal ideal, but that $\mathfrak{p}^2 = (2)$. Prove that the ring $\mathbb{Z}[i\sqrt{2k}]$ is not factorial.

(3.15) Show that 3 and 5 are irreducible members of $\mathbb{Z}[i\sqrt{14}]$ that are not prime.

(3.16) Consider the ring $\mathbb{Z}[i\sqrt{14}]$. Prove that

$$3^4 = (5 + 2i\sqrt{14})(5 - 2i\sqrt{14})$$

and show that all the involved elements are irreducible elements of $\mathbb{Z}[i\sqrt{14}]$.

(3.17) Assume that d is an integer such $d \equiv 1 \mod 4$. Let $\alpha = (1 + \sqrt{d})/2$. Show that $\alpha^2 = \alpha + (d-1)/4$. Prove that $A = \mathbb{Z}[\alpha]$ is free $\mathbb{Z}$-module of rank 2 with $1, \alpha$ as a basis. Determine the matrix of the map $x \mapsto \alpha x$ in this basis and compute the characteristic polynomial. Describe the norm-map.

⁂ (3.18) *The ring of real trigonometric polynomials.* The ring $A = \mathbb{R}[x, y]$ with constituting relation $x^2 + y^2 = 1$ is the ring of real polynomial functions on the unit circle in $\mathbb{R}^2$, or if you wish, you may view it as the ring of trigonometric polynomials; just put $x = \sin t$ and $y = \cos t$.

a) Show that $\mathbb{R}[x]$ is a polynomial ring and that A is a free module of rank two over $\mathbb{R}[x]$ with 1 and y as a basis;

b) Let N denote the norm defined by the extension $\mathbb{R}[x] \subseteq A$. Show that the norm is given as $N(f(x) + g(x)y) = f^2(x) - g^2(x)(1 - x^2)$, and that the non-zero constants are the only units in A;

c) Show that y, $(1 - x)$ and $(1 + x)$ are irreducible elements in A and conclude that A is not a UFD. HINT: $y^2 = (1 - x)(1 + x)$.

(3.19) *The ring of complex trigonometric polynomials.* Contrary to the ring A from the previous exercise, the ring $B = \mathbb{C}[x, y]$ with constituting relation $x^2 + y^2 = 1$ is a UFD, it is even a PID.

a) Show that all non-zero and proper ideals in B are of the form $(x - a, y - b)$ with a, b complex numbers such that $a^2 + b^2 = 1$. HINT: Theorem 3.32 on page 78;

b) Show that $B = \mathbb{C}[u, \bar{u}]$ with $u = x + iy$ and $\bar{u} = x - iy$ and that $u\bar{u} = 1$;

c) Show that $(x - a, y - b) = (u - c)$ where $u = x + iy$ and $c = a + ib$.

(3.20) *Hyperelliptic curves.* Let k be a field. Let p be a polynomial in $k[x]$ of degee n without multiple roots, and let $A = k[x, y]$ with constituting relation $y^2 = p$. This ring is the coordinate ring of a so-called *hyperelliptic curve*

Hyperelliptic curves (hyperelliptiske kurver)

a) Show that A is an integral domain which is a free $k[x]$-module of rank two;

b) Compute the norm map $A \to k[x]$;

c) Assume that the degree of p is odd. Show that the units in A coincide with the non-zero constants and that A is not factorial;

d) Assume the characteristic of k is not two. Show by way of examples that for each even n there are hyperelliptic curves with non-constant units. HINT: Let q be a polynomial of degree ν assuming each of the values 1 and -1 in ν distinct points and consider $p = (1+q)(1-q)$.

⁂ (3.21) Determine the fundamental unit in $\mathbb{Z}[2]$, $\mathbb{Z}[\sqrt{3}]$ and $\mathbb{Z}[\sqrt{5}]$.

Lecture 4

Modules

Along with every ring comes a swarm of objects called modules; they are the additive groups on which the ring acts. The axioms for modules resemble the axioms for a vector spaces, and modules over fields are in fact just vector spaces. Over general rings however, they are much more diverse and seriously more complicated. Ideals for instance, are modules, and any over-ring is a module over the subring, to mention two instances. An abelian group is nothing but a $\mathbb{Z}$-module, and a module over the polynomial ring $k[t]$ over a field k is just a vector space over k endowed with an endomorphism; so the module theory encompasses the theory of abelian groups and the entire linear algebra!

4.1 The axioms

(4.1) A *module* M over the ring A, or an *A-module* as one also says, has two layers of structures. It is endowed with an underlying structure as an abelian group, which will be written additively, on top of which lies a linear action of the ring A. Such an action is specified by a map $A \times M \to M$, whose value at (a, m) will be denoted by $a \cdot m$ or simply by am. It is subjected to the following four conditions: *Modules (moduler)*

i) $a(m + m') = am + am'$;
ii) $(a + a')m = am + a'm$;
iii) $1 \cdot m = m$;
iv) $a \cdot (a'm) = (aa') \cdot m$.

where $a, a' \in A$ and $m, m' \in M$ are arbitrary elements. The first condition requires the action to be A-linear; or in other words, the map $A \to \mathrm{Hom}_{\mathsf{Sets}}(M, M)$ that sends a to the "multiplication-by-a-map" $m \mapsto am$ must take values in the ring $\mathrm{Hom}_{\mathsf{Ab}}(M, M)$ of group-homomorphism. The last three requirements ensure that this map is a ring homomorphism; it is the *ring* A that acts.

(4.2) One recognizes these conditions from linear algebra; they are word for word the same as the vector space axioms, the sole difference being that A is not required to be a

field, but can be any ring. So in case A is a field k, there is nothing new; a k-module is just a vector space over k. However, one should not draw this analogy to far; general modules are creatures that behave very differently from vector spaces.

Examples

(4.1) The primordial examples of modules over a ring A are the ideals $\mathfrak{a}$ in A and the quotients $A/\mathfrak{a}$. Already here, the difference from the case of vector spaces surfaces; fields have no non-zero and proper ideals. There are also the "subquotients" $\mathfrak{b}/\mathfrak{a}$ of two nested ideals. Of course, these examples include the ring itself; every ring is a module over itself.

(4.2) Another examples more in the flavour of vector spaces are the direct sums of copies of A. The underlying additive group is just the direct sum $A \oplus A \oplus \ldots \oplus A$ of a finite number, say r, copies of A. The elements are r-tuples $(a_1, \ldots, a_r)$, and addition is performed componentwise. The action of A is also defined componentwise: $a \cdot (a_1, \ldots, a_r) = (a \cdot a_1, \ldots, a \cdot a_r)$. We insist on this being an additive* construction and shall write rA for this module, as distinguished from the common usage A^r in linear algebra.

*The direct sum plays an additive role in the category of A-modules, the multiplicative role is taken by another construct, the tensor product. But that's for a later chapter.

(4.3) Another familiar class of modules are the abelian groups. They are nothing but modules over the ring $\mathbb{Z}$ of integers. An integer n acts on an element from the abelian group by just adding up the appropriate number of copies of the element and then correcting the sign.

(4.4) Over-rings form an abundant source of examples—when A is a subring of B, multiplication by elements from A makes B into an A-module. So for instance, $k[x, y]$ is a $k[x]$-module as is $k[x, x^{-1}]$. And $k[x]$ will be a module over the subring $k[x^2, x^3]$. If $\eta \in \mathbb{C}$ is a root of unity, $\mathbb{Z}[\eta]$ is a $\mathbb{Z}$-module .

More generally, any ring homomorphism $\phi\colon A \to B$ induces an A-module structure on B through the action $a \cdot b = \phi(a)b$ of an element $a \in A$ on $b \in B$. This gives B the structure of an *A-algebra* as defined in Paragraph 1.12 on page 19.

Algebras (algebraer)

(4.5) Suppose that k is a field. Giving a $k[t]$-module is the same as giving a k-vector space M and an endomorphism of M; that is, a linear map $\tau\colon M \to M$. A polynomial $p(t)$ in $k[t]$ acts on M as $p(t) \cdot m = p(\tau)(m)$.

☆

✲ EXERCISE 4.1 Let A and B be two rings and assume that B has an A-module structure compatible with the ring structure; *i. e.* $a \cdot bb' = b(a \cdot b')$. Show that there is ring homomorphism $A \to B$ inducing the module structure. ★

Homomorphisms between modules

A new concept in mathematics is always followed by a fresh class of relevant maps; so also in our present case of modules. An *A-module homomorphism* $\phi\colon M \to N$ between

Module homomorphisms (modulhomomorfier)

two A-modules M and N is a homomorphism of the underlying abelian groups that respects the action of A; that is, $\phi(am) = a\phi(m)$ for all a's in A and all m's in M. Simply said, a module homomorphism is just *an A-linear map* from M to N. With this notion of morphisms, the A-modules form a category Mod_A. (It is easily checked that the composition of two A-linear maps is A-linear and, of course, identity maps are A-linear as well.) An A-module homomorphism $\phi\colon M \to N$ is said to be an *isomorphism* if there is another one $\psi\colon N \to M$ which is a two-sided inverse to ϕ; *i. e.* it holds true that $\phi \circ \psi = \mathrm{id}_N$ and $\psi \circ \phi = \mathrm{id}_M$. One easily verifies that it suffices to be bijective for ϕ to be an isomorphism.

Isomorphisms of modules (modulisomorfier)

(4.3) The set $\mathrm{Hom}_A(M,N)$ of A-linear maps from M to N is naturally contained in the set $\mathrm{Hom}_{\mathbb{Z}}(M,N)$ of group homomorphism from M to N (which are just the additive maps), consisting of those commuting with the actions of A on M and N. It is well-known that the sum of two additive maps is additive, and when both commute with the actions of A, the sum does so as well. Therefore $\mathrm{Hom}_A(M,N)$ is an abelian group, and defining $a \cdot \phi$ as the map that sends m to $a\phi(m)$, gives it an A-module structure. One must of course verify that $a \cdot \phi$ is A-linear, but this is easy: if $b \in A$ is another element, one finds

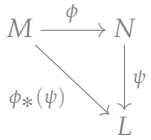
$$a \cdot \phi(bm) = a(b\phi(m)) = b(a\phi(m)),$$

where the first equality holds since ϕ is A-linear and the second because A is commutative*. The module axioms follow readily.

For non-commutative rings A the set $\mathrm{Hom}_A(M,N)$ is in general merely an abelian group; it does not carry an A-module structure unless further hypotheses are imposed on A or the involved modules.

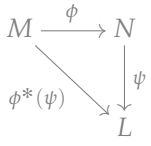
(4.4) The module $\mathrm{Hom}_A(M,N)$ depends functorially on both variables M and N, and historically, it was one of the very first functors to be studied. The dependence on the first variable is contravariant—the direction of arrows are reversed— whereas the dependence on the second is covariant—directions are kept. The induced maps are just given by composition. Of course, such constructions are feasible in all categories, what is special in Mod_A is that $\mathrm{Hom}_A(M,N)$ is an A-module and the induced maps are A-linear. The technical name is that category Mod_A has *internal homs*—the set of maps stay within the family To be precise let M, N and L be three A-modules and $\psi\colon N \to L$ an A-linear map. Sending ϕ to $\psi \circ \phi$ yields an associated map

$$\psi_*\colon \mathrm{Hom}_A(M,N) \to \mathrm{Hom}_A(M,L).$$



It is A-linear, and if $\psi\colon L \to L'$ is another A-linear map, one has $(\psi' \circ \psi)_* = \psi'_* \circ \psi_*$. In a similar fashion, the contravariant upper-star version

$$\phi^*\colon \mathrm{Hom}_A(N,L) \to \mathrm{Hom}_A(M,L),$$



which sends ψ to $\psi \circ \phi$, is A-linear as well, and it is functorial; *i. e.* $(\phi' \circ \phi)^* = \phi^* \circ \phi'^*$ for composable maps ϕ and ϕ'.

(4.5) It follows readily from the involved maps being A-linear that composition of

*That is, ϕ and ϕ' are maps from M to N and ψ and ψ' from N to L.

composable maps is an A-bilinear operation. That is, one has

$$\phi \circ (a\psi + a'\psi') = a\phi \circ \psi + a'\phi \circ \psi' \text{ and } (a\phi + a'\phi') \circ \psi = a\phi \circ \psi + a'\phi' \circ \psi,$$

where the maps are composable*, and a and a' denote ring elements.

Submodules

Submodules (Undermoduler)

(4.6) A *submodule* N of an A-module M is a subgroup closed under the action of A; in other words, for arbitrary elements $a \in A$ and $n \in N$ it holds true that $an \in N$, and of course, N being a subgroup the sum and the difference of two elements from N belong to N.

Examples

(4.6) Ideals in the ring A are good examples of submodules, and in fact, by definition, they are all the submodules of A.

(4.7) If $\mathfrak{a} \subseteq A$ is an ideal and M an A-module, the subset $\mathfrak{a}M$ of M formed by all finite linear combinations $\sum_i a_i m_i$ with $a_i \in \mathfrak{a}$ and $m_i \in M$ is a submodule.

(4.8) Given an ideal $\mathfrak{a}$ in A. The set $(0 : \mathfrak{a})_M$ of elements in M annihilated by all members of $\mathfrak{a}$ form a submodule. It holds true that $\phi \mapsto \phi(1)$ gives an isomorphism $\mathrm{Hom}_A(A/\mathfrak{a}, M) \simeq (0 : \mathfrak{a})_M$. Indeed, the value $\phi(1)$ of a map $\phi\colon A/\mathfrak{a} \to M$ is killed by $\mathfrak{a}$ since for $a \in \mathfrak{a}$ it holds true that $a \cdot \phi(1) = \phi(a \cdot 1) = \phi([a]) = 0$. To obtain a resiproque map, assume that $m \in (0 : \mathfrak{a})_M$ is given. The product $x \cdot m$ does only depend on the class $[x]$ of x as $(x + a) \cdot m = x \cdot m$ for elements a that kill m. Hence $[x] \mapsto x \cdot m$ is a legitimate definition of a map $A/\mathfrak{a} \to M$, and it takes the value m at 1.

☆

Exercises

☼ (4.2) Show that there is a canonical isomorphism $\mathrm{Hom}_A(A, M) \simeq M$. HINT: The correspondence is $\phi \leftrightarrow \phi(1)$.

(4.3) Show that $\mathrm{Hom}_A(A/\mathfrak{a}, A) = 0$ whenever $\mathfrak{a}$ is a non-zero ideal in a domain A. Show, *e.g.* by giving examples, that the equality does not necessarily hold true when A is not a domain. Find an example with A having just four elements.

(4.4) Let p and q be two prime numbers. Show that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/q\mathbb{Z}) = 0$ if $p \neq q$, and that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/p\mathbb{Z}) \simeq \mathbb{Z}/p\mathbb{Z}$.

(4.5) Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in the ring A. Show that there is a canonical isomorphism $\mathrm{Hom}_A(A/\mathfrak{a}, A/\mathfrak{b}) \simeq (\mathfrak{b} : \mathfrak{a})/\mathfrak{b}$. HINT: The correspondence is $\phi \leftrightarrow \phi(1)$.

(4.6) Assume that M is an A-module. For each element $a \in A$ let $[a]$ denote the "multiplication–by–a" map in M. Let N be a second A-module. Show that $[a]_* = [a]$ and $[a]^* = [a]$ (for each occurrence of $[a]$, it should self-explanatory in which of the modules M, N or $\mathrm{Hom}_A(M, N)$ multiplication by a takes place).

★

The lattice of submodules

(4.7) Just as the ideals in A the submodules of a given A-module M form a partially ordered set under inclusion*, which we shall denoted $\mathcal{I}(M)$.

The intersection $\bigcap_{i\in I} N_i$ of a collection $\{N_i\}_{i\in I}$ of submodules of M is a submodule. It is the largest submodule of M contained in all the submodules from the collection. In the similar way, the smallest submodule containing all the modules in the collection is the sum $\sum_{i\in I} N_i$, whose elements are finite A-linear combinations of elements from the N_i's; that is, the elements are shaped like

$$\sum a_i m_i, \tag{4.1}$$

where $m_i \in N_i$, and the a_i's are elements from A only finitely many of which are non-zero. More generally, for any set $S \subseteq M$ there is a smallest submodule of M containing S; it is called *the submodule generated* by S and consists of elements as in (4.1) but with the m_i's confined to S.

The set $\mathcal{I}(M)$ forms what is called a complete lattice. *A partially ordered set is called a* lattice *when every pair of elements possesses a least upper bound and a greatest lower bound, and it is said to be* complete *if the same is true for any subset. In our case the greatest lower bound is the intersection and the least upper bound the sum.*

Submodules generated by elements
(undermoduler generert av elementer)

EXERCISE 4.7 This exercise parallels the list of properties of direct and inverse images of ideals in Propositions 2.11 and 2.13 on pages 33 and 33. Let $\phi\colon M \to N$ be an A-linear map.

a) Show that for any submodule $L \subseteq N$ the inverse image $\phi^{-1}(L)$ is a submodule of M, and that the induced map $\phi^{-1}\colon \mathcal{I}(N) \to \mathcal{I}(M)$ respects inclusions and takes arbitrary intersections to intersection. What about sums of ideals?

b) Moreover, if $\mathfrak{a}$ is an ideal show that $\mathfrak{a}\phi^{-1}(L) \subseteq \phi^{-1}(\mathfrak{a}L)$, and give examples that strict inclusion may occur (one can even find examples with ϕ an inclusion);

c) In the same vein, show that $\phi(L)$ is a submodule of N for each submodue $L \subseteq M$, and that the induced map $\mathcal{I}(N) \to \mathcal{I}(M)$ respects inclusions. What happens with intersections and sums? And what about $\phi(\mathfrak{a}L)$?

★

Kernels and images

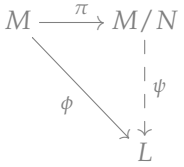
(4.8) An A-module homomorphism $\phi\colon M \to N$ is in particular a group homomorphism and as such has a kernel and an image. Both these subgroups are submodules as well; this ensues from the equality $a\phi(m) = \phi(am)$ satisfied by A-linear maps. Indeed, one immediately sees that the image is closed under multiplication by elements from A, and if $\phi(m) = 0$, it follows that $\phi(am) = a\phi(m) = 0$ as well.

Quotients

(4.9) Just as with ideals in a ring, one can form quotient of a module by a submodule and the construction is word for word the same. Let M be the module and N the submodule. From the theory of abelian groups we know that the two underlying

additive groups have a quotient group M/N, which is formed by the cosets $[m] = m + N$ for $m \in M$. Endowing M/N with an A-module structure amounts to telling how elements $a \in A$ act on M/N, and one does this simply by putting $a \cdot [m] = [am]$. Of course, some verifications are needed. The first is to check that the class $[am]$ only depends on the class $[m]$ and not on the representative m, which is the case since $a(m + N) = am + aN \subseteq am + N$. Secondly, the module axioms in Paragraph 4.1 must be verified; this is however straightforward and left to the zealous students.

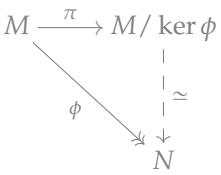
(4.10) The quotient group M/N comes together with the canonical defined additive map $\pi\colon M \to M/N$ that sends m to the class $[m]$. Moreover, by the very definition of the module structure on M/N, this map is A-linear, and it enjoys the important universal property that any A-linear map that vanishes on N, factors through it:



PROPOSITION 4.11 (UNIVERSAL PROPERTY OF QUOTIENTS) *Let N be a submodule of the A-module M. The quotient map $\pi\colon M \to M/N$ enjoys the following universal property. For every A-module homomorphism $\phi\colon M \to L$ with $N \subseteq \ker\phi$ there exists a unique A-linear map $\psi\colon M/N \to L$ so that $\phi = \psi \circ \pi$.*

PROOF: The proof is *mutatis mutandis* the same as for (abelian) groups. The map ϕ vanishes on N and is therefore constant on the residue classes $[m] = m + N$, and $\psi([m])$ is defined as (and compelled to be) that constant value. Since ϕ is A-linear and vanishes on N, the constant value on $[m + m'] = m + m' + N$ equals $\phi(m) + \phi(m')$, and on $[am] = am + N$ it is $a\phi(m)$. Hence ψ is A-linear. ❑

COROLLARY 4.12 (THE FIRST ISMORPHISM THEOREM) *An A-linear map $\phi\colon M \to N$ which is surjective, induces an isomorphism $M/\ker\phi \simeq N$.*



PROOF: By the universal property ϕ factors through a map $\psi\colon M/\ker\phi \to N$. This is surjective since ϕ is, and injective since it kills the kernel. (That a class $[x]$ goes to zero, implies that $\phi(x) = 0$, hence $x \in \ker\phi$). ❑

(4.13) One easily establishes the two following results. The analogue assertions for abelian groups are well known, and the proofs persist being valid for modules as well. The proofs are left as exercises; for inspiration either recall the proofs for abelian groups (these proofs go through *mutatis mutandis*) or take a look at the corresponding statements for ideals.

PROPOSITION 4.14 *Let $\pi\colon M \to M/N$ be the quotient map. The "inverse-image-map"*

$$\pi^{-1}\colon \mathcal{I}(M/N) \to \mathcal{I}(\mathcal{M})$$

that sends and ideal $\mathfrak{a}$ to $\pi^{-1}(\mathfrak{a})$, is a one-to-one correspondence between submodules of M containing N and submodules of M/N. It respect inclusions, arbitrary intersections and arbitrary sums.

PROOF: The proof is similar to the proof of Proposition 2.19 on page 36 and is left to the students as an exercise. ❑

PROPOSITION 4.15 (THE SECOND ISOMORPHISM THEOREM) *Assume N and N' are two submodules of M. There are then canonical isomorphisms where in the second one assumes that $N' \subseteq N$:*

i) $(N + N')/N' \simeq N/N \cap N'$;
ii) $(M/N')/(N/N') \simeq M/N$.

PROOF: The proof is similar to the proof of the isomorphism theorem for ideals (Theorem 2.21 on page 37) and is left as a DIY-proof. ❑

Cokernels

(4.16) In a famous paper Alexander Grothendieck introduced axiomatically the notion of an *abelian category*. The axioms reflect the main categorical properties of the module category Mod_A. Among the requirements is that there is a zero-objects, that all hom-sets are abelian groups and that all composition maps are bilinear (as we discussed in Paragraph 4.4). Moreover, all maps are requested to have kernels and a cokernels, and finally, there is an axiom which fabulously can be formulated as "the kernel of the cokernel equals the cokernel of the kernel".

Abelian category (Abelske kategorier)

Let us also mention that a category with a zero object whose hom-sets are abelian groups and whose compositions are bilinear is called an *additive category*. When, as is true of Mod_A, the hom-sets are A-modules and the compositions A-bilinear, it is said to an *A-linear category*. And not to forget, a *zero-object* in a category is an objcet 0 so that for each object C in the category, there is precisely one arrow from 0 to C (so that 0 is an *initial* object) and dually, for each C there is precisely one arrow from C to 0 (and 0 is thus a *final* or a *terminal* object).

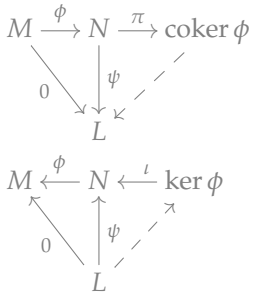
Additive categories (Additive kategorier)

A-linear category (A-lineære kategorier)

Zero object (nullobjekt)

(4.17) The definition of the *cokernel* in a categorical vernacular must be formulated exclusively in terms of arrows and therefore given as a universal property. The cokernel of an A-linear map $\phi\colon M \to N$ is an A-linear map $\pi\colon N \to \operatorname{coker}\phi$ such that $\pi \circ \phi = 0$, and which is universal with respect to that property. By The First Isomorphism Theorem (Theorem 4.12) the quotient $N/\operatorname{im}\phi$ fulfils that requirement, and hence serves as the cokernel of ϕ. The two stablemates kernel and cokernel are from a categorical viewpoint dual concepts, and a definition of the kernel just in terms of arrows is as indicated with a diagram in the margin: The kernel is an arrow $\iota\colon \ker\phi \to M$ satisfiying $\phi \circ \iota = 0$ with the universal property that any A linear map $\psi\colon L \to N$ such that $\phi \circ \psi = 0$, factors through it.

The cokernel (Kokjernen)



PROPOSITION 4.18 *Every A-module homomorphism $\phi\colon M \to N$ has a kernel, an image and a cokernel.*

PROOF: As mentioned above, the quotient $N/\operatorname{im}\phi$ serves as the cokernel. The kernel is the usual tangible subset of M consisting of the elements sent to zero. ❑

Examples

(4.9) In Mod_A the fabulous axiom cited above boils down to the obvious: The kernel of the cokernel and the cokernel of the kernel both equal the image. (If you find this rather more cryptical than obvious, think twice.)

(4.10) The submodules $\mathfrak{a}M$ where $\mathfrak{a}$ is an ideal in A form a particular important class of submodules of M. A quotient $M/\mathfrak{a}M$ inherits a natural structure of module over the quotient ring $A/\mathfrak{a}$; indeed, the product $x \cdot m$ between elements $x \in A$ and $m \in M/\mathfrak{a}M$ only depends on the residue class $[x]$ of x modulo $\mathfrak{a}$ since $(x+a)m = xm + am = xm$ for any $a \in \mathfrak{a}$. So for instance, the module M itself is in a canonical way a module over $A/\operatorname{Ann} M$; or for that matter, over $A/\mathfrak{a}$ for any ideal $\mathfrak{a}$ that kills M.

(4.11) Given a ring map $\phi\colon A \to B$ between two rings A and B, allows one to consider any B-module M as an A-module just by letting members a of A act on elements $m \in M$ as $\phi(a) \cdot m$. In this ways one obtains a natural functor from Mod_B to Mod_A. Sometimes one sees the notation M_ϕ or M_A for this module, but to avoid overdecorated symbols letting the A-module structure be tacitly understood and simply writing M is to prefer in most instances.

☆

4.2 *Direct sums and direct products*

There are two important and closely related constructions one can make in the category Mod_A of A-modules, namely the direct product and the direct sum. There is no restriction on the cardinality of the involved families, but during practical work in algebraic geometry or number theory one mostly meets finite families, and in that case the two constructs agree.

Direct products

Direct products of modules (direkte produkter av moduler)

(4.19) In this section we work with a collection $\{M_i\}_{i\in I}$ of modules over the ring A. The underlying abelian group of the *the direct product* $\prod_{i\in I} M_i$ will be the direct product of the abelian groups underlying the M_i's, which should be well known from earlier courses. The elements are *strings* or *tuples* $(m_i)_{i\in I}$ indexed by the set I, and the addition is performed componentwise; *i. e.* $(m_i) + (m_i') = (m_i + m_i')$. In case I is finite, say $I = \{1, \dots, r\}$, an alternative notation for a tuple is $(m_1, \dots, m_r)$. The actions of A on the different M_i's induce an action on the direct product, likewise defined component for component: a ring element a acts like $a \cdot (m_i) = (a \cdot m_i)$. The module axioms in paragraph (4.1) are easily verified component by component, and we have an A-module structure on $\prod_{i\in I} M_i$.

The projections $\pi_i\colon \prod_{i\in I} M_i \to M_i$ are A-linear simply because the module operations in the product are performed componentwise.

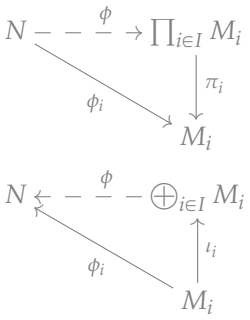
Direct sums

Direct sums of modules (direkte summer av moduler)

(4.20) The *direct sum* of the module collection $\{M_i\}_{i\in I}$ is denoted by $\bigoplus_{i\in I} M_i$ and is defined as the submodule of the direct product consisting of strings $m = (m_i)_{i\in I}$ with all but a finite number of the m_i's vanishing.

When the index set I is finite, requiring strings to merely have finitely many non-zero components imposes no constraint, so in that case the direct sum and the direct product coincide. However, when the index set I is infinite, they are certainly not isomorphic; they are not even of the same cardinality. For instance, the direct sum of countably many copies of $\mathbb{Z}/2\mathbb{Z}$ is countable (being the set of finite sequences of zeros and ones) whereas the direct product of countably many copies of $\mathbb{Z}/2\mathbb{Z}$ has the cardinality of the continuum (the elements my be considered to be 2-adic expansions of real numbers).

Universal properties



(4.21) Both the product and the direct sum are characterised by a *universal properties*. It is noticeable that these properties are dual to each other; reversing all arrows in one, yields the other. For this reason the direct sum is frequently called the *co-product* in the parlance of category theory.

We first describe the universal property the direct product has. The set-up is an A-module N and a collection of A-linear maps $\phi_i\colon N \to M_i$, and the outcome is that there exists a unique A-linear map $\phi\colon N \to \prod_{i\in I} M_i$ such that $\pi_i \circ \phi = \phi_i$. Indeed, this amounts to the map $\phi(n) = (\phi_i(n))_{i\in I}$ being A-linear.

In the case of the direct sum the universal property does not involve the projections, but rather the natural inclusions $\iota_j\colon M_j \to \bigoplus_{i\in I} M_i$ that send an $m \in M_j$ to the string having all entries equal to zero but the one in slot j which equals m. The given maps are maps $\phi_i\colon M_i \to N$, and the conclusion is that there exists a unique map $\phi\colon \bigoplus_{i\in I} M_i \to N$ so that $\phi \circ \iota_j = \phi_j$. The map ϕ is compelled to be defined as

$$\phi\big((m_i)\big) = \sum_{i\in I} \phi_i(m_i),$$

and this is a legitimate definition since merely finitely many of the m_i's are non-zero.

EXERCISE 4.8 Work out all the details in the above reasoning. ★

(4.22) With the stage rigged as in the previous paragraphs we round off the discussion of the universal properties of direct products and direct sums by offering equivalent formulations in terms of the hom-modules:

PROPOSITION 4.23 *There are canonical isomorphisms*

i) $\operatorname{Hom}_A(\bigoplus_{i\in I} M_i, N) \simeq \prod_{i\in I} \operatorname{Hom}_A(M_i, N)$;

ii) $\mathrm{Hom}_A(N, \prod_{i\in I} M_i) \simeq \prod_{i\in I} \mathrm{Hom}_A(N, M_i)$.

Notice that in the first isomorphism, which involves the contravariant slot, the direct sum is transformed into a direct product. It further warrants a special comment that when the index set is finite, the direct product coincides with the direct sum, and the proposition may be summarized by saying that the hom-functor commutes with finite direct sums. In the vernacular of category theory one says that it is *additive* in both variables.

EXERCISE 4.9 Figure out the precise definitions of the isomorphisms in Proposition 4.23 above. HINT: The key word is universal properties. ★

(4.24) We shall identify each module M_j with the image $\iota_j(M_j)$ in $\bigoplus_{i\in I} M_i$ under the natural inclusion ι_j; that is, with the submodule of elements having all entries zero except in slot j.

Fix one of the indices, say ν. Forgetting the ν-th entry in string $(m_i)_{i\in I}$ gives a string $(m_i)_{i\in I\setminus\{\nu\}}$ indexed by the subset $I\setminus\{\nu\}$ of indices different from ν. The operations in direct sums being performed component-wise, this is clearly an A-linear assignment; hence it gives an A-linear map

$$\bigoplus_{i\in I} M_i \longrightarrow \bigoplus_{i\in I\setminus\{\nu\}} M_i \ .$$

The kernel is obviously equal to M_ν (identified with the submodule of the direct sum where merely the ν-th entry is non-zero), and the Isomorphism Theorem (Theorem 4.12 on page 90) yields an isomorphism

$$\Big(\bigoplus_{i\in I} M_i\Big) / M_\nu \simeq \bigoplus_{i\in I\setminus\{\nu\}} M_i.$$

The slogan is: Killing one addend of a direct sum yields the sum of the others.

EXERCISE 4.10 Generalize the slogan above to any sub-collection: Let $J \subseteq I$ be a subset. Prove that there is a canonical isomorphism

$$\bigoplus_{i\in I} M_i / \bigoplus_{j\in J} M_J \simeq \bigoplus_{i\in I\setminus J} M_i,$$

and that there is a corresponding isomorphism for direct products:

$$\prod_{i\in I} M_i / \prod_{j\in J} M_J \simeq \prod_{i\in I\setminus J} M_i.$$

★

Split submodules, direct sums and idempotent maps

It is well known from linear algebra that every sub-vector space is a direct summand in the surrounding space; bases of the subspace can be extended to the entire containing space. This stands in contrast to submodules of modules over general rings, most of which are not direct summands. It is therefore of interest to have criteria for a submodule to be a direct summand of the surrounding module.

(4.25) A synonym for a submodule $N \subseteq M$ to be a direct summand, is that N *lies split* in M—this of course means that there is another submodule N' so that $M \simeq N \oplus N'$— and just as in linear algebra, the submodule N' is called a *complement* to N. Equivalently, every element m from M can be unambiguously expressed as a sum $m = n + n'$ with $n \in N$ and $n' \in N$; or phrased differently, the two conditions $N \cap N' = 0$ and $N + N' = M$ are fulfilled. With a slightly sloppy notation, one usually writes $M = N \oplus N'$.

Split submodules (splitt undermoduler)

Complementary submodules (komplementære undermoduler)

(4.26) When we treated direct products of other rings, the notion of idempotent elements turned out to be quite useful. This notion can be generalizations in several directions and in various contexts, the virtue of idempotents always being that they express some kind of "direct decomposition". In our present context of modules over a ring A an A-linear map $\epsilon\colon M \to M$ is said to be *idempotent* if $\epsilon^2 = \mathrm{id}_M$.

Idempotent endomorphisms (idempotente endomorfier)

PROPOSITION 4.27 (IDEMPOTENTS AND DIRECT SUMS) *Let M be an A-module.*

i) If ϵ is an idempotent endomorphism of M, then M decomposes as the direct sum $M = \ker\epsilon \oplus \operatorname{im}\epsilon$;

ii) When ϵ is an idempotent endomorphism of M, one has $\operatorname{im}\epsilon = \epsilon M$ and $\ker\epsilon = (\mathrm{id}_M - \epsilon)M$;

iii) A submodule N of M lies split if and only if there is an idempotent endomorphism $\epsilon\colon M \to M$ with $\operatorname{im}\epsilon = N$.

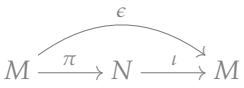
Note that $\mathrm{id}_M - \epsilon$ is idempotent precisely when ϵ is, so the two appear in a completely symmetric way in the proposition.

PROOF: Suppose to begin with that ϵ is an idempotent endomorphism of M. We contend that $M = \operatorname{im}\epsilon \oplus \ker\epsilon$. Indeed, it holds true that $x = (x - \epsilon(x)) + \epsilon(x)$. Obviously $\epsilon(x)$ lies in $\operatorname{im}\epsilon$ and $x - \epsilon(x)$ lies in the kernel $\ker\epsilon$ because ϵ is idempotent:

$$\epsilon(x - \epsilon(x)) = \epsilon(x) - \epsilon^2(x) = \epsilon(x) - \epsilon(x) = 0.$$

On the other hand, $\operatorname{im}\epsilon \cap \ker\epsilon = 0$ since if $x = \epsilon(y)$ lies in $\ker\epsilon$, it holds that $0 = \epsilon(x) = \epsilon^2(y) = \epsilon(y) = x$. This takes care of *i)*. As to *ii)*, we only need to add that if $\epsilon(x) = 0$, clearly $x - \epsilon(x) = x$ so that $\ker\epsilon \subseteq (\mathrm{id}_M - \epsilon)M$, the other inclusion already being taken care of.

Finally we attack *iii)* and suppose that $M = N \oplus N'$. Let $\pi\colon M \to N$ and $\iota\colon N \to M$ be respectively the projection and the inclusion map. Then $\pi \circ \iota = \mathrm{id}_N$. Putting $\epsilon = \iota \circ \pi$



we find

$$\epsilon^2 = (\iota \circ \pi) \circ (\iota \circ \pi) = \iota \circ (\pi \circ \iota) \circ \pi = \iota \circ \pi = \epsilon,$$

and it follows readily that $\ker \epsilon = N$ and $\operatorname{im} \epsilon = M$. ❑

Examples

(4.12) A principal ideal (n) i $\mathbb{Z}$, with n neither being zero nor plus-minus one, is not a direct summand of $\mathbb{Z}$ since every other ideal contains multiples of n.

(4.13) Cheap but omnipresent examples of non-split submodules are ideals $\mathfrak{a}$ in domains A. By Paragraph 4.24 above, any complement of $\mathfrak{a}$ would be isomorphic to $A/\mathfrak{a}$. If $\mathfrak{a}$ is a non-zero proper ideal, the quotient $A/\mathfrak{a}$ contains non-zero elements killed by $\mathfrak{a}$, which is absurd since A was assumed to be a domain.

(4.14) A ring that is not a domain, may possess non-zero proper ideals lying split. The simples example is the direct product $A = k \times k$ of two fields. The subspaces $k \times (0)$ and $(0) \times k$ are both ideals.

☆

Exercises

(4.11) Let M_1 and M_2 be two submodules of the A-module M whose intersection vanishes; that is, $M_1 \cap M_2 = (0)$. Prove that $M_1 + M_2$ is naturally isomorphic with the direct sum $M_1 \oplus M_2$. HINT: Establish that any $m \in M_1 + M_2$ can be expressed as $m = m_1 + m_2$ with m_1 and m_2 unambiguously defined elements in respectively M_1 and M_2.

(4.12) Let $\{M_i\}_{i \in I}$ be a family of submodules of the A-module N. Assume that they comply to the following rule: For any index $\nu \in I$ and any finite subset $J \subseteq I$ not containing ν, the intersection of M_ν and $\sum_{j \in J} M_j$ vanishes; that is, $M_\nu \cap \sum_{i \in J} M_j = (0)$. Prove that $\sum_{i \in I} M_i$ is isomorphic with the direct sum $\bigoplus_{i \in I} M_i$.

(4.13) Assume that for each $i \in I$ there is given a submodule $N_i \subseteq M_i$. Prove that $\bigoplus_{i \in I} N_i$ is a submodule of $\bigoplus_{i \in I} M_i$ in a natural way and that there is a natural isomorphism $\bigoplus_{i \in I} M_i/N_i \simeq (\bigoplus_{i \in I} M_i)/(\bigoplus_{i \in I} N_i)$.

(4.14) Generalize Proposition 4.27 in the following way. Assume that $\epsilon_1, \ldots, \epsilon_r$ are mutually orthogonal idempotent endomorphisms of the A-module M. Suppose they satisfy $\sum \epsilon_i = \mathrm{id}_M$. Show that putting $M_i = \epsilon_i(M)$ one obtains a decomposition $M = \bigoplus_i M_i$. Prove the converse: If such a decomposition exists, exhibit a collection of idempotents inducing it.

★

Modules over direct products

We aim at describing all modules over a direct product $A = A_1 \times \ldots \times A_r$ of a finite collection $\{A_i\}_{1 \leqslant i \leqslant r}$ of rings in terms of modules over the factors A_i. The description

is based on a very natural construction: Suppose given an A_i-module M_i for each i. The additive group $\bigoplus_i M_i$ has a natural A-module structure; a string $a = (a_i)_{i\in I}$ of ring elements acts on a string $m = (m_i)_{i\in I}$ of module elements according to the rule $a \cdot m = (a_i \cdot m_i)$, and once more the axioms come for free, the action being defined component-wise. We contend that all A-modules are shaped like this.

PROPOSITION 4.28 *Let $A_1, \ldots, A_r$ be rings and put $A = A_1 \times \ldots \times A_r$. Assume that M is an A-module. Then there are canonically defined A-submodules M_i of M that are A_i-modules, and are such that $M \simeq \bigoplus_i M_i$.*

PROOF: The point is that a decomposition of 1 as a sum of orthogonal idempotents in A induces a decomposition of M. To be precise, let $e_1, \ldots, e_r$ be the idempotents $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 located in slot i. Let $\mathfrak{a}_i$ be the kernel of the projection $A \to A_i$; that is, $\mathfrak{a}_i$ is the ideal generated by the idempotents e_j other than e_i. The set $M_i = e_i M$ is an A-submodule of M killed by $\mathfrak{a}_i$; hence it is an A_i-module. From $\sum_i e_i = 1$ we infer that $M = \sum_i M_i$: It holds that $x = \sum_i xe_i$, and the sum is direct since if $x = \sum_i m_i e_i$, it follows readily from the e_i's being orthogonal that $m_i e_i = xe_i$; thence the terms $m_i e_i$ depend unambiguously on x. ❑

(4.29) Let us take closer look at the case when A is a direct product of finitely many fields; say $A = k_1 \times \ldots \times k_r$. Then Proposition 4.28 above tells us that all modules over A are shaped like direct sums $V_1 \oplus \ldots \oplus V_r$ with each V_i a vector space over k_i.

EXERCISE 4.15 Extend Proposition 4.28 to an arbitrary direct product $A = \prod_{i\in I} A_i$: Prove that any A module M is isomorphic to a product $\prod_{i\in I} M_i$ where each M_i is an A_i-module unambiguously associated with M. ★

4.3 *Finitely generated modules*

(4.30) The *finitely generated modules* form a particularly important class of modules. For short, they are also called *finite modules*. As the name indicates, these modules have finite sets of generators; that is, sets of elements $m_1, \ldots, m_r$ such that each $m \in M$ can be expressed as a linear combination $m = \sum_i a_i m_i$ of the m_i's with coefficients a_i's from A. The generators are by no means unique, and in general there might very well be non-trivial linear relations among them. The subcategory of the category Mod_A of A-modules whose objects are the finitely generated A-modules and morphisms are the A-linear maps* will be denoted by mod_A.

Finitely generated modules (endeliggenererte moduler)
Finite modules (endelige moduler)

The maps in mod_A bewteen two objects in mod_A are thus the same as those in Mod_A. One says that mod_A is a full subcategory of Mod_A

EXAMPLE 4.15 Several important and natural occurring modules are not finitely generated. One example can be the $\mathbb{Z}$-module $B = \mathbb{Z}[p^{-1}]$ where p is a natural number. For each non-negative integer i consider the submodule $B_i = \mathbb{Z} \cdot p^{-i}$. Because $p^{-i} = p \cdot p^{-i-1}$, it holds true that $B_i \subseteq B_{i+1}$, and the B_i's form an ascending chain of submodules. Now, every element in B is of the form $a \cdot p^{-n}$ for some n, in other

words one has $B = \bigcup_i B_i$. Any finite set of elements from B is contained in B_N for some N sufficiently large (just take N larger than all exponents of p appearing in the denominators), so if finitely many elements generated B, it would hold true that $B_N = B$ for some N. This is obviously absurd, as p can appear to any power in the denominator. ☆

Cyclic modules

Cyclic modules (sykliske moduler)

Monogenic modules (monogene moduler)

(4.31) Modules requiring only a single generator are said to be *cyclic* or *monogenic*. Among the ideals the principal ideals are precisely the cyclic ones, and more generally, if M is any module and $m \in M$ an element*, the submodule $A \cdot m = \{ a \cdot m \mid a \in A \}$ is cyclic.

*The zero module is counted among the cyclic ones, so $m = 0$ is admitted.

Now, assume that M is a cyclic A-module and let $m \in M$ be a generator. Multiplication induces an A-linear map $\phi\colon A \to M$ that sends a to am, and this map is surjective since m was chosen to be a generator. The kernel of ϕ consists by definition of those a's that kill m, or which amounts to the same, that kill M. Hence $\ker \phi = \operatorname{Ann} M$, and by Corollary 4.12 on page 90, we arrive at an isomorphism $M \simeq A/\operatorname{Ann} M$.

LEMMA 4.32 *A cyclic A-module M is isomorphic to $A/\operatorname{Ann} M$.*

So the cyclic modules are up to isomorphism precisely the quotients $A/\mathfrak{a}$ of A by ideals $\mathfrak{a}$. Notice that A itself is cyclic corresponding to $\mathfrak{a} = 0$. The ideal $\mathfrak{a}$ is of course uniquely determined by the isomorphism class of M as an A-module (it equals the annihilator $\operatorname{Ann} M$ of M), but different ideals may give rise to quotient that are isomorphic as rings (but of course not as modules). For instance, the quotients $\mathbb{C}[x]/(x-a)$ with $a \in \mathbb{C}$ are all isomorphic to $\mathbb{C}$.

The name cyclic is inherited from the theory of groups; the cyclic groups being those generated by a single element; in other words, those shaped like $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}$.

Simple modules

Simple modules (Simple moduler)

(4.33) The simplest modules one can envisage are the ones without other submodules than the two all modules have—the zero submodule and the module itself—and they are simply called simple: A non-zero* A-module M is said to be *simple* if it has no non-zero proper submodules. Simple modules are cyclic, and any non-zero element generates; indeed, if $m \neq 0$, the cyclic submodule $A \cdot m$ is non-zero (as m lies in it) and hence equals M since M has no proper non-zero submodules. Lemma 4.32 above gives that M is of the form $A/\operatorname{Ann} M$. Moreover, the annihilator ideal $\operatorname{Ann} M$ must be maximal since if $\operatorname{Ann} M \subseteq \mathfrak{a}$, the quotient $\mathfrak{a}/\operatorname{Ann} M$ is a submodule of M which either equals 0 or M. In the former case $\operatorname{Ann} M = \mathfrak{a}$, and in the latter it holds that $\mathfrak{a} = A$. Thus simple A-modules are characterized as follows:

*It might appear paradoxical that the simplest of all modules is not counted among the simple ones; the reason is found in the upcoming Proposition 4.34.

PROPOSITION 4.34 *An A-module M is simple if and only if it is cyclic and its annihilator* $\operatorname{Ann} M$ *is a maximal ideal; i. e. M is simple if and only if M is isomorphic to $A/\mathfrak{m}$ for some maximal ideal $\mathfrak{m}$.*

Exercises

(4.16) Let N be a submodule of M. Show that if N and M/N both are finitely generated, then M is finitely generated as well. Give an example of modules M and N so that M and M/N are finitely generated but N is not.

(4.17) Let N and L be submodules of the A-module M. If $N \cap L$ and $N + L$ are finitely generated, show that both N and L are finitely generated.

(4.18) Show that $k[x, x^{-1}]$ is not a finitely generated module over $k[x]$.

(4.19) Assume that k is a field. Consider the polynomial ring $k[x]$ as a module over the subring $k[x^2, x^3]$. Prove it is finitely generated by exhibiting a set of generators. Determine the annihilator of the quotient $k[x]/k[x^2, x^3]$. What can you say about $k[x]/k[x^2, x^p]$ where p is an odd prime?

(4.20) Assume that k is a field. Consider the polynomial ring $k[x]$ as a module over the subring $k[x^3, x^7]$. Prove it is finitely generated by exhibiting a set of generators. Determine the annihilator of the quotient $k[x]/k[x^3, x^5]$.

(4.21) *Schur's lemma.* Assume that M and N are two simple A-module that are not isomorphic. Prove that $\operatorname{Hom}_A(M, N) = 0$. Prove that $\operatorname{Hom}_A(M, M) = A/\operatorname{Ann} M$.

★

4.4 *Bases and free modules*

Bases for modules (basis for en modul)

Just like for vector spaces one says that a set of generators $\{m_i\}_{i \in I}$ (not necessarily finite) is a *basis* for M if every element from M can be written as a (finite) linear combination of the m_i's in only one way; that is, the coefficients a_i in an expression $m = \sum_i a_i m_i$ are unambiguously determined by m. Be aware, however, that unlike what is the case for vector spaces, most modules do not have a basis.

EXAMPLE 4.16 The two elements x and y generate the ideal (x, y) in the polynomial ring $k[x, y]$, but do not form a basis since the element xy can be expressed as two different linear combinations, namely* one has $xy = x \cdot y = y \cdot x$. And of course, x and y form a minimal set of generators, one can not do without either, so even minimal generator sets are not necessarily bases. The natural question then arises: Can the ideal (x, y) be generated by one element? The answer is no! A generator would divide both x and y which is absurd.

*In the first expression x is the coefficient and y the generator, where as in the second it is the other way around; y is the coefficient and x the generator.

The gist of this example is that x and y commute, which indicates that the phenomenon is inherent in commutative rings. Any set of generators for an ideal consisting

of at least two elements can never be a basis simply because the generators commute. ☆

EXERCISE 4.22 Show that the property, familiar from the theory of vector spaces, that $\sum_i a_i m_i = 0$ implies that $a_i = 0$ is sufficient for a generating set $m_1, \dots, m_r$ to be a basis. HINT: Consider the difference of two equal linear combinations of the m_i's. ★

Free modules

Free modules (fri moduler)

(4.35) The lack of bases for most modules leads to a special status of those that have one. One says that an A-module F is *free* if it has a basis. The reason behind the suggestive name "free" is that one may freely prescribe values to linear map on the basis elements—a principle that goes under the name of the *Universal Mapping Principle*:

PROPOSITION 4.36 (THE UNIVERSAL MAPPING PRINCIPLE) *Suppose that a given A-module F is free and has a basis $\{f_i\}_{i\in I}$, and let M be another A-module. For any subset $\{m_i\}_{i\in I}$ of M indexed by I, there is a unique A-linear map $\phi\colon F \to M$ such that $\phi(f_i) = m_i$.*

PROOF: Every element $x \in F$ is expressible as $x = \sum_{i\in I} a_i f_i$ with coefficients a_i from A, merely finitely many of which are non-zero, and most importantly, the a_i's are uniquely determined by x. Hence sending x to $\sum_{i\in I} a_i m_i$ gives a well defined map $\phi\colon F \to M$. That this yields an A-linear map amounts to the coefficients of a linear combination being the corresponding linear combination of the coefficients, which ensues from coefficients being unique. ❑

(4.37) We return to Example 4.16 on the preceding page about the question when ideals are free, to give a precise statement:

PROPOSITION 4.38 *An ideal $\mathfrak{a}$ in the ring A is a free A-module if and only if it is principal and generated by a non-zero divisor.*

PROOF: We saw in Example 4.16 on the previous page that when $\mathfrak{a}$ requires at least two generators, it has no basis and therefore is not free. Nor can principal ideals generated by a zero divisor be free since if $a \cdot f = 0$ with $a \neq 0$, the relations $a \cdot f = 0 \cdot f = 0$ give two representations of 0. The other way around, if the non-zero divisor f is a generator for $\mathfrak{a}$, it is a basis; indeed f being a non-zero divisor it can be cancelled from an equality like $af = bf$. ❑

EXAMPLE 4.17 Another kind of non-free modules ubiquitously present in algebra are the *the torsion modules*; among them we find the cyclic modules of the form $A/\mathfrak{a}$ where $\mathfrak{a}$ is a non-zero ideal. Since $a \cdot 1 = 0 \cdot 1$ for any $a \in \mathfrak{a}$, such a module can not be free: Any map $\phi\colon A/\mathfrak{a} \to M$ must have image in the submodule $(0 : \mathfrak{a})_M$ consisting of elements killed by $\mathfrak{a}$ which violates the Universal Mapping Principle (Proposition 4.36 above).

If the ring A is a PID, all non-zero ideals will be free modules, but of course there will still be torsion modules. However, these are, at least among the finitely generated modules, the ones that prohibits modules from being free, since any finitely generated

A-module is isomorphic to a finite direct sum $\bigoplus_i M_i$ with the M_i's either being A or $A/(a_i)$ for some $a_i \in A$. This nice behaviour does not persist for modules that are not finitely generated, as Exercise 4.27 on page 103 below shows. ☆

(4.39) Archetypes of free modules are the direct sums $nA = A \oplus \ldots \oplus A$ of n copies of the ring A which we already met in Example 4.2 on page 4.2. They come equipped with the so-called *standard basis* familiar from courses in linear algebra. The basis elements e_i are given as $e_i = (0, \ldots, 0, 1, 0, \ldots 0)$ with the one sitting in slot number i.

There is no reason to confine these considerations to direct sums of finitely many copies of A. For any set I, the direct sum $\bigoplus_{i\in I} A$ has a *standard basis* $\{e_i\}_{i\in I}$ and is a free module; the basis element e_i is the string with a one in slot i and zeros everywhere else.

PROPOSITION 4.40 *Assume that F is a free A-module with basis $\{f_i\}_{i\in I}$. Then there is an isomorphism between F and the direct sum $\bigoplus_{i\in I} A$ that sends each basis vector f_i to the standard basis vector e_i.*

PROOF: By Proposition 4.36 above, we may define a map $\phi\colon F \to \bigoplus_{i\in I} A$ by sending f_i to the standard basis vector e_i; conversely, since $\bigoplus_{i\in I} A$ is free, sending e_i to f_i sets up a map $\psi\colon \bigoplus_{i\in I} A \to F$. These two maps are obviously mutual inverses. ❑

COROLLARY 4.41 *Any two bases of a free module have the same cardinality. Two free modules are isomorphic if and only if they possess bases of the same cardinality.*

Rank of a free module (Rangen til en fri modul)

The common cardinality of the bases for a free module is called the *rank* of the module. The rank is the sole invariant of free modules; up to isomorphism it determines the module. When the module is a vector space over a field and the rank is finite, the rank is just the dimension of the vector space.

PROOF: After Proposition 4.40 above we need merely to verify that when two direct sums $\bigoplus_{i\in I} A$ and $\bigoplus_{j\in J} A$ are isomorphic as A-module, the index sets I and J are of the same cardinality. This is well known from the theory of vector spaces, and we reduce the proof to the case that A is a field, so take any maximal ideal in A and consider the isomorphic vector spaces $\bigoplus_{i\in I} A/\mathfrak{m}$ and $\bigoplus_{j\in J} A/\mathfrak{m}$ over $A/\mathfrak{m}$ (isomorphic in view of Exercise 4.13 on page 96). One has a basis of the same cardinality as I, the other one of cardinality that of J; hence I and J are equipotent. ❑

EXAMPLE 4.18 *Free modules with given basis:* From time to time itis convenient to operate with free A-modules with a given set S as basis. There is no constraint on the set S, it can be whatever one finds useful. The formal way to construct such a module, denoted A^S, is as the set of maps $\alpha\colon S \to A$ with finite support; that is, the maps such that $\alpha(s) \neq 0$ for at most finitely many members s of S; in symbols

$$A^S = \{\, \alpha\colon S \to A \mid \alpha \text{ of finite support } \}.$$

The module structure of A^S is given point-wise: $(\alpha+\alpha')(s) = \alpha(s)+\alpha'(s)$ and $(a\cdot\alpha)(s) = a\alpha(s)$.

In A^S there is a collection of function termed *generalized Kronecker-δ's* that constitute a natural basis and which are in a one-to-one correspondence with the set S. For each member $s \in S$ there is one such function δ_s defined as

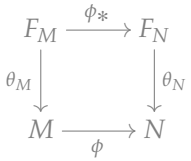
$$\delta_s(t) = \begin{cases} 0 & \text{when } s \neq t \\ 1 & \text{when } t = s\,. \end{cases}$$

It is a trivial matter to verify they form a basis. They generate A^S because any α can be expressed as $\alpha = \sum_{s\in S}\alpha(s)\delta_s$, and if $\sum a_s\delta_s = 0$ is a dependence relation, one just plugs in any t from S to find that $a_t = 0$.

A suggestive way of denoting elements from A^S is as linear combinations $\sum_s a_s\cdot s$ of elements from S, which merely amounts to writing s for the function δ_s.

The module A^S depends functorially on S. Indeed, given any map $\phi\colon S\to S'$. Because the δ_s's form a basis for A^S, we obtain according to the Universal Mapping Principle for free modules, a map $\phi_*\colon A^S\to A^{S'}$ by sending each basis element δ_s to the element $\delta_{\phi(s)}$ of $A^{S'}$. In the alternative notation, the map ϕ_* takes the form $\phi_*(\sum_s a_s\cdot s) = \sum_s a_s\cdot\phi(s)$. It is pretty obvious that $(\psi\circ\phi)_* = \psi_*\circ\phi_*$ when ψ is another map composable with ϕ, so that A^S is a covariant functor from the category Sets to the category Mod_A. ☆

EXERCISE 4.23 Let M be an A-module and let $F_M = A^M$ be the free module with elements from M as a basis. Then M is a quotient of F_M in a canonical way: Define a map $\theta_M\colon F_M\to M$ by sending δ_m to m (in the alternative notation it takes the hypertautological form $m\mapsto m$). Show that θ_M is surjective and that $\phi\circ\theta_M = \theta_N\circ\phi_*$ whenever $\phi\colon M\to N$ is an A-linear map. ★



EXERCISE 4.24 Show that a finitely generated A-module is the quotient of a finite free module. ★

Matrices and maps of free modules

Just like linear maps between two vector spaces of finite dimension, A-linear maps between two *free* A-modules can be described by matrices, and the mechanism works exactly in the same way. Be aware however, that describing maps between modules that are not free, is substantially more complicated, if a good description at all is possible.

(4.42) The representation of a map by a matrix depends on the choice of bases for each module. So let E and F two finitely generated and free A-module and let $\{e_j\}$ and $\{f_i\}$ be bases for E and F respectively. When each $\phi(e_j)$ is expressed in terms of the basis $\{f_i\}$, the coefficients in the expressions make up the matrix. If E is of rank n and F of rank m the matrix is the $m\times n$-matrix given* as $M(\phi) = (a_{ij})$ where $\phi(e_j) = \sum_i a_{ij}f_i$.

(4.43) The familiar property that compositions of maps correspond to products of

*The precise notation $M^{\mathcal{E}}_{\mathcal{F}}(\phi)$, where $\mathcal{E}$ and $\mathcal{F}$ designate the two bases, is cumbersome, but sometimes helpful in that the rule for composition takes the form $M^{\mathcal{E}}_{\mathcal{F}}(\phi)\cdot M^{\mathcal{F}}_{\mathcal{G}}(\psi) = M^{\mathcal{E}}_{\mathcal{G}}(\phi\circ\psi)$

matrices still holds true, and the verification is *mutatis mutandis* the same as for linear maps between vector spaces (we leave it to students needing to fresh up their knowledge of linear algebra); that is, if ψ is A-linear map from F to a third free module G (equipped with a basis), one has

$$M(\psi \circ \phi) = M(\psi) \cdot M(\phi).$$

Likewise, associating a matrix to a map persists being a linear operation in that

$$M(\alpha\phi + \beta\phi') = \alpha M(\phi) + \beta M(\phi'),$$

whenever ϕ and ϕ' are A-linear maps from E to F and α and β are two elements from the ring A.

Exercises

The next series of exercises, which culminates with problem 4.27, is aimed at giving an example that countable products of free modules are not necessarily free.

(4.25) Show that in a free $\mathbb{Z}$-module every element is divisible* by at most finitely many integers.

*One says that an element x of a module M is divisible by n if there is a $y \in M$ so that $ny = x$.

(4.26) Show that the direct sum of countably many copies of $\mathbb{Z}$ is countable, whereas the direct product of countably many copies is not (it has the cardinality of the continuum).

(4.27) *Infinite products are not free.* The task is to show that the direct product $P = \prod_{i\in\mathbb{N}} \mathbb{Z}$ of a countable number of copies of $\mathbb{Z}$ is not a free $\mathbb{Z}$-module. The crucial point is to show that if P were free, the direct sum $\bigoplus_{i\in\mathbb{N}} \mathbb{Z}$ would be contained in a proper direct summand Q of P. The quotient P/Q would then be free which is absurd since it has infinitely divisible elements.

Aiming for a contradiction, we suppose that the product has a basis $\{f_i\}_{i\in I}$. The direct sum $\bigoplus_{i\in\mathbb{N}} \mathbb{Z}$ lies in P and has the standard basis elements e_i (with a one in slot i as sole non-zero component). Each e_i can be developed as a finite sum $e_i = \sum_j a_{ij} f_j$ in terms of the basis elements f_j with coefficients $a_{ij} \in \mathbb{Z}$.

a) Prove that I cannot be countable (cfr. Exercise 4.26).

b) Prove that there is a countable subset $J \subseteq I$ so that the module Q generated by the f_j's with $j \in J$ contains the direct sum $\bigoplus_{i\in\mathbb{N}} \mathbb{Z}$. Conclude that Q is a proper direct summand in P. HINT: Let j be in J when the coefficient $a_{ij} \neq 0$ for at least one i. Observe that for each i it holds that $a_{ij} \neq 0$ only for finitely many j.

c) For any element $x = (n_1, n_2, \ldots)$ in P and any $i \in \mathbb{N}$ prove that the element $y = (0, 0, \ldots, 0, n_i, n_{i+1}, \ldots)$ has the same image in P/Q as x.

d) Show that there are strictly increasing sequences $\{n_k\}$ of natural numbers with $n_k | n_{k+1}$ and so that $a = (n_1, n_2, \ldots)$ does not lie in Q. HINT: Q is countable.

e) Show that the image of a in P/Q is divisible by infinitely many numbers and hence P/Q cannot be free.

4.5 Graded modules

Graded modules (graderte moduler)

(4.44) Let $A = \bigoplus_{i\in\mathbb{Z}} A_i$ be a graded ring. A *graded module* over A is a module whose underlying additive group decomposes as $M = \bigoplus_{i\in\mathbb{Z}} M_i$ in way compatible with the action of A on M; that is, the following condition is satisfied

$$A_i M_j \subseteq M_{i+j}$$

Homogeneous maps of degree zero (homogene avbildninger av grad null)

for all i and j. Note that each homogenous component M_j will be a module over A_0.

It turns out to be important to allow elements of negative degree, and as long as the degrees are bounded away from $-\infty$, this does not pose serious problem; we say that M is*bounded from below* if $M_i = 0$ for $i << 0$.

(4.45) As always, a new concept is followed by the concept of the corresponding "morphisms" preserving the new structure. In the present case a "morphism" between two graded A-modules M and M' is an A-homomorphism $\phi\colon M \to M'$ that respects the grading; it sends homogeneous elements to homogenous elements of the same degree. It is common usage to say that such a homomorphism ϕ is a *homogeneous of degree zero*, or a *homomorphism of graded modules*, or just a *map of graded modules*. It may be decomposed as a sum $\phi = \sum_i \phi_i$ where each $\phi_i\colon M_i \to M'_i$ is an A_0-linear map. And as usual, two graded modules are *isomorphic* if there is a homomorphism of graded modules $\phi\colon M \to M'$ that has an inverse.

Isomorphic graded modules (isomorfe graderte moduler)

Homomorphisms of graded modules (homomorfier av graderte moduler)

(4.46) The composition of two maps of degree zero is obviously of degree zero, as is any linear combination of two. The three identities $\ker\phi = \bigoplus_i \ker\phi_i$, $\operatorname{coker}\phi = \bigoplus_i \operatorname{coker}\phi_i$ and $\operatorname{im}\phi = \bigoplus_i \operatorname{im}\phi_i$ are close to trivial to verify, and hence the kernels, cokernels and imegas of maps of graded modules are graded in a canonical manner. One leisurely verifies that these are kernels, cokernel and images also in the category of graded A-modules; moreover the requirement that "the kernel of the cokernel equals the cokernel of the kernel" is fulfilled (maps have images) so that GrMod_A is an *abelian category*. A sequence in GrMod_A is exact if and only if it is exact in $\mathsf{Mod}\, A$.

Exercise 4.28 Show that if $\{M_i\}_{i\in I}$ is a collection of graded modules, the sum $\bigoplus_{i\in I} M_i$ and the product $\prod_{i\in I} M_i$ are graded in a natural way. Show that with this grading they are the sum and the product also in the graded category GrMod_A. ★

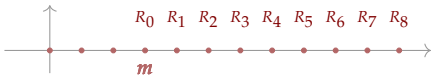
Shift operators

(4.47) There is a collection of *shift operators* acting on the category of graded S-modules. For each graded module M and each integer $m \in \mathbb{Z}$ there is graded module $M(m)$ associated to a graded module M. The shift do not alter the module structure of M, not even the sets of homogeneous elements is affected, but they are given new degrees. The new degrees are defined by setting

$$M(m)_d = M_{m+d}.$$

In other words, one declares the degree of the elements in M_m to be equal to $d - m$. Any map $\phi\colon M \to N$ between two graded modules which is homogeneous of degree zero, persists being so when the degrees are shifted uniformly in M and N, hence it induces a map $\phi[m]\colon M(m) \to N(m)$. This means that the shifts (m) are functors from GrMod_A to itself. Obviously they are exact and of course $(m) \circ (m') = (m + m')$.

EXAMPLE 4.19 For instance, when $m > 0$, the shifted polynomial ring $A(-m)$ has no elements of degree d when $d < m$, indeed, $A(m)_d = A_{d-m}$, and the ground field k sits as the graded piece of degree m. Whereas the twisted algebra $A(m)$ has non-zero homogeneous elements of degrees down to $-m$ with the ground field sitting as the piece of degree $-m$. ☆



4.6 *Nakayama's lemma*

Nakayama's lemma is a workhorse in commutative algebra, and is applied over and over again. As often is the case with popular courses, it comes in quit a lot of different flavours, and we shall present the ones most frequently met.

One way of viewing this famous result—which is the most basic and in our view the best, and which we shall adopt as our point of departure—is as an extension of the fundamental existence result for maximal ideals (Theorem 2.49 on page 49) to finitely generated modules: Every non-zero finitely generated module has a maximal proper submodule, or what amounts to the same, every non-zero finitely generated module has a simple quotient. Indeed, if N is a submodule of M, the quotient M/N is simple if and only if N is a maximal proper submodule (simply because submodules of M/N correspond to submodules of M containing N). Notice, that formulated in this way, Nakayama's lemma comes for free for the large an important class of Noetherian modules (which not yet have been defined, but will be in Chapter 9).

One version of Nakayama's lemma is best proved using a localization technique, and is therefore postponed until after localization has been treated.

Tadashi Nakayama (1912–1964) Japanese mathematician

Nakayama's lemma and simple quotients

(4.48) Here comes our first version of Nakayama's lemma:

PROPOSITION 4.49 (NAKAYAMA'S LEMMA I) *Every finitely generated A-module M which is non-zero, has a non-zero simple quotient. In other words, there exists a maximal ideal $\mathfrak{m}$ and an A-linear surjection $M \to A/\mathfrak{m}$. Equivalently, $\mathfrak{m}M$ is a proper submodule of M.*

Not all modules have simple quotients; to find an example we need look no further than to the rationals $\mathbb{Q}$ considered a $\mathbb{Z}$-module. For any ideal $\mathfrak{a}$ in $\mathbb{Z}$ It holds true that $\mathfrak{a}\mathbb{Q} = \mathbb{Q}$, and hence there are no non-zero maps $\mathbb{Q} \to \mathbb{Z}/\mathfrak{a}$. A constituting property of Noetherian modules (which we soon come to) is that every non-empty set of submodules has a maximal member, so in a Noetherian module maximal proper submodules exist almost by definition.

Proof: Assume first that M is cyclic. It is then of the form $A/\mathfrak{a}$ for some proper ideal and thus has $A/\mathfrak{m}$ as a quotient for any maximal ideal $\mathfrak{m}$ containing $\mathfrak{a}$. If M is not cyclic, chose generators $x_1, \dots, x_n$ for M with n minimal and $n \geqslant 2$. The submodule N generated by $x_2, \dots, x_n$ is a proper submodule of M. Consequently M/N is non-zero and cyclic and has a simple quotient by the first part of the proof. ❑

(4.50) We can not resist giving another argument for M having a maximal proper submodule tailored to the same pattern as the proof of the Basic Existence Theorem for ideals (Theorem 2.49 on page 49). If $\{M_i\}$ is an ascending chain of proper submodules and M is finitely generated, the union $\bigcup_i M_i$ is a proper submodule; indeed, the finite number of generators of M would all be contained in an M_i for i large enough and thence $M_i = M$, which is not the case. Zorn's lemma then ensures there is a maximal proper submodule.

Nakayama classic

(4.51) To assure anyone (hopefully there are none) that finds our approach a blasphemous assault on their most cherished tradition, we surely shall include Nakayama classic; and here it comes. Recall that the Jacobson radical of A equals the intersection of all the maximal ideals in A.

Proposition 4.52 (Nakayama classic) *Let $\mathfrak{a}$ be an ideal in A contained in the Jacobson radical of A. Let M be a finitely generated A-module and assume that $\mathfrak{a}M = M$. Then $M = 0$.*

Proof: Assume $M \neq 0$. By Nakayama I (Proposition 4.49 above) there is a maximal ideal $\mathfrak{m}$ such that $\mathfrak{m}M$ is a proper submodule, which is impossible since $\mathfrak{a} \subseteq \mathfrak{m}$ and $\mathfrak{a}M = M$ by assumption. ❑

The by far most common situation when Nakayama's lemma is applied is when A is a local ring. The Jacobson radical then equals the maximal ideal $\mathfrak{m}$, and, when M is finitely generated, an equality $\mathfrak{m}M = M$ implies that $M = 0$.

One may rephrase Nakayama's lemma as follows.

Proposition 4.53 (Nakayama's lemma II) *Assume that M is a finitely generated A-module and that $\mathfrak{a}$ is an ideal contained in the Jacobson-radical of A. If $M/\mathfrak{a}M = 0$, then $M = 0$.*

Other formulations

(4.54) There are several other reformulations of Nakayama's lemma, and here we offer a few of the most frequently applied ones.

PROPOSITION 4.55 (NAKAYAMA'S LEMMA III) *Let M is be a finitely generated A-module. Assume that $\mathfrak{a}$ is an ideal contained in the Jacobson radical of A and that N a submodule of M such that $N + \mathfrak{a}M = M$. Then $N = M$.*

PROOF: The quotient M/N is finitely generated since M is, and it holds true that $\mathfrak{a} \cdot M/N = M/N$ because any m from M lies in $\mathfrak{a}M$ modulo elements in N. ❑

PROPOSITION 4.56 *Assume that $\phi\colon N \to M$ is A-linear between two A-modules and that M is finitely generated. Moreover, let $\mathfrak{a}$ be an ideal contained in the Jacobson radical of A. If the induced map $\bar{\phi}\colon N/\mathfrak{a}N \to M/\mathfrak{a}M$ is surjective, then ϕ is surjective.*

PROOF: That $\bar{\phi}$ is surjective means that $x \in M$ there is a $y \in N$ such that $x = \phi(y) + z$ with $z \in \mathfrak{a}M$. Hence $M = \phi(N) + \mathfrak{a}M$, and we conclude that $M = \phi(N)$ by the previous proposition. ❑

PROPOSITION 4.57 *Assume that $\mathfrak{a} \subseteq A$ is an ideal contained in the Jacobson-radical of A. Let M be a finitely generated A-module and assume that $\{m_i\}_{i\in I}$ are elements in M whose residue classes generate $M/\mathfrak{a}M$. Then the m_i's generate M.*

PROOF: Let N be the submodule of M generated by the m_i's. The hypothesis that the residue classes generate $M/\mathfrak{a}M$ translates into the statement that $M = N + \mathfrak{a}M$, and the proposition follows from Proposition 4.55. ❑

Exercises

⁂ (4.29) Let Φ be an $n \times n$-matrix with coefficients in a local ring A and denote by $\overline{\Phi}$ the matrix whose entries are the classes of the entries of Φ in the residue class field k of A. Show that if the determinant $\det\overline{\Phi}$ does not vanish, then Φ is invertible. HINT: $\Phi \cdot \Phi^{\dagger} = \det\Phi \cdot I$ and $\det\Phi$ does not belong to the maximal ideal.

⁂ (4.30) *Demystifying Nakayama's lemma.* Let A be a local ring with residue class field k. Assume that $\phi\colon E \to F$ is an A-linear map between free modules of finite rank, and let Φ be the matrix of ϕ in some bases.

a) Show that if one of the maximal minors of $\bar{\Phi}$ does not vanish, one of the maximal minors of Φ is invertible in A. Conclude that Φ is surjective when $\bar{\Phi}$ is.

b) Show the classical Nakayama's lemma for finitely presented modules over a local ring by using the previous subproblem.

c) *(Mystifying the demystification)* Show Nakayama's lemma for finitely generated modules over a local ring by using subproblem *a*). HINT: The key word is

"right sections" of linear maps, if you don't prefer juggling maximal minors of $n \times \infty$-matrices!

(4.31) Let A be a local ring with residue class field k. Let $\phi\colon E \to F$ be a map between finitely generated free A-modules, and suppose that the induced map $\bar{\phi}\colon E/\mathfrak{m}E \to F/\mathfrak{m}F$ is injective. Prove that ϕ is a split injection. HINT: Prove that at least one maximal minor of the matrix of ϕ in some bases is invertible in A. Then the projection $\pi\colon F \to E$ corresponding to that minor furnishes a section.

(4.32) Let M an A-module such that $\mathfrak{m}M = M$ for every maximal ideal $\mathfrak{m}$. Show that M has the property that if one discards any finite part from a generating set one still has a generating set.

(4.33) Let M be a finitely generated A-module and let $\phi\colon M \to M$ be a *surjective* A-linear map. Show that ϕ is injective. Show by exhibiting examples that this is no longer true if M is not finitely generated. HINT: Regard M as a module over the polynomial ring $A[t]$ with t acting on $x \in M$ as $t \cdot x = \phi(x)$. Use the extended version of Nakayama's lemma with $\mathfrak{a} = (t)A[t]$.

(4.34) *Nilpotent Nakayama.* This exercise is about a result related to Nakayama's lemma, but of a much more trivial nature. Let A be a ring M an A-module. Assume that $\mathfrak{a}$ is a nilpotent ideal in A. Show that if $\mathfrak{a}M = M$, then $M = 0$.

(4.35) *Graded Nakayama.* Let $M = \bigoplus_i M_i$ be a graded module over the graded ring $R = \bigoplus_i R_i$. Assume that $M_{-i} = 0$ for i sufficiently big; that is, the degrees of the non-zero homogeneous elements from M are bounded below. Let $\mathfrak{a}$ be a homogenous ideal whose generators are of positive degree. Assume that $\mathfrak{a}M = M$ and show that $M = 0$. HINT: Consider the largest n so that $M_{-n} \neq 0$.

(4.36) Let A be ring and P a finitely generated projective module. Show that there is a set of elements $\{f_i\}$ in A such that the distinguished open subsets $D(f_i)$ cover Spec A, and such that each localized module P_{f_i} is a free module over A_{f_i}.

(4.37) Let A a ring and let e be a non-trivial idempotent element. Show that the principal ideal $I = (e)A$ is projective, and that a direct sum $\bigoplus_i I$ of a any number, finite or not, of copies of I never can be free. HINT: Such sums are killed by $1 - e$.

★

4.7 *Appendix: The determinant and the characteristic polynomial*

Several notions from classical linear algebra, (and from other branches of mathematics, for that matter) are of a universal nature. They have a meaning whatever the ground ring is. Their basic properties are universal too; they persist being true over all rings. The all important determinant is an examples of such a universal creature.

The classical construction of determinants generalizes word for word to A-linear maps between free modules over any ring A, and all the fundamental properties

continue to hold; like multiplicativity, alternation in rows and columns and the rules for expansion along rows or columns. Moreover, the classical proofs still hold water over general (commutative *)rings.

*Except over a few mildly non-commutative rings where there is a kind of substitute, determinants need commutative rings-

The classical approach to determinants is a pedestrian's—there is also a Formula 1 way based on the so-called exterior powers. It has the disadvantage of requiring a rather advanced machinery and being rather opaque for beginners, but has the great advantage of being completely functorial. It also opens up for defining determinants of endomorphisms of a wider class of modules than the free ones.

(4.58) Let $C = (c_{ij})$ be an $n \times n$-matrix with entries from a ring A. Recall that the determinant $\det C$ is defined as the sum

$$\det C = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) c_{1\,\sigma(1)} \cdot \ldots \cdot c_{n\,\sigma(n)} \tag{4.2}$$

where S_n denotes the symmetric groups on n letters and $\operatorname{sgn}(\sigma)$ is the sign of the permutation σ. This is formally the same as definition over a field, and all the usual elementary properties of the determinant persist being valid; *e.g.* linearity in rows and columns, sign change on rows or columns being swapped, and the expression for the determinant developing along a row or a column.

The adjunction formula and the determinant trick

(4.59) As mentioned at the top of the section, the basic properties of the determinant hold true over a general ring, and their classical proofs go through *mutatis mutandis* (the students are encouraged to brush up their knowledge of linear algebra by rereading or, better, reconstruction the proofs). In particular, we would like to point out the *adjunction formula*

Adjunction formula (adjungsjonsformelen)

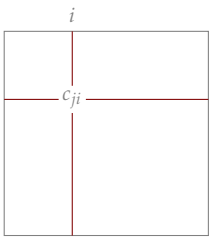
$$C \cdot C^{\dagger} = \det C \cdot I, \tag{4.3}$$

valid for a square $n \times n$-matrix C, where $C^{\dagger}$ is the so-called *cofactor-matrix* of C and I denotes the $n \times n$ identity matrix. The ij-th entry of $C^{\dagger}$ is the sub-determinant of C with the j-th row and i-th column struck out adjusted with the sign $(-1)^{i+j}$. The formula (4.3) follows from (and is in fact equivalent to) the rules for expanding a determinant along a row.

The cofactor-matrix (kofaktormatrisen)

Contrary to the case of vector spaces it does not suffice that the determinant $\det C$ be non-zero for C to be invertible, the determinant must be invertible in A. In that case it ensues from (4.3) that the inverse is given as

$$C^{-1} = (\det C)^{-1} \cdot C^{\dagger}.$$



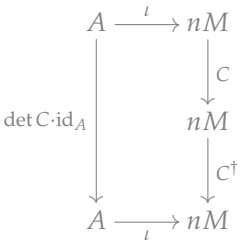
(4.60) The adjunction formula immediately gives that a complex square matrix with a non-trivial kernel has a vanishing determinant. There is a reformulation adapted to

modules of this arch-classical fact called the *determinant trick* (so named by Miles Reid in his book [**?**]).

Let M be an A-module and $C = (c_{ij})$ an $n \times n$-matrix with entries from A. In the same way as C induces an endomorphism of the free module nA, it induces an endomorphism of the iterated direct sum nM; if $m = (m_i) \in nM$, we just let $C \cdot m$ be the n-tuple $(\sum_j c_{ij} m_j)_i$.

LEMMA 4.61 (THE DETERMINANT TRICK) *Let C be an $n \times n$-matrix with entries in the ring A, and let M be an A-module. Assume that the module M has generators $m_1, \dots, m_n$ such that $C \cdot (m_1, \dots, m_n) = 0$. Then the determinant $\det C$ kills M.*

PROOF: Consider the A-linear map $\iota\colon A \to nM$ that sends x to $(x \cdot m_1, \dots, x \cdot m_n)$. The hypothesis of the lemma translates into the relation $C \circ \iota = 0$ (where we confuse the matrix C with the map it induces), and citing the adjunction formula (4.3) we find



$$\det C \cdot \iota = C^\dagger \circ C \circ \iota = 0.$$

This means that $(\det C \cdot m_1, \dots, \det C \cdot m_n) = \det C \cdot \iota(1) = 0$, and hence the determinant $\det C$ kills M as the m_i's generate M. ❏

The characteristic polynomial

(4.62) An endomorphism γ of a finitely generated free A-module E has a canonically defined determinant. Indeed, let C and C' be the matrices of γ in two bases for E. If D denotes the base-change matrix, it holds true that $C' = DCD^{-1}$, and consequently

$$\det C' = \det D \cdot \det C \cdot \det D^{-1} = \det C,$$

which makes $\det \gamma = \det C$ a legitimate definition; the determinant of a matrix representing γ is independent of which basis is used.

The characterstic polynomial of an endomorphism (det karakteristiske polynomet til en endomorfi)

This opens the way for the definition of *the characteristic polynomial* of an endomorphism γ, namely as $P_\gamma(t) = \det(t \cdot \mathrm{id}_E - \gamma)$. It is an element of the polynomial ring $A[t]$.

For any matrix $C = (c_{ij})$ with entries in a ring A and any ring homomorphism $\phi\colon A \to B$ we let $\phi(C) = \big(\phi(c_{ij})\big)$ accepting a slight ambiguity in the notation. The canonical extension of ϕ to a map $A[t] \to B[t]$ between the polynomial rings will as well be denoted by ϕ; that is, $\phi(\sum a_i t^i) = \sum \phi(a_i) t^i$. The following lemma is an almost trivial observation:

LEMMA 4.63 *It holds true that $\phi\big(P_C(t)\big) = P_{\phi(C)}(t)$ where $\phi\colon A \to B$ is any ring homomorphism and C is any square matrix with entries from A.*

PROOF: The determinant is a polynomial in the entries of the matrix, hence it holds true that $\det \phi(D) = \phi(\det D)$ for all ring homomorphisms ϕ and all square matrices D with entries in the source of ϕ. We infer that

$$\phi\big(P_C(t)\big) = \phi(\det(t \cdot I - C) = \det(t \cdot I - \phi\big(C\big)) = P_{\phi(C)}(t).$$

❑

The Cayley–Hamilton theorem and the generic matrix

The Cayley-Hamilton theorem is one of the subtler results from elementary linear algebra. It seems that Frobenius was the first to give a proof in some generality, but much earlier Cayley and Hamilton did the 2×2- and 3×3-cases, which turned out to be sufficient for the theorem to be named after them.

(4.64) The statement involves the characteristic polynomial of a matrix square matrix $C = (c_{ij})$ is any with entries c_{ij} from any (commutative) ring A. Recall that it is given as $P_C(t) = \det(t \cdot I - C)$ where t is a variable and I the identity matrix of the same size as C. The general Cayley–Hamilton theorem reads as follows.

Ferdinand Georg Frobenius (1849–1917)

German mathematician

THEOREM 4.65 (GENERAL CAYLEY–HAMILTON) *Let A be any (commutative) ring and C a square matrix with entries from A. Then C satisfies its characteristic polynomial; in precise terms, if $P_C(t) = \det(t \cdot I - C)$ denotes the characteristic polynomial of C, then $P_C(C) = 0$.*

The characteristic polynomial has the virtue of being canonically associated with the matrix C, and thus it depends functorially on C, in contrast to any arbitrary polynomial equation satisfied by C.

EXAMPLE 4.20 It might be instructive to consider the special case when $A = k$ is a field. If $v \in k^n$ is an eigenvector of C corresponding to the eigenvalue λ in k, then obviously $P_C(C)v = 0$ since $t - \lambda$ is a factor of $P_C(t)$. Therefore the Cayley–Hamilton theorem follows whenever k^n has a basis of eigenvectors of C. In particular it holds true if C has n distinct eigenvalues; *e.g.* if $P_C(t)$ has n distinct roots.

It will even suffice that for some field extension K of k, the space K^n has a basis of eigenvector for C (for instance, this will be the case if $P_C(t)$ has distinct roots in a field extension k). Matrices for which this holds, are said to be *semi-simple*. ☆

Semi-simple matrices (semisimple matriser)

(4.66) There are of course mountains of proofs for such a central result in elementary linear algebra as is the Cayley–Hamilton theorem. The proof we shall offer is very simple. With some knowledge of rudimentary field theory and of polynomials over UFD's, one will find that it almost reduces to bare common sense*. Moreover, it is a low hanging illustration of techniques—specialisation and generalisation— frequently used in modern mathematics.

**But remember Einstein's words: According to common sense, the earth is flat.*

The universal $n \times n$-matrix

The universal matrix (den universelle matrisen)

(4.67) The salient point of the proof we present is that there is a *universal* $n \times n$-matrix $C_n = (x_{ij})$ with entries in a ring R_n, and it is universal in the sense that every other $n \times n$-matrix $C = (c_{ij})$ with entries in any ring A is obtained as $C = (\phi(x_{ij}))$ for an unambiguously defined ring homomorphism $\phi\colon R_n \to A$. There is no hocus-pocus about this; the ring R_n will simply be the polynomial ring $R_n = \mathbb{Z}[x_{ij} | 1 \leqslant i, j \leqslant n]$ where the x_{ij}'s are variables double indexed as entries in a matrix, and of course, the universal matrix will be $C_n = (x_{ij})$. Clearly any $n \times n$-matrix $C = (c_{ij})$ with entries in any ring A is of the announced form $(\phi(x_{ij}))$; just let $\phi\colon R_n \to A$ be defined by the assignments $x_{ij} \mapsto c_{ij}$. The characteristic polynomial depends functorially on the matrix (Lemma 4.63 on page 110) and we infer that $P_C(t) = P_{\phi(C_n)}(t) = \phi(P_{C_n}(t))$, and consequently it suffices to verify the Cayley–Hamilton theorem for the single matrix C_n:

LEMMA 4.68 *If $P_{C_n}(C_n) = 0$, then $P_C(C) = 0$ for all $n \times n$-matrices C.*

(4.69) One may consider the following theorem as the ultimate formulation the Cayley Hamilton theorem; anyhow, in view of the above, it implies Cayley–Hamilton as formulated in Theorem 4.65:

THEOREM 4.70 *The universal matrix C_n is semi-simple. Hence $P_{C_n}(C_n) = 0$.*

PROOF: Let $K = \mathbb{Q}(x_{ij} | 1 \leqslant i, j \leqslant n)$ be the fraction field of R_n. Aiming at an absurdity, we assume that the characteristic polynomial $P(t) = P_{C_n}(t)$ has a multiple root in some extension L of K. The ring R_n is a UDF, and Exercise 3.9 on page 78 yields that $P(t)$ and its derivative $P'(t)$ have a common factor $Q(t)$ in $R_n[t]$, which is a monic polynomial since $P(t)$ is (the units in R_n are just ± 1). Hence for any $\phi\colon R_n \to A$, the two polynomials $\phi\big(P(t)\big)$ and $\phi\big(P'(t)\big)$ have the nontrivial common factor $\phi\big(Q(t)\big)$, but $\phi(P(t))$ being the characteristic polynomial of $\phi(C)$ and $\phi\big(P(t)'\big)$ its derivative, it follows that $\phi(P(t))$ has a multiple root. Consequently, no matrix at all can have distinct eigenvalues, which is utterly absurd (matrices with distinct eigenvalues exist!). Hence C_n is semi-simple. ❑

An epilogue

A technique that permeates modern algebraic geometry is to try to represent functors; *i. e.* to find a universal object of the kind one is interested in, which dictates the behavior of all the crowd. As illustrated above, knowing properties of a universal object can have strong implications. That universal objects exist, is however by no means always true, and lot of activity has gone into trying to give criteria fo existence and into coping with situations where substitutes, almost universal object, may be found. What we did in this sections, is very simple case indeed, but can serve as a leisurely introduction to the formalism.

(4.71) Consider the functor $\mathcal{M}_n\colon \mathsf{Rings} \to \mathsf{NCRings}^*$ that send a (commutative) ring A to the ring of $n \times n$-matrices with entries in A, and whose action on a ring homomorphism ϕ is the map that sends $M = (a_{ij})$ to $\phi(M) = \big(\phi(a_{ij})\big)$. Since matrix products and sums are preserved, $\mathcal{M}_n$ takes values in NCRings. This functor is as one says, *representable* and it is represented by the universal matrix, which means there is an isomorphism of functors

$$\mathrm{Hom}_{\mathsf{Rings}}(R_n, -) \xrightarrow{\simeq} \mathcal{M}_n(-).$$

*This denotes the category of not necessarily commoutative rings

Representable functors (representerbare funktorer)

Well, it assigns $\phi(M_n)$ to the homomorphism $\phi\colon R_n \to A$; and we checked it is bijective for each A. It being functorial boils down to the obvious formula $\psi(\phi(M)) = (\psi \circ \phi)(M)$; or more precisely, to the following diagram being commutative for every ring homomorphism $\psi\colon A \to B$:

$$\begin{array}{ccc} \mathrm{Hom}_{\mathsf{Rings}}(R_n, A) & \longrightarrow & \mathcal{M}_n(A) \\ {\scriptstyle \psi_*}\downarrow & & \downarrow{\scriptstyle \psi(-)} \\ \mathrm{Hom}_{\mathsf{Rings}}(R_n, B) & \longrightarrow & \mathcal{M}_n(B) \end{array}$$

The horizontal arrows are the ones above, and the two vertical ones acts respectively as $\phi \mapsto \psi \circ \phi$ and $M \mapsto \psi(M)$.

Functorial maps (funktorielle avbildninger)

(4.72) NATURAL TRANSFORMATIONS In a more general setting, if $F, G\colon \mathsf{C} \to \mathsf{D}$ are two functors, a *functorial map* or a *map of functors*, also called a *natural transformation*, from F to G is just a collection of maps θ_A (*i. e.* arrows in D)—one for each object A in C—such that the diagram

Natural transformations (naturlige transformasjoner)

$$\begin{array}{ccc} F(A) & \xrightarrow{\theta_A} & G(A) \\ {\scriptstyle F(\phi)}\downarrow & & \downarrow{\scriptstyle G(\phi)} \\ F(B) & \xrightarrow[\theta_B]{} & G(B) \end{array}$$

commutes for each arrow $\phi\colon A \to B$ in C. And of course, a natural transformation is called an isomorphism, or a *natural equivalence* as one prefers to say, if there is another from G to F such that the two compositions equal the identity.

Natural equivalences (naturlige ekvivalenser)

4.8 *Appendix: Direct and inverse limits*

There is no reason that the union of a collection of submodules in general should be a submodule; no more than the union of two lines through the origin in space is a plane. This is an additive issue (unions of submodules are obvious closed under multiplication by ring elements) and concerns most abelian groups. In Paragraph 2.31 on page 41 we

argued that the union of two subgroups of an abelian group, neither contained in the other, is not a subgroup.

(4.73) There is however a natural condition that ensures the union to be a submodule. One says that the collection is *directed* if for any two members there is a third containing both; that is, for any pair M_i and M_j from the collection $\{M_i\}_{i\in I}$ there should be an index k so that $M_i \subseteq M_k$ and $M_j \subseteq M_k$. The union $\bigcup_{i\in I} M_i$ will then be closed under addition (and as multiplication poses no problem will be a submodule); indeed, let x and y be two members of the union. This means that there are indices i and j so that $x \in M_i$ and $y \in M_j$, and since the collection is directed, one may find an index k so that $M_i \cup M_j \subseteq M_k$. Both elements x and y then lies in M_k, and their sum does as well. So the sum belongs to the union. We have proven:

Directed collection of submodules (rettede samlinger av undermoduler)

PROPOSITION 4.74 *Let $\{M_i\}_{i\in I}$ be a directed collection of submodules of the A-module M. Then the union $\bigcup_{i\in I} M_i$ is a submodule.*

Direct limits

Direct limits (direkte grenser)

(4.75) There is a general construct called *the direct limit* inspired by the above reasoning It permits us to form "the limit" of certain "directed systems" of modules. The direct limit is a vast generalization, but under certain circumstances it resembles the "union".

As an illustration, imagine a chain of sets S_i indexed by the natural numbers $\mathbb{N}$ such that each S_i is a subset of the succeeding set S_{i+1}. They form an ascending chain which may be displayed as

$$S_1 \subseteq \ldots \subseteq S_i \subseteq S_{i+1} \subseteq \ldots.$$

In the traditional set theory there is no way of defining the union of the S_i's unless they all are subsets of given set. The introduction of the direct limit of the S_i's remedies this, and the direct limit fills the role as their union. But remember, this is just a motivating example; the direct limit is a much more general construct and can be quite subtle. The index set can be any ordered set (with some conditions, though) and the inclusions may be replaced by any maps (with some compatibility conditions).

*Recall that a relation $i \leqslant j$ on a set I is called a preorder *if it is reflexive and transitive; in contrast to a partial order, it is not necessarily antisymmetric; i. e. it might well be that $i \leqslant j$ and $j \leqslant i$ without i and j being equal.*

One may state the definition of the direct limit in any category, and of course, it is expressed by way of a universal property. To fix the ideas we shall only work with modules over a ring A. However, what we shall do is easily translated into several other categories including Sets, Rings and Alg_A.

(4.76) The key notion is that of a *directed systems of modules*. Such a system has two ingredients. The first is a collection $\{M_i\}_{i\in I}$ of modules over A. The index set I is supposed to be a preordered set* whose ordering is *directed*: for any two indices i and j there is a third larger than both; *i. e.* there is a $k \in I$ such that $k \geqslant i$ and $k \geqslant j$. The second ingredient is a collection of A-linear maps $\phi_{ij} \colon M_j \to M_i$, one for each pair (i, j) of elements from I so that $i \geqslant j$, which are subjected to the following two conditions:

Directed orderings (direkte ordninger)

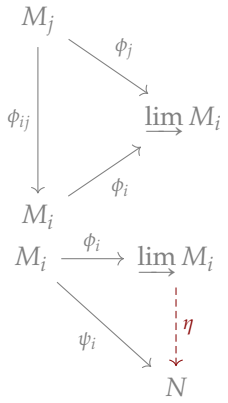
$$M_k \xrightarrow[\phi_{jk}]{} M_j \xrightarrow[\phi_{ij}]{} M_i, \qquad M_k \xrightarrow{\phi_{ik}} M_i$$

- $\phi_{ij} \circ \phi_{jk} = \phi_{ik}$ whenever $k \leqslant j \leqslant i$;
- $\phi_{ii} = \mathrm{id}_{M_i}$.

The system will be denoted $(M_i, \phi_{ij})_I$. If you prefer working in a general category C, just replace the words "A-module" with "object" and A-linear by "arrow". The definition is by way of a universal property: The *direct limit* of the system (M_i, ϕ_{ij}) is an A-module $\varinjlim M_i$ together with a collection of A-linear maps

Direct limits (direkte grenser)

$$\phi_i \colon M_i \to \varinjlim M_i$$



that satisfy $\phi_i \circ \phi_{ij} = \phi_j$, and which are universal with respect to this. In other words, for any given collection $\{N_i\}_{i\in I}$ of A-modules and any given system of A-linear maps

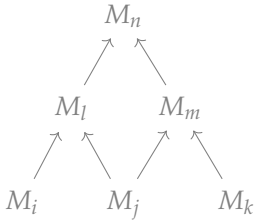
$$\psi_i \colon M_i \to N$$

such that $\psi_i \cdot \phi_{ij} = \psi_j$, there is a unique map $\eta \colon \varinjlim M_i \to N$ such that $\psi_i = \phi_i \circ \eta$. (4.77) Even though the definition of the direct limit may be formulated in any category, whether it exists or not is quit another question. Every direct system possessing a limit, is an exclusive quality most categories do not enjoy. However, the category of modules is among the privileged ones.

The main ideas of the construction are quit transparent, but as most proofs of this type it includes a tiresome list of more or less trivial verifications, whose details we gladly skip. However, to get the mechanism of the limit under the skin, the students are urged to do that work.

PROPOSITION 4.78 *Let A be any ring. Every directed system $(M_i, \phi_{ij})_I$ of modules over A has a direct limit, which is unique up to a unique isomorphism.*

PROOF: We begin with introducing an equivalence relation on the disjoint union $\bigcup_i M_i$. Loosely phrased, two elements are to be equivalent if they become equal somewhere out in the hierarchy of the M_i's. In precise terms, $x \in M_i$ and $y \in M_j$ are defined to be equivalent when there is an index k dominating both i and j such that x and y map to the same element in M_k; that is, $\phi_{ki}(x) = \phi_{kj}(y)$. We shall write $x \sim y$ to indicate that x and y are equivalent.



Obviously this relation is symmetric, since $\phi_{ii} = \mathrm{id}_{M_i}$ it is reflexive, and it being transitive ensues from the system being directed: Assume that $x \sim y$ and $y \sim z$, with x, y and z sitting in respectively M_i, M_j and M_k. This means that there are indices l dominating i and j, and m dominating j and k so that the two equalities $\phi_{li}(x) = \phi_{lj}(y)$ and $\phi_{mj}(y) = \phi_{mk}(z)$ hold true. Because the system is directed, there is an index n larger than both l and m, and by the first requirement above, we find

$$\phi_{ni}(x) = \phi_{nl}(\phi_{li}(x)) = \phi_{nl}(\phi_{lj}(y)) = \phi_{nm}(\phi_{mj}(y)) = \phi_{nm}(\phi_{mk}(z)) = \phi_{nk}(z)$$

and so $x \sim z$. The underlying set of the A-module $\varinjlim M_i$ is the quotient $\bigcup_i M_i / \sim$ and the maps ϕ_i are the ones induced by the inclusions of the M_i's in the disjoint union.

The rest of the proof consists of putting an A-module structure on $\varinjlim M_i$ and checking the universal property. To this end, the salient observation is that any two elements $[x]$ and $[y]$ in the limit may be represented by elements x and y from the same M_k; indeed, if $x \in M_i$ and y in M_j, chose a k that dominates both i and j and replace x and y by their images in M_k. Forming linear combinations is possible by the formula $a[x] + b[y] = [ax + by]$ where the last combination is formed in any M_k where both x and y live; this is independent of the particular k used (the system is directed, and the ϕ_{ij}'s are A-linear). The module axioms follow suit since any equality involving a finite number of elements from the limit may be checked in an M_k where all involved elements have representatives.

Finally, checking the universal property is straightforward: the obvious map from the disjoint union $\bigcup_i M_i$ into N induced by the ψ_i's is compatible with the equivalence relation and hence passes to the quotient; that is, it gives the searched for map $\eta \colon \varinjlim M_i \to N$. ❑

(4.79) Apart from the universal property — which should be the favoured tool for anybody working with direct limits — there are two principles one should have in mind. Firstly, every element in $\varinjlim M_i$ is induced from an element $x \in M_j$ for some index j; that is, it is of the form $\phi_j(x)$ — in fact every finite collection of elements may be represented by elements from a common M_j — and secondly an element $x \in M_j$ maps to zero in the limit if and only if it maps to zero in an M_i for some $i \geqslant j$.

PROPOSITION 4.80 (WORKING PRINCIPLES) *With the notations above, the following two statements hold true:*

i) Every element in $\varinjlim M_i$ is of the form $\phi_j(x)$ for some j and some $x \in M_j$;
ii) An element $x \in M_i$ maps to zero in $\varinjlim M_i$ if and only if $\phi_{ji}(x) = 0$ for som $j \geqslant i$.

PROOF: Clearly every element y in $\bigcup_{i \in I} M_i / \sim$ is the class $[x]$ of some x in some M_i, so that $y = \phi_i(x)$ and *i)* is checked. That $x \sim 0$, means by definition that $\phi_{ji}(x) = \phi_{ji}(0)$ for some $j \geqslant i$, and of course $\phi_{ji}(0) = 0$ so *ii)* holds true. ❑

EXAMPLE 4.21 Let p be any non-zero integer. Consider the directed system indexed by the natural numbers $\mathbb{N}_0$ for which $M_i = \mathbb{Z}$ and $\phi_{ji}(x) = p^{j-i}x$ when $j \geqslant i$; to give it a name, let us denote it by $(\mathbb{Z}, p^{j-i})$. It may depicted as

$$\mathbb{Z} \xrightarrow{p} \mathbb{Z} \xrightarrow{p} \ldots\ldots \xrightarrow{p} \mathbb{Z} \xrightarrow{p} \mathbb{Z} \xrightarrow{p} \ldots$$

where the drawn maps are those shaped like $\phi_{i+1,i}$ which each is multiplication by p. The maps ϕ_{ji} are just the compositions of $j - i$ consecutive maps from the sequence.

$\phi_0(1) = 1$; indeed, $\phi_i(x)\phi_0(1) =$ We contend that $p \cdot \phi_1 = 1$ in $\varinjlim M_i$.

We contend that there is a natural isomorphism $\varinjlim(\mathbb{Z}, p^{j-i}) \simeq \mathbb{Z}[1/p]$. Indeed, define $\psi_i\colon \mathbb{Z} \to \mathbb{Z}[1/p]$ by $\psi_i(x) = p^{-i}x$. Then clearly $\phi_{ij} \circ \psi_j = \psi_i$ and so by the universal property of direct limits there is induced a map $\psi\colon \varinjlim(\mathbb{Z}, p^{j-i}) \to \mathbb{Z}[1/p]$ that satisfies $\psi \circ \phi_i = \psi_i$ for all i. This map is surjective: each element in $\mathbb{Z}[1/p]$ is shaped like ap^{-i} with $a \in \mathbb{Z}$ and hence $\psi([a]) = \psi(\phi_i(a)) = \psi_i(a) = ap^{-i}$. And it is injective: assume that $\psi([a]) = 0$ and chose an i so that $[a] = \phi_i(a)$. It follows that $0 = \psi(\phi_i(a)) = \psi_i(a)$, but then $a = 0$ as ψ_i is injective. ☆

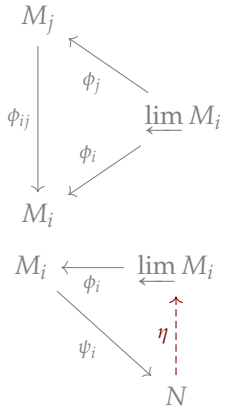
Inverse limits

Most concept in category theory has a dual counterpart, and the dual notion of a direct limit is the *inverse limit* (also called the *projective limit* or just the *limit*). We suppose given a directed set I and for each $i \in I$ an A-module M_i. Moreover, for every pair i, j from I with $i \leqslant j$ we are given maps $\phi_{ij}\colon M_j \to M_i$ which comply with the conditions:

Inverse limits (inverse grenser)

Projective limits (projektive grenser)

- ❑ $\phi_{ij} \circ \phi_{jk} = \phi_{ik}$;
- ❑ $\phi_{ii} = \mathrm{id}_{M_i}$.



The definition is by way of a universal property: the inverse limit $\varprojlim M_i$ is a module together with maps $\phi_i\colon \varprojlim M_i \to M_i$ that satisfy $\phi_i = \phi_{ij} \circ \phi_j$ when $i \leqslant j$, and which are universal in this regard; that is to say, for any other module N together with maps $\psi_i\colon N \to M_i$ with $\psi_i = \phi_{ij} \circ \psi_j$ there is a unique map $\eta\colon N \to \varprojlim M_i$ such that $\psi_i = \phi_i \circ \eta$.

PROPOSITION 4.81 *Let A be a ring. Every directed inverse system of A-modules has an inverse limit.*

PROOF: Consider the product $\prod_i M_i$ and define a submodule by

$$L = \{\, (x_i) \mid x_i = \phi_{ij}(x_j) \text{ for all pairs } i, j \text{ with } i \leqslant j \,\},$$

The projections induce maps to $\phi_i\colon L \to M_i$, and we claim that N together with those maps constitute the inverse limit of the system. A family of maps $\psi_i\colon N \to M_i$ defines a map $\eta\colon N \to \prod_i M_i$ by $x \mapsto (\psi_i(x))$. When the ψ_i's satisfy the compatibility constraints $\psi_i = \phi_{ij} \circ \psi_j$ this maps takes values in L. It is clearly unique, and that gives the desired universal property. ❑

EXAMPLE 4.22 *Power series:* Let A be a ring and x a variable. For each $i \in \mathbb{N}$ let $M_i = A[x]/(x^i)$, and if $i \leqslant j$ let $\phi_{ij}\colon M_j \to M_i$ be the canonical reduction map $A[x]/(x^j) \to A[x]/(x^i)$. We contend that the projective limit $\varprojlim M_i$ equals the formal powers series ring $A[[x]]$: indeed, for each i it holds that $A[[x]]/(x^i) = A[x]/(x^i)$ so by the universal property of $\varprojlim M_i$, there is map $\eta\colon A[[x]] \to \varprojlim M_i$ satisfying

$\phi_i(\eta(\sum_j a_j x^j)) = \sum_{j<i} a_j x^j$, where $\phi_i \colon \varprojlim M_i \to M_i$ is the canonical map. It is straightforward to verify that this is an isomorphism. ☆

EXAMPLE 4.23 *p-adic integers:* Let p be a prime number. The modules of the system are $M_i = \mathbb{Z}/p^i\mathbb{Z}$ and the maps ϕ_{ij} are just the canonical reduction maps $\mathbb{Z}/p^j\mathbb{Z} \to \mathbb{Z}/p^i\mathbb{Z}$ that send a class $[x]_{p^j}$ mod p^j to the class $[x]_{p^i}$ mod p^i. It may be illustrated by the sequence of maps

$$\ldots \longrightarrow \mathbb{Z}/p^{i+1}\mathbb{Z} \longrightarrow \mathbb{Z}/p^i\mathbb{Z} \longrightarrow \ldots \longrightarrow \mathbb{Z}/p^2\mathbb{Z} \longrightarrow \mathbb{Z}/p\mathbb{Z}$$

where each map is the canonical reduction; *i. e.* $\phi_{i+1,i}$ and the other maps ϕ_{ij} from the system are just compositions of $j - i$ consecutive such.

The inverse limit of the system is denoted by $\mathbb{Z}_p$ and is called the ring of *p-adic integers*. ☆

Exercises

(4.38) Let A be a ring. Convince yourself that direct limits exist unconditionally in the category Alg_A of A-algebras.

(4.39) Show that any finite directed set has a largest element. What will a direct limit indexed by such an ordered set be?

(4.40) Let $S \subseteq A$ be a multiplicative system. Define a preorder on S by declaring that $t \leqslant s$ if there is a $u \in S$ such that $s = ut$. For each such pair there is a morphism $\psi_{st} \colon A_t \to A_s$ given by $\psi_{ts}(at^{-n}) = au^n s^{-n}$.

a) Show that S is directed set under the given preorder;
b) Show that (A_s, ψ_{ts}) is a directed system;
c) Show that there is a natural isomorphism $\varinjlim A_s \simeq S^{-1}A$.

(4.41) Let $M = (M_i, \phi_{ij})$ and $N = (N_i, \psi_{ij})$ be two directed systems indexed by the same preordered set I. A morphism α from M to N is a sequence of maps $\alpha_i \colon M_i \to N_i$ commuting with the system maps; that is, $\alpha_i \circ \phi_{ij} = \psi_{ij} \circ \alpha_j$ for all pairs i and j with $i \geqslant j$.

a) Show that this makes the set of directed systems of A-modules indexed by I into a category $\mathsf{DirMod}_{A,I}$.
b) Show that α induces a map $\widetilde{\alpha} \colon \varinjlim M \to \varinjlim N_i$ so that $\alpha \circ \phi_i = \psi_i \circ \alpha_i$, and this this construction is functorial.
c) Show that $(\ker \alpha_i, \phi_{ij}|_{\ker \alpha_i})$ is directed system and that the ψ_{ij} induce maps $\bar{\psi}_{ij} \colon \operatorname{coker} \alpha_j \to \operatorname{coker} \alpha_i$ that makes $(\operatorname{coker} \alpha_i, \bar{\psi}_{ij})$ a directed system.
d) Show that the category of directed system $\mathsf{DirMod}_{A,I}$ is an abelian category.
e) Show that the direct limit of an exact sequence in $\mathsf{DirMod}_{A,I}$ is exact.

(4.42) Let I be a directed set. A subst $K \subseteq I$ is called *cofinal* if for each $i \in I$ there is an $k \in K$ with $i \leqslant k$. Let $(M_i, \psi_{ij})_I$ be a directed system. Show that the restricted

system $(M_k, \psi_{kl})_K$ also is a directed system and that the two direct limits are naturally isomorphic; *i. e.* one has $\varinjlim_{i\in I} M_i \simeq \varinjlim_{k\in K} M_k$.

(4.43) Let p be a prime. For $i, j \in \mathbb{N}$ let $M_i = \mathbb{F}_p[x]$ and let ϕ_{ij} be the map given by $\phi_{ij}(a) = a^{(i-j)p}$. Show that this is a directed system, and that the limit is isomorphic to $\mathbb{F}_p[x^{1/p}]$.

(4.44) Consider the set $\mathbb{N}$ of natural numbers equipped with the divisibility order; that is, $i \geqslant j$ if and only if $j|i$. Prove that this order makes $\mathbb{N}$ a directed set. Consider the system (M_i, ϕ_{ij}) indexed by $\mathbb{N}$ where $M_i = \mathbb{Z}$ for all i, and $\phi_{ij}(a) = ij^{-1}a$ when $j|i$. Check that this is a directed system and show that its limit is isomorphic to the field $\mathbb{Q}$ of rational numbers.

(4.45) Let A is PID with fraction field K and let $p \in A$ be an irreducible element. Consider the directed system (A, p^{j-i}) indexed by $\mathbb{N}$. Show that $\varinjlim(A, p^{j-i}) = A[1/p]$. Consider the system $(A/p^iA, \psi_{ij})$ where ψ_{ij} is the natural inclusion $\psi_{ij}\colon A/p^iA \hookrightarrow A/p^{i+1}A$ that sends x to px. Show that $\varinjlim(A/p_iA, \psi_{ij})$ is isomorphic to $A[1/p]/A$.

Lecture 5

A touch of homological algebra

One of the founders of homological algebra, Saunders Mac Lane, once referred to the subject as "General abstract nonsense", a term that many may find offensive. However, it has no pejorative connotation, but is rather a light-hearted way to warn the readers that arguments are of a very abstract nature far from the specific context—often formulated in the vernacular of homological algebra or category theory. Notions or arguments deserving this honorary title are ubiquitous; they are found all over mathematics—hence their general nature and importance.

We shall in the present chapter lightly touch upon the important notion of a complex, but most of the chapter will be about the special case of short exact sequences. Playing with short exact sequences is a formal pathfinders game, and before seeing it applied and experiencing the force of the method, one may find the nickname "General abstract nonsense" appropriate.

5.1 Exact sequences

Let ϕ and ϕ be two composable A-linear maps and display them as a sequence

$$M \xrightarrow{\phi} N \xrightarrow{\psi} L.$$

Exact sequences (eksakte følger)

This sequence is said two be *exact* if $\ker \psi = \operatorname{im} \psi$, which in particular implies that $\psi \circ \phi = 0$. It frequently happens that such a sequence is part of a longer sequence of maps, extending to the left or to the right, and the extended sequence is then said to be *exact at* N as well. If the composition of any two consecutive maps in the extended sequence equals zero, the sequence is called a *complex*; we shall come back to those later in the chapter. A sequence exact at all places, is simply said to be *exact*.

Complex (kompleks)

(5.1) Two special cases warrant mentioning; the first being when $M = 0$:

$$0 \longrightarrow N \xrightarrow{\psi} L$$

The image of the zero map being the zero submodule (0), exactness boils down to ψ being injective. Similarly, when $L = 0$, the sequence is shaped like

$$M \xrightarrow{\phi} N \xrightarrow{\psi} 0,$$

and it is exact if and only if ϕ is surjective.

EXAMPLE 5.1 Every A-linear map $\alpha\colon M \to N$ lives in the exact sequence

$$0 \longrightarrow \ker\alpha \longrightarrow M \longrightarrow N \longrightarrow \operatorname{coker}\alpha \longrightarrow 0.$$

☆

Short exact sequences

By far the most often met exact sequences, are the so-called short exact sequences; they are the easiest to handle and a long exact sequences can be split into a sequence of short ones.

They are a valuable tool for several purposes; for instance, when one tries to study a module by breaking it down into smaller (and presumptive simpler) pieces.

(5.2) A three-term sequence (or a five-term sequence if you count the zeros)

$$0 \longrightarrow M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \longrightarrow 0 \tag{5.1}$$

Short exact sequences (korteksakte følger)

is called a *short exact sequence* when it is exact. This means that α is injective, that β surjective and that $\operatorname{im}\alpha = \ker\beta$. Of course, the term "short" in the name implies there are long exact sequence as well, and indeed there are, as we shall see later on.

It ensues from the First Isomorphism Theorem (Corollary 4.12 on page 90) that there is a unique isomorphism $\theta\colon M'' \simeq M/\alpha(M')$ shaped in a way that β corresponds to the quotient map. In other words, θ enters into the following commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M' & \xrightarrow{\alpha} & M & \xrightarrow{\beta} & M'' & \longrightarrow & 0 \\ & & {\scriptstyle \alpha|_{M'}}\downarrow & & \| & & \downarrow{\scriptstyle \theta} & & \\ 0 & \longrightarrow & \alpha(M') & \longrightarrow & M & \longrightarrow & M/\alpha(M') & \longrightarrow & 0 \end{array} \tag{5.2}$$

where the maps in the bottom row are respectively the quotient map and the inclusion of $\alpha(M')$ into M. In short, up to isomorphisms all short exact sequence appear as

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

where $N \subseteq M$ is a submodule, and the two maps are respectively the inclusion and the quotient map.

Examples

(5.2) *Direct sums:* The direct sum $M \oplus N$ of two A-modules fits naturally into the short exact sequence

$$0 \longrightarrow N \longrightarrow N \oplus M \longrightarrow M \longrightarrow 0 \tag{5.3}$$

where the left-hand map is the natural inclusion sending x to $(x, 0)$ and the one to the right is the projection onto M, which maps (x, y) to y. In particular, if N and N' are two submodules of a module M, then there is a short exact sequence

$$0 \longrightarrow N \cap N' \xrightarrow{\iota} N \oplus N' \xrightarrow{\sigma} N + N' \longrightarrow 0$$

where ι is the "diagonal inclusion map" $\iota(x) = (x, x)$ and $\sigma(x, y) = x - y$.

(5.3) *A Chinese squence:* The Chinese Remainder Theorem (Theorem 2.72 on page 57) for two ideals may be generalized by saying that the sequence

$$0 \longrightarrow \mathfrak{a} \cap \mathfrak{b} \longrightarrow A \longrightarrow A/\mathfrak{a} \oplus A/\mathfrak{b} \longrightarrow A/\mathfrak{a} + \mathfrak{b} \longrightarrow 0 \tag{5.4}$$

is exact where the two maps in the middle are given by the assignments $x \mapsto ([x]_{\mathfrak{a}}, [x]_{\mathfrak{b}})$ and $([x]_{\mathfrak{a}}, [y]_{\mathfrak{b}}) \mapsto [x]_{\mathfrak{a}+\mathfrak{b}} - [y]_{\mathfrak{a}+\mathfrak{b}}$. Having four non-zero terms it is to long to be called short exact, but it may be obtained by splicing together the two short exact sequences

$$0 \longrightarrow \mathfrak{a} \cap \mathfrak{b} \longrightarrow A \longrightarrow A/\mathfrak{a} \cap \mathfrak{b} \longrightarrow 0$$

and

$$0 \longrightarrow A/\mathfrak{a} \cap \mathfrak{b} \longrightarrow A/\mathfrak{a} \oplus A/\mathfrak{b} \longrightarrow A/\mathfrak{a} + \mathfrak{b} \longrightarrow 0.$$

☆

Exercises

(5.1) Let A be a UFD and x and y two elements. Let $\mathfrak{a}$ be the ideal $\mathfrak{a} = (x, y)$. Show that the sequence

$$0 \longrightarrow A \xrightarrow{\iota} A \oplus A \xrightarrow{\pi} \mathfrak{a} \longrightarrow 0$$

where $\iota(a) = (ya, -xa)$ and $\pi(a, b) = ax + by$, is exact if and only if x and y are without common factors.

(5.2) Let p be prime. Show that for every pair of natural numbers n and m there is a short exact sequence of abelian groups

$$0 \longrightarrow \mathbb{Z}/p^m\mathbb{Z} \longrightarrow \mathbb{Z}/p^{n+m}\mathbb{Z} \longrightarrow \mathbb{Z}/p^n\mathbb{Z} \longrightarrow 0.$$

(5.3) Verify that the two maps defined in the Example 5.3 above are well defined and that the sequence is exact. Deduce the Chinese Remainder Theorem from it.

(5.4) Write down a "Chinese sequence" involving three ideals that generalizes the sequence (5.4) above. Prove it is exact and deduce the Chinese Remainder Theorem for three ideals. HINT: The sequence will have six non-zero terms.

★

Split exact sequences

Some short exact sequence stand out from all the crowd, to wit, the so-called split exact sequences. A short exact sequence

$$0 \longrightarrow M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \longrightarrow 0. \tag{5.5}$$

Split exact sequences (Splitteksakte følger)

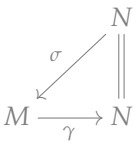
is *split exact* when being isomorphic to the standard sequence (5.3) above. This not only requires that M be isomorphic to the direct sum $M' \oplus M''$, but the somehow stronger requirement that there be an isomorphism inducing the identity on M' and pairing β with the projection, must be met: that is, the isomorphism must fit into the following commutative diagram

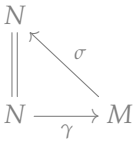
$$\begin{array}{ccccccccc} 0 & \longrightarrow & M' & \xrightarrow{\alpha} & M & \xrightarrow{\beta} & M'' & \longrightarrow & 0 \\ & & \| & & \downarrow\wr & & \| & & \\ 0 & \longrightarrow & M' & \longrightarrow & M' \oplus M'' & \longrightarrow & M'' & \longrightarrow & 0, \end{array} \tag{5.6}$$

where the maps in the bottom sequence are the projection and the inclusion.

(5.3) Of course, all sequences are not split exact, and even if two short exact sequences have the same two extreme modules, the middle modules need not be isomorphic. The easiest example is found among finite abelian groups: Both $\mathbb{Z}/p^2\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}$ appear in the midst of short exact sequences with both extreme modules being $\mathbb{Z}/p\mathbb{Z}$. In general, it is an unsurmountable challenge to classify all possible middle modules given the two extreme ones.

Right amd left sections (høyre- og venstreseksjoner)





(5.4) There is a nice criterion for a short exact sequence to be split involving only one of the maps α or β; to formulate it we need two new concepts. Let $\gamma\colon M \to N$ be A-linear. An A-linear map $\sigma\colon N \to M$ is said to be a *right section* for γ if $\gamma \circ \sigma = \mathrm{id}_N$, and it is called a *left section* if $\sigma \circ \gamma = \mathrm{id}_M$. A map having a right section will be surjective and one with a left section will be injective.

PROPOSITION 5.5 (SPLITTING CRITERION) *Let the short exact sequence*

$$0 \longrightarrow M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \longrightarrow 0$$

of A-modules be given. Then the following three statements are equivalent:

i) The sequence is split;
ii) The map α has a left section;
iii) The map β has a right section.

PROOF: *i) $\Longrightarrow$ ii)* and *i) $\Longrightarrow$ iii)*: If $M = M' \oplus M''$, the canonical inclusion of M'' into M is a right section for the projection onto M'', and dually, the projection onto M' is a left section for the inclusion of M' into M.
iii) $\Longrightarrow$ i): Let σ be a right section of β so that $\beta \circ \sigma = \mathrm{id}_{M''}$. It is good practise (as explained in Proposition 4.27 on page 95) to search for idempotents when trying to decomposing modules into direct sums, and in the present case $\epsilon = \sigma \circ \beta$ is one:

$$(\sigma \circ \beta) \circ (\sigma \circ \beta) = \sigma \circ (\beta \circ \sigma) \circ \beta = \sigma \circ \beta$$

since $\beta \circ \sigma = \mathrm{id}_{M''}$. One has $\epsilon(M) = \sigma(M'')$ and $(\mathrm{id}_M - \epsilon)M = \alpha(M')$, which yields the decomposition

$$M = \epsilon(M) \oplus \mathrm{id}_M - \epsilon M = \alpha(M') \oplus \sigma(M'').$$

Mapping $x \mapsto (\alpha^{-1}(x - \epsilon(x)), \beta(\epsilon(x)))$ gives an isomorphism $\gamma\colon M \simeq M' \oplus M''$ rendering the following diagram commutative

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \| & & \gamma\downarrow\wr & & \| & & \\ 0 & \longrightarrow & M' & \xrightarrow[\alpha]{} & M' \oplus M'' & \xrightarrow[\beta]{} & M'' & \longrightarrow & 0, \end{array}$$

where the upper short exact sequence is the standard "direct sum sequence".
ii) $\Longrightarrow$ i): Let τ be a left section of α so that $\tau \circ \alpha = \mathrm{id}_{M'}$. The idempotent endomorphism of M giving rise to the decompostion is in this case $\epsilon = \alpha \circ \tau$; indeed, ϵ is idempotent: one finds

$$\epsilon^2 = (\alpha \circ \tau) \circ (\alpha \circ \tau) = \alpha \circ (\tau \circ \alpha) \circ \tau = \alpha \circ \mathrm{id}_{M'} \circ \tau = \epsilon.$$

It follows that M decomposes as $M = \epsilon M \oplus (\mathrm{id}_M - \epsilon)M$, and one verifies that $\alpha(M') = \epsilon M$ (obvious) and that β maps $(\mathrm{id}_M - \epsilon)M$ isomorphically onto M''. Tis ensues from the equality $\beta \circ (\mathrm{id}_M - \alpha \circ \tau) = \beta$. Indeed, it holds that

$$\ker \beta \cap (\mathrm{id}_M - \epsilon)M = \alpha(M') \cap (\mathrm{id}_M - \epsilon)M = \epsilon(M) \cap (\mathrm{id}_M - \epsilon)M = 0$$

and obviously β maps $(\mathrm{id}_M - \epsilon)M$ surjectively onto M''.

If $\gamma\colon M \to M' \oplus M''$ is defined by the assignment $x \mapsto \big(\alpha^{-1}(\epsilon(x)), x - \epsilon(x)\big)$, it enters into the digram (5.6), and that ends the affair. ❑

EXERCISE 5.5 *Ambiguity of summands.* The two submodules M' and M'' of M occurring in a decomposition $M = M' \oplus M''$ are seldom unique, *e.g.* just remember that a vector

*To be entirely correct, one should write $\sigma + \iota \circ \gamma$ where ι denotes the inclusion of $\ker\beta$ in M

spaces has many different bases. And in fact, not even their isomorphism classes are determined; but examples of that will be for later (see Examples 5.8 on page 133 and 8.19 on page 227). Neither are sections unique, and this exercise examines this issue. Let $\beta\colon M \to N$ be A-linear and let σ be a right section.

a) Show that for any A-linear $\gamma\colon N \to \ker\beta$ the map $\sigma + \gamma$ is another section*;
b) Show that σ is unique if and only if $\mathrm{Hom}_A(N, \ker\beta) = 0$;
c) Assume that M' and M'' are two complementary split submodules of M. Show that if they are non-isomorphic simple modules, they are unique as submodules.

*To make things simple, we contend ourselves to module categories. There is however a notion of additive categories (where hom-sets are abelian groups and compositions bilinear), and between such categories the term additive functor is meaningful.

★

Additive functors and direct sums

Additive functors (additive funktorer)

An *additive functor* $F\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$ between the categories of modules* over two rings A and B is one that takes sums of maps to sums of maps; that is, $F(\phi+\psi) = F(\phi) + F(\psi)$ for every pair of A-linear maps between to A-modules. Additive functors can be either co- or contra-variant. The functors $\mathrm{Hom}_A(-, N)$ and $\mathrm{Hom}_A(N, -)$ are prototypical examples of such animals.

(5.6) A basic property of an additive functor is that it preserves direct sums, a property of diverse functors that will be fundamental at several later occasions. In stead of repeating multiple ad hoc proofs, we prefer giving one general version. Additionally, the functors will often depend on parameters, as the ones in our applications will do, and the notation may easily appear rather decorated; we believe it easier to grasp the salient points when both notation and context are stripped down to bear essentials.

PROPOSITION 5.7 (ADDITIVE FUNCTORS PRESERVE DIRECT SUMS) *Given rings A and B. Let $F\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$ be an additive functor. Assume that M' and M'' are submodules of an A-module of an A-module M and that $\alpha\colon M \to N$ is an A-linaere map.*

i) If M decomposes as a direct sum $M = M' \oplus M''$, then $F(M)$ decomposes as $F(M) = F(M') \oplus F(M'')$;
ii) The decomposition of $F(M)$ is functorial in that $F(\alpha) = F(\alpha)|_{F(M')} + F(\alpha)|_{F(M'')}$.

PROOF: The natural way to approach direct sum decompositions is through idempotent endomorphism. If M is an A-module and ϵ an idempotent endomorphism of M—as holds for any functor—$F(\epsilon)$ is an idempotent endomorphism of $F(M)$; indeed, ϵ being idempotent means that $\epsilon \circ \epsilon = \epsilon$ and applying F to that equality yields $F(\epsilon) \circ F(\epsilon) = F(\epsilon \circ \epsilon) = F(\epsilon)$.

Now, the equality $\mathrm{id}_M = \epsilon + (\mathrm{id}_M - \epsilon)$ incarnates the decomposition $M = \epsilon M \oplus (\mathrm{id}_M - \epsilon)M$, and when F is as applied to it, it becomes transformed into

$$\mathrm{id}_{F(M)} = F(\mathrm{id}_M) = F\big(\epsilon + (\mathrm{id}_M - \epsilon)\big) = F(\epsilon) + (\mathrm{id}_{F(M)} - F(\epsilon)).$$

because F is additive. Consequently F(M) decomposes into the direct sum $F(M) = F(\epsilon)F(M) \oplus (\mathrm{id}_{F(M)} - F(\epsilon))F(M)$.

We proceed to verify that $F(\epsilon M) = F(\epsilon)F(M)$, which is an essential part of the statement that F respects direct sums. To that end consider the factorisation

$$M \xrightarrow[\pi]{} \epsilon M \xrightarrow[\iota]{} M \qquad (5.7)$$

where π is just the map ϵ but considered to take values in its image ϵM, and ι denotes the inclusion, and as ϵ is idempotent, $\pi \circ \iota = \mathrm{id}_{\epsilon M}$. Applying F to (5.7) yields

$$F(M) \xrightarrow[F(\pi)]{} F(\epsilon M) \xrightarrow[F(\iota)]{} F(M) \qquad (F(5.7))$$

From this ensues that $F(\iota)$ maps $F(\epsilon M)$ onto $F(\epsilon)F(M)$, but since $\pi \circ \iota = \mathrm{id}_{\epsilon M}$, it holds true that $F(\pi) \circ F(\iota) = \mathrm{id}_{F(\epsilon M)}$ so that $F(\iota)$ is an injection, and hence $F(\epsilon M) = F(\epsilon)F(M)$. We have proven the first assertion using an idempotent ϵ such that $M' = \epsilon M$ and $M'' = (\mathrm{id}_M - \epsilon)M$.

What is left to verify is the assertion about functoriality. Now, α decomposes as $\alpha = \alpha|_{M'} + \alpha|_{M''}$ and since F is additive we will be through, once we have proven that $F(\alpha)|_{F(M')} = F(\alpha|_{M'})$ (by symmetry, the same will then hold for $\alpha|_{M''}$), but checking that is effortless: Applying the functor F to the following commutative diagram where ϵ and ι are as above,

$$M' = \epsilon M \xrightarrow[\iota]{} M \xrightarrow[\alpha]{} N$$

we obtain the commutative diagram

$$F(M') = \epsilon F(M) \xrightarrow[\iota]{} F(M) \xrightarrow[F(\alpha)]{} F(N),$$

and that's it. ❑

(5.8) We round off this little excursion into category theory with a result on natural transformations between additive functors, which basically asserts that such transformations preserves direct sums. Assume that $F, G\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$ are two additive functors and that $\eta\colon F \to G$ is a natural transformation, which is just a collection of B-linear maps $\eta_M\colon F(M) \to G(M)$, one for each A-module M, such that $\eta_N \circ F(\phi) = G(\phi) \circ \eta_M$ for each A-linear $\phi\colon M \to N$.

$$\begin{array}{ccc} F(M) & \xrightarrow{\eta_M} & G(M) \\ {\scriptstyle F(\phi)}\downarrow & & \downarrow{\scriptstyle G(\phi)} \\ F(N) & \xrightarrow[\eta_N]{} & G(N) \end{array}$$

PROPOSITION 5.9 *Assume that we are given two additive functors $F, G\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$ and a natural transformation $\eta\colon F \to G$. If an A-module M decomposes as $M = M' \oplus M''$, then $\eta_M = \eta_{M'} + \eta_{M''}$.*

PROOF: For any direct summand L' in an A-module L and any additive functor H we shall denote by $\iota_{L'}$ the inclusion of L' in L; then $H(\iota_{L'}) = \iota_{H(L')}$, which is the essential content of Proposition 5.7 above.

Bearing this in mind, we may decompose each $x \in F(M) = F(M') \oplus F(M'')$ as

$$x = \iota_{F(M')}y + \iota_{F(M'')}z = F(\iota_{M'})y + F(\iota_{M''})z,$$

and applying η to that relation, we obtain since η by definition is a module homomorphism

$$\eta_M(x) = \eta_M\big(F(\iota_{M'})y\big) + \eta_M\big(F(\iota_{M''})z\big) = G(\iota_{M'})\eta_{M'}(y) + G(\iota_{M''})\eta_{M''}(z),$$

which is precisely the identity we want. ❑

Note that the assertions remains true if one *e.g.* replaces one or both of the categories by the categories mod_A and mod_B of finitely generated modules (or for that matter by any additive categories; our proofs are entirely arrow based).

Exercises

(5.6) Let F be a covariant functor $\mathsf{Mod}_A \to \mathsf{Mod}_B$. Let $\phi\colon M \to N$ be linear and assume it has a left (respectively right) section. Show that $F(\phi)$ has a left (respectively right section). What if F is contravariant? Give an example of a functor F and a module with a decomposition $M = N \oplus N'$ such that $F(M)$ is not isomorphic to $F(N) \oplus F(N')$.

(5.7) Additive functors do not necessarily commute with *infinite* direct sums. Prove that $\mathrm{Hom}_{\mathbb{Z}}(\bigoplus_{i\in\mathbb{N}} \mathbb{Z}, \mathbb{Z}) = \prod_{i\in\mathbb{N}} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}) = \prod_{i\in\mathbb{N}} \mathbb{Z}$. Prove that $\prod_{i\in\mathbb{N}} \mathbb{Z}$ is not isomorphic to $\bigoplus_{i\in\mathbb{N}} \mathbb{Z}$. HINT: For the first question go back up to Proposition 4.23 on page 93; for the second verify that the two abelian groups are of different cardinalities, or resort to Exercise 4.27 on page 103.

(5.8) Short exact sequences which are not split, may cease bing exact when exposed to an additive functor. Describe the resulting sequence when one applies the functor $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/p\mathbb{Z}, -)$ to the standard short exact sequence

$$0 \longrightarrow \mathbb{Z}/p\mathbb{Z} \longrightarrow \mathbb{Z}/p^2\mathbb{Z} \longrightarrow \mathbb{Z}/p\mathbb{Z} \longrightarrow 0.$$

What happens when $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/p^2\mathbb{Z}, -)$ is applied?

★

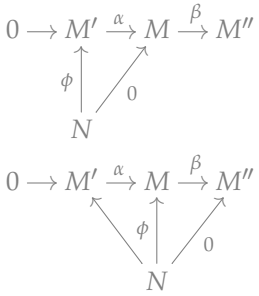
5.2 *Left exactness of hom-functors*

That hom-functors are left exact, is not a very deep result—one checks it by just doing what is needed—but is a fundamental property of hom's.

(5.10) Suppose given a short exact sequence like (5.1) above and let N be any A-module. Applying the covariant hom-functor $\operatorname{Hom}_A(N,-)$ to the sequence we obtain an induced sequence shaped like

$$0 \longrightarrow \operatorname{Hom}_A(N,M') \xrightarrow{\alpha_*} \operatorname{Hom}_A(N,M) \xrightarrow{\beta_*} \operatorname{Hom}_A(N,M'') \,. \tag{5.8}$$

The maps are simply given by composition; *i. e.* $\alpha_*\phi = \alpha\circ\phi$ and $\beta_*\phi = \beta\circ\phi$, and since $\beta_*\circ\alpha_* = (\beta\circ\alpha)_*$ and $\beta\circ\alpha = 0$, it holds true that $\beta_*\circ\alpha_* = 0$. More is true, the sequence (5.8) will in fact be exact. There are two spots where exactness needs to be checked. The first point is that α_* is injective. But $\alpha_*(\phi) = \phi\circ\alpha$, and since α is assumed to be injective, α_* is injective as well; indeed, $\alpha\circ\phi = 0$ means that the image $\operatorname{im}\phi$ lies in the kernel of α. Secondly, to verify that the sequence (5.8) is exact at the middle spot, assume that $\beta\circ\phi = 0$ for a map $\phi\colon N\to M$. Then ϕ factors through the image $\alpha(M')$, and α being injective, ϕ can be represented as $\alpha\circ\phi'$ for a map $\phi'\colon N\to M'$, which is precisely what we desire.



(5.11) In a similar vein, the contravariant version $\operatorname{Hom}_A(-,N)$ applied to (5.1) yields the sequence

$$0 \longrightarrow \operatorname{Hom}_A(M'',N) \xrightarrow{\beta^*} \operatorname{Hom}_A(M,N) \xrightarrow{\alpha^*} \operatorname{Hom}_A(M',N) \,, \tag{5.9}$$

where the arrows are reversed, and repeating *mutatis mutandis* the argument above one shows that this also is an exact sequence.

Left exact functors (venstre-eksakte funktorer)
Exact functors (eksakte funktorer)

(5.12) It is common usage to refer to the phenomena described above as saying that $\operatorname{Hom}_A(N,-)$ and $\operatorname{Hom}_A(-,N)$ are *left exact functors*. The two functors $\operatorname{Hom}_A(N,-)$ and $\operatorname{Hom}_A(-,N)$ are however, seldom *exact functors* in the sense that they take short exact sequences to short exact sequences. There are crowds of examples that β_* and α^* are not surjective.

A large part of homological algebra was developed just to describe the "missing cokernels" $\operatorname{coker}\beta_*$ and $\operatorname{coker}\alpha^*$. In general, the answer to this challenge is that the two sequences can be extended *ad infinitum* to the right to yield long exact sequences which involve so-called *Ext*-modules. These modules depend only on the modules involved, not on the maps in the original short exact sequence, and of course, they depend on N as well. However, the maps in the long exact sequence depend on the entire short exact sequence. In some good cases the Ext-modules can be computed and the long exact sequences be controlled.

Right exact functors (høyreeksakte funktorer)

(5.13) There is of course the symmetric notion of *right exact functors*, with the lack of exactness appearing at the left end of a sequence. The tensor product, which shortly will be introduced, will be of this kind.

One meets these semi exact functors in a variety of contexts and defined in different abelian categories. The modules—or one should rather say *objects*— involved in long exact sequences associated with short exact ones, depend functorially on the objects and are the famous *derived functors*. Most cohomology theories in the universe can be constructed like this.

(5.14) In paragraph 5.10 above we proved the "only-if-part" of the following proposition (although we worked with short exact sequences like in (5.1), we never used that β was surjective).

PROPOSITION 5.15 (LEFT EXACTNESS I) *Let the sequence*

$$0 \longrightarrow M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \tag{5.10}$$

be given and assume that $\beta \circ \alpha = 0$. The sequence is exact if and only if for all A-modules N the sequence

$$0 \longrightarrow \mathrm{Hom}_A(N, M') \longrightarrow \mathrm{Hom}_A(N, M) \longrightarrow \mathrm{Hom}_A(N, M'') \tag{5.11}$$

is exact.

PROOF: To attack the remaining "if-part" assume that (5.11) is exact for all A-modules N. If α is not injective, take $N = \ker \alpha$, which is non-zero, and let ι be the inclusion of $\ker \alpha$ in M'. Then $\alpha \circ \iota = 0$, but ι is non-zero so α_* is not injective.

In a similar vein, if the image $\mathrm{im}\,\alpha$ is strictly smaller than the kernel $\ker \beta$, take $N = \ker \beta$ and consider the inclusion map ι of N in M. By choice it holds that $\beta \circ \iota = 0$, but ι cannot factor though α since $\mathrm{im}\,\alpha$ is strictly contained in $\mathrm{im}\,\iota$. ❑

(5.16) There is also an assertion dual to the one of Proposition 5.15 above. The proofs of the two being quit similar, we leave all the checking to the zealous students; it is a good training for these diagram-arguments. The assertion reads as follows:

PROPOSITION 5.17 (LEFT EXACTNESS II) *Let the sequence*

$$M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \longrightarrow 0 \tag{5.12}$$

be given and assume that $\beta \circ \alpha = 0$. The sequence is exact if and only if for all A-modules N the sequence

$$0 \longrightarrow \mathrm{Hom}_A(M', N) \longrightarrow \mathrm{Hom}_A(M, N) \longrightarrow \mathrm{Hom}_A(M'', N) \tag{5.13}$$

is exact.

EXAMPLE 5.4 As alluded to in Paragraph 5.12 above, even if the map β is surjective, the induced map β_* will in most cases not be surjective. The simples examples are the short exact sequences of abelian groups

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \xrightarrow{\beta} \mathbb{Z}/n\mathbb{Z} \longrightarrow 0, \tag{5.14}$$

where the left map is multiplication* by an integer n, and β the canonical projection. It obviously holds true that $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) = 0$ and $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$, so the induced sequence becomes

$$0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbb{Z}/n\mathbb{Z},$$

and, of course, a map $0 \to \mathbb{Z}/n\mathbb{Z}$ cannot be surjective! We may as well apply the functor $\operatorname{Hom}_{\mathbb{Z}}(-, \mathbb{Z}/n\mathbb{Z})$ to (5.14) and obtain the sequence

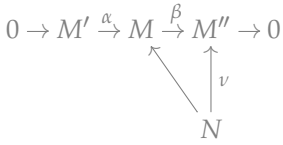
$$0 \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \xrightarrow{n} \mathbb{Z}/n\mathbb{Z}.$$

Of course, multiplication by n (upper star of multiplication by n is multiplication by n) is the zero map on $\mathbb{Z}/n\mathbb{Z}$ and is not surjective. ☆

*True to earlier notation one should write the write $[n]$ for the multiplication map, but grown up algebrists don't do that

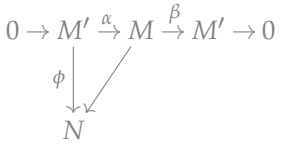
Exercises

(5.9) Convince yourself that β_* being surjective means that any A-linear map $\phi\colon N \to M''$ can be lifted to an A-linear map into M, as the diagram in the margin illustrates.



(5.10) Give the argument referred to in the previous paragraph in detail.

(5.11) In most cases the map α^* will not be surjective even if α is. Convince yourself that α^* being surjective means that any map $\phi\colon M' \to N$ can be extended to a map $M \to N$, as in the marginal diagram



★

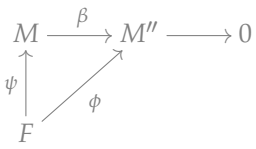
Projective and injective modules

(5.18) In view of the two left exactness theorems two classes of modules stand out, namely the ones such that the functor $\operatorname{Hom}_A(N, -)$ is exact and those such that $\operatorname{Hom}_A(-, N)$ exact. The former are called *projective modules* (they are ubiquitous in commutative algebra, and we shall come back to them) and the latter are the so-called *injective modules*.

Projective modules (projektive moduler)

Injective modules (injektive moduler)

(5.19) The universal mapping property that free modules enjoy, entails that they are projective, but they are by no means the only ones. When the base ring has non-trivial idempotents, there are cheap examples, but finding examples over say integral domains requires some effort. Some classes of rings, as the local rings or the polynomial rings, enjoy the property that all projective modules are free. Over local rings this is an easy



consequence of Nakayma's lemma when the modules are finitely generated and a result of Kaplansky's in general, but over polynomial rings it is a deep theorem, first conjectured by Jean Pierre Serre in 1955 and proved by Daniel Quillen and Andrei Suslin about twenty years later.

PROPOSITION 5.20 *If F is a free A-module, the functor $\mathrm{Hom}_A(F,-)$ is exact; in other words, free modules are projective.*

PROOF: It suffices to prove that β_* is surjective when $\beta\colon M \to M''$ is surjective. Suppose that $\phi\colon F \to M''$ is given, and let $\{f_i\}_{i\in I}$ be a basis for F. Since β is surjective, there are elements $\{m_i\}_{i\in I}$ such that $\beta(m_i) = \phi(f_i)$. By the Universal Mapping Property of free modules (on page 100), there is a map $\psi\colon F \to M$ with $\psi(f_i) = m_i$. Then $\beta\circ\psi = \phi$ since $\beta(\psi(f_i)) = \beta(m_i) = \phi(f_i)$, and two maps agreeing on a basis are equal. ❑

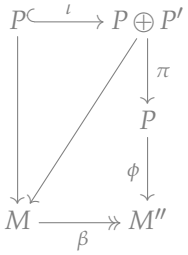
(5.21) Although projective modules are not necessarily free, they are closely related to free modules; they will always be a direct summand in a free module:

PROPOSITION 5.22 *An A-module is projective if and only if it is a direct summand in a free module.*

PROOF: Let P be the module, and assume to begin with that P is projective. Let $\{p_i\}_{i\in I}$ be a generating set for P (finite or not, we do not care about the size) and consider the exact sequence

$$0 \longrightarrow K \longrightarrow \bigoplus_{i\in I} A \xrightarrow{\phi} P \longrightarrow 0,$$

where a string $(a_i)_{i\in I}$ is sent to $\sum_i a_i p_i$ (which is meaningful since merely finitely many of the a_i's are non-zero). Because the p_i's generate P, this map is surjective. The point is that because P is projective, the identity map $\mathrm{id}_P\colon P \to P$ can be lifted to a map $\sigma\colon P \to \bigoplus_{i\in I} A$. This means that $\phi\circ\sigma = \mathrm{id}_P$, and the lifting σ is a right section of ϕ. Hence P lies split in $\bigoplus_{i\in I} A$ by the Splitting Criterion on page 124.



To prove the converse implication let $\beta\colon M \to M''$ be a surjection and let $\phi\colon P \to M''$ be given. Assume further that P' is a complement to P in a free module; that is, $P\oplus P'$ is free. Consider the map $P\oplus P' \to M''$ sending a pair (x,y) to $\phi(x)$. This map can be lifted as $P\oplus P'$ is free, and restricting the lifted map to P yields a lifting of ϕ. ❑

Examples

(5.5) *A projective module which is not free:* Let $A = \mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ and consider $M = \mathbb{Z}/2\mathbb{Z}\times(0)$ which has a natural structure as an A-module. Then M is projective, since if $N = (0)\oplus\mathbb{Z}/2\mathbb{Z}$, we have $M\oplus N \simeq A$ (as A-modules!). However, M is clearly not free, since any free A-module $M \simeq A^I$ must have at least four elements!

(5.6) *Another projective module which is not free:* A similar example can be constructed over the ring $A = \prod_{i=0}^{\infty}\mathbb{Z}$. We may regard $M = \mathbb{Z}$ as an A-module embedding it

as the 0-th component $\mathbb{Z} \subseteq \prod_{i=0}^{\infty} \mathbb{Z}$. Then $\mathbb{Z}$ is projective, since $\mathbb{Z} \oplus \left(\prod_{i=1}^{\infty} \mathbb{Z}\right) \simeq A$. However, note that the ring A is uncountable. In particular, this means that $\mathbb{Z}$ certainly is not isomorphic to any module of the form A^I. The same argument shows that the A-module $M = \bigoplus_{i=0}^{\infty} \mathbb{Z}$ is projective, but not free.

(5.7) *More non-free projective modules:* The modules emerging in the previous examples are of a similar character; they are instances of modules arising when the ring A is a non-trivial direct product $A = A_1 \times A_2$. The factor $A_1 \times \{0\}$ (or $\{0\} \times A_2$) lies split in A as an A-module, and consequently it is projective. However, is not free as the product is non-trivial. One way to see this is to consider the annihilator ideal $(0 : A_1 \times \{0\})$. For free modules the annihilator ideal is the zero ideal (indeed, if e is a basis element, $xe = 0$ implies $x = 0$.), where as for $A_1 \times \{0\}$ it equals $\{0\} \times A_2$ which is non-zero since the product is non-trivial.

(5.8) *More sophisticated examples:* There is a large and all important class of rings called Dedekind rings in which all ideals are projective. A rich source of Dedekinds rings are the coordinate rings of the so-called affine regular curves in algebraic geometry, and the core activity of algebraic number theory is the study of Dedekind rings which are finitely generated $\mathbb{Z}$-modules. The quadratic extensions $\mathbb{Z}[\sqrt{n}]$ and $\mathbb{Z}[(1+\sqrt{n})/2]$ according to n being congruent to one modulo 4 or not, are examples of such*. And in most of these rings you will find ideals that are not principal; that is, ideals that are not free modules. We might as well have given examples from geometry, and for the geometers we offer a treatment of the ideals in the coordinate ring of an affine elliptic curve at a later occasion (Exercise 8.19 on page 227). The two cases are strikingly similar, and of course, there is a common theory behind—but that will also be for later.

*They are the integral closures of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{n})$

For the moment we content ourself with giving just one illustrating example, the ideal $\mathfrak{a} = (2, 1 + i\sqrt{5})$ in the ring $A = \mathbb{Z}[i\sqrt{5}]$. We shall give an explicit construction of $\mathfrak{a}$ as a direct summand in the free module $A \oplus A$. Hence it will be projective, but not being principal it is not free. In fact, we shall prove that $\mathfrak{a} \oplus \mathfrak{a} \simeq A \oplus A$, which also serves as an example of two isomorphic direct sums whose summands are not isomorphic. To ease the notation, we let $z = 1 + i\sqrt{5}$; then $\bar{z} = 1 - i\sqrt{5}$ and $z\bar{z} = 6$.

The free module $A \oplus A$ has the usual basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$. The gist of the construction is the map

$$\alpha \colon A \oplus A \to \mathfrak{a} \subseteq A$$

defined by the assignments $e_1 \mapsto 2$ and $e_2 \mapsto z$. We shall identify a submodule M inside $A \oplus A$ that α maps isomorphically onto $\mathfrak{a}$ and thereby proving that $\mathfrak{a}$ lies split in $A \oplus A$. The submodule M in question is generated by the two elements $a_1 = -2e_1 + \bar{z}e_2$ and $a_2 = -ze_1 + 3e_2$.

We begin with checking that the restriction $\alpha|_M$ is surjective. This ensues from the very definition of α as the following little calculations show:

$$\alpha(a_1) = -2 \cdot 2 + z\bar{z} = -4 + 6 = 2 \qquad \alpha(a_2) = -z \cdot 2 + 3 \cdot z = z.$$

To prove that α is injective on M, pick an element $a = xa_1 + ya_2$ in M that maps to zero, which means that $2x + zy = 0$ in A. Now, one has

$$xa_1 + ya_2 = -(2x + zy)e_1 + (\bar{z}x + 3y)e_2 = (\bar{z}x + 3y)e_2,$$

and since $0 = \bar{z}(2x + zy) = 2\bar{z}x + \bar{z}zy = 2(\bar{z}x + 3y)$, it follows that $a = 0$.

This shows that $A \oplus A \simeq \mathfrak{a} \oplus \ker \alpha$, but to see that $A \oplus A \simeq \mathfrak{a} \oplus \mathfrak{a}$ some further effort is required. Since $\mathfrak{a} = (2, z) = (2, \bar{z})$, the twins z and $\bar{z}$ enter symmetrically into the picture, so swapping z and $\bar{z}$ and e_1 and e_2, we may as well apply what we just did to the submodule $N \subseteq A \oplus A$ generated by $-ze_1 + 2e_2$ and $-3e_1 + \bar{z}e_2$, and conclude that N lies split and is isomorphic to $\mathfrak{a}$. But one directly verifies that $N \subseteq \ker \alpha$, and since both lie split, they must coincide.

☆

Exercises

(5.12) Let n be a natural number. Decide for which natural numbers m the resulting sequence is exact when the functor $\mathrm{Hom}_{\mathbb{Z}}(-, \mathbb{Z}/m\mathbb{Z})$ is applied to the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow 0\,.$$

(5.13) *Multiplicativity of the characteristic polynomial.* Assume given a commutative diagram of A-modules

ψ	$\star\star$
0	θ

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E & \longrightarrow & F & \longrightarrow & G & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \theta} & & \\ 0 & \longrightarrow & E & \longrightarrow & F & \longrightarrow & G & \longrightarrow & 0 \end{array}$$

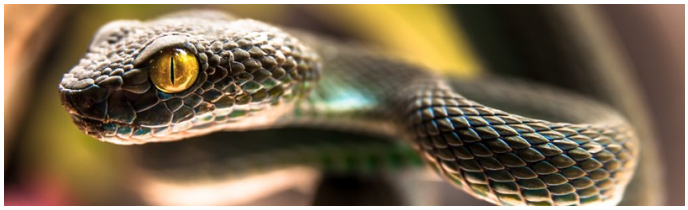
where the rows are exact and the involved modules are free of finite rank. Show the equality $P_\phi(t) = P_\psi(t) \cdot P_\theta(t)$. Conclude that $\det \phi = \det \theta \cdot \det \psi$ and that $\operatorname{tr} \phi = \operatorname{tr} \theta + \operatorname{tr} \psi$. HINT: Exhibit a basis for F in which the matrix of ϕ has an appropriate block decomposition (as in the margin).

(5.14) Consider the exact sequence of finite abelian groups

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0.$$

Show that $\#B = \#A \cdot \#C$.

★



5.3 *Snakes and alike*

(5.23) An all important feature in homological algebra are the so-called *connecting maps* which relate homology modules of complexes in various ways. A simple but very useful instance of this feature is described in the Snake Lemma. The name is of "bourbakistic origin", and mnemotechnical efficient. See the next section below for the reason behind the name.

LEMMA 5.24 (THE SNAKE LEMMA) *Given a diagram*

$$\begin{array}{ccccccccc} & & M_1 & \xrightarrow{\phi_1} & M_2 & \xrightarrow{\phi_2} & M_3 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & & \downarrow{\scriptstyle \alpha_3} & & \\ 0 & \longrightarrow & N_1 & \xrightarrow[\psi_1]{} & N_2 & \xrightarrow[\psi_2]{} & N_3 & & \end{array} \tag{5.15}$$

where the rows are exact and the squares are commutative. Then there exists a map $\delta\colon \ker \alpha_3 \to \operatorname{coker} \alpha_1$ *rendering the following sequence exact*

$$\ker \alpha_2 \longrightarrow \ker \alpha_3 \xrightarrow{\delta} \operatorname{coker} \alpha_1 \longrightarrow \operatorname{coker} \alpha_2, \tag{5.16}$$

where the two unmarked maps respectively are the ones induced by and ϕ_2 *and* ψ_1*.*

PROOF: The proof of the Snake Lemma is an example of a sport called *diagram chasing*, which when homological algebra arose, was extensively practised among homological algebraists. We have two missions to complete; firstly, the map δ must be constructed, and secondly, we must verify that the sequence (5.16) is exact.

We begin with the first and most interesting task. A short and dirty mnemotechnical definition of δ is $\psi_1^{-1} \circ \alpha_2 \circ \phi_2^{-1}$ which of course is meaningless as it stands since neither ϕ_2 nor ψ_1 is invertible, but it gives a hint of how to construct δ. For each $x \in \ker \alpha_3$ it holds true, with a liberal interpretation of the inverses, that $\delta(x) = [\psi_1^{-1}(\alpha_2(\phi_2^{-1}(x)))]$, where $[y]$ designates the class in $\operatorname{coker} \alpha_1$ of an element $y \in N_1$.

After this heuristics, the fun is starting: Pick an element $x \in M_3$ so that $\alpha_3(x) = 0$ and lift it to an element y in M_2; that is, pick an element $y \in M_2$ with $\phi_2(y) = x$. The rightmost square of (5.15) being commutative, we infer that $\psi_2(\alpha_2(y)) = 0$; the bottom

$$\begin{array}{ccccc} y & \longmapsto & x & & \\ \downarrow & & \downarrow & & \\ z & \longmapsto & \alpha_2(y) & \longmapsto & 0 \\ \downarrow & & & & \\ \delta(x) & & & & \end{array}$$

line of (5.15) being exact, there is thence a z in N_1 with $\psi_1(z) = \alpha_2(y)$. And that is it; the image of z in $\operatorname{coker}\alpha_1$ is the wanted guy $\delta(x)$.

We made a choice on the way—the choice of a lift of x to M_2—and for the definition of δ to be legitimate, the image of the trapped z in $\operatorname{coker}\alpha_1$ must be independent of that choice. So assume that y' is another element of M_2 that maps to x; then we may write $y' = y + w$ with $\phi_2(w) = 0$. Since the top line of (5.15) is exact, it holds that $w = \phi_1(u)$ for some u, and we find

$$\alpha_2(y') = \alpha_2(y) + \alpha_2(\phi_1(w)) = \alpha_2(y) + \psi_1(\alpha_1(u))$$

Luckily, ψ_1 is injective (the bottom line of (5.15) is exact), so if $z' \in N_1$ is such that $\psi_1(z') = y'$ one has

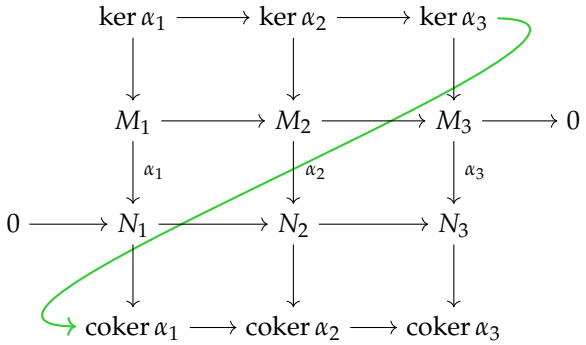
$$z' = z + \alpha_1(u),$$

and finally, this means the images of z and z' in $\operatorname{coker}\alpha_1$ agree, and δ is well defined!

The big game has been snared, and it remains only to check exactness of (5.16): We shall do half of the job and check exactness at $\ker\alpha_3$, letting the zealous students have the fun of checking the other half. So assume that $\delta(x) = 0$. This means that $z = \alpha_1(v)$ for some $v \in M_1$; hence $\alpha_2(y) = \psi_1(z) = \psi_1(\alpha_1(v)) = \alpha_2(\phi_1(v))$. It follows that $y = \phi_1(v) + t$ with $t \in \ker\alpha_2$ and consequently $x = \phi_2(y) = \phi_2(t)$, which is what we need. ❏

Why snake?

(5.25) The reason for the name "Snake Lemma" is apparent when one considers the diagram below. There the map δ connecting $\ker\alpha_3$ to $\operatorname{coker}\alpha_1$ we constructed in the Snake Lemma, zig-zags like a green snake through the diagram.



The Snake Lemma is frequently applied in situations where the map $\phi_1 : M_1 \to M_2$ is injective, and the map $\psi_2 : N_2 \to N_3$ is surjective so that the diagram we depart from

is shaped like

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_1 & \xrightarrow{\phi_1} & M_2 & \xrightarrow{\phi_2} & M_3 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & & \downarrow{\scriptstyle \alpha_3} & & \\ 0 & \longrightarrow & N_1 & \xrightarrow[\psi_1]{} & N_2 & \xrightarrow[\psi_2]{} & N_3 & \longrightarrow & 0. \end{array} \tag{5.17}$$

Such a diagram induces two three term exact sequences, one formed by the kernels of the α_i's and one by their cokernels, and the point is that the snake map δ connects these two sequences. In other words, we have a six term exact sequence

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker \alpha_1 & \longrightarrow & \ker \alpha_2 & \longrightarrow & \ker \alpha_3 & & \\ & & & & & \overset{\delta}{\searrow} & & & \\ & & \operatorname{coker} \alpha_1 & \longrightarrow & \operatorname{coker} \alpha_2 & \longrightarrow & \operatorname{coker} \alpha_3 & \longrightarrow & 0. \end{array} \tag{5.18}$$

LEMMA 5.26 (SNAKE LEMMA II) *Assume given a commutative diagram with exact row as in (5.17). Then the six term sequence (5.18) above is exact.*

PROOF: The sequence is trivially exact at the two extreme slots $\ker \alpha_1$ and $\operatorname{coker} \alpha_3$, and that the snake-part is exact, is just the Snake Lemma. What remains to be done is checking exactness at $\ker \alpha_2$ and $\operatorname{coker} \alpha_2$, and this follows by two simple hunts in the diagram. We shall check exactness at $\ker \alpha_2$, but leave exactness at $\operatorname{coker} \alpha_2$ for the students to practise diagram chasing. So assume that $x \in \ker \alpha_2$ is such that $\phi_2(x) = 0$. Then $x = \phi_1(y)$ for some $y \in M_1$, and $\psi_1(\alpha_1(y)) = \alpha_2(\phi_1(y)) = \alpha_2(x) = 0$. Since ψ_1 is assumed to be injective, it follows that $y \in \ker \alpha_1$, and we are through. ❑

Exercises

(5.15) *The five lemma I.* Use the Snake Lemma to prove the following abbreviated and preliminary version of the five lemma. Assume given a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & & \downarrow{\scriptstyle \alpha_3} & & \\ 0 & \longrightarrow & N_1 & \longrightarrow & N_2 & \longrightarrow & N_3 & \longrightarrow & 0 \end{array}$$

with exact rows. If two of the α_i's are isomorphisms, then the third one is as well.

(5.16) *The five lemma II.* Given a commutative diagram

$$\begin{array}{ccccccccccc} M_0 & \longrightarrow & M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & M_4 \\ \downarrow{\scriptstyle \alpha_0} & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & & \downarrow{\scriptstyle \alpha_3} & & \downarrow{\scriptstyle \alpha_4} \\ M_0 & \longrightarrow & N_1 & \longrightarrow & N_2 & \longrightarrow & N_3 & \longrightarrow & N_4 \end{array}$$

of A-modules with exact rows. Show that α_2 is an isomorphism whenever the four other α_i's are.

There is a slightly stronger assertion namely that if α_1 and α_3 are isomorphisms, α_0 surjective and α_4 injective, then α_2 is an isomorphism. Prove this. ★

There is a plethora of small results like this involving diagrams of different geometric shapes and with suggestive names like the Star Lemma and the Diamond Lemma. Once you have grasped the essence of diagram chasing and remember the Snake Lemma, you should be safe in that corner of the territory of homological algebra. The most important use of connecting homomorphisms is when constructing long exact sequences of homology associated with a complexes; but that will be for a later occasion.

In the next two exercises one may take the following statement for granted:

PROPOSITION 5.27 *Assume that A is a* PID *and that $\phi\colon E \to E$ is an endomorphism of a finitely generated free A-module E. Then ϕ lives in a commutative diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F & \longrightarrow & E & \longrightarrow & A & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \phi} & & \downarrow & & \\ 0 & \longrightarrow & F & \longrightarrow & E & \longrightarrow & A & \longrightarrow & 0 \end{array}$$

where the rows are exact and where F is free.

(5.17) Infer from the Proposition that if $A = \mathbb{Z}$ and ϕ has non-vanishing determinant, it holds true that the cokernel $\operatorname{coker}\phi$ is finite and that $|\det\phi| = \#\operatorname{coker}\phi$. HINT: Use the Snake Lemma.

(5.18) In the same vein as in Exercise 5.17 above, assume that $A = k[t]$ is the polynomial ring over the field k and that ϕ has non-vanishing determinant. Prove that $\operatorname{coker}\phi$ is of finite dimension over k and that $\deg\det\phi = \dim_k \operatorname{coker}\phi$ HINT: Again, the snake is the solution.

Modules of finite presentation (moduler av endelig presentasjon)

(5.19) *Modules of finite presentation.* One says that an A-module M is of *finite presentation* if it sits in an exact sequence

$$A^n \longrightarrow A^m \longrightarrow M \longrightarrow 0$$

where A^n and A^m are finitely generated free modules. In general this is a more restrictive condition on a module than being finitely generated. Over Noetherian rings however, the two are equivalent. Let $N \subseteq M$ be a submodule. Prove that if both N and M/N are of finite presentation, then the same holds for M. HINT: Establish a diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \longrightarrow & M & \longrightarrow & M/N & \longrightarrow & 0 \\ & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & \longrightarrow & A^r & \longrightarrow & A^{r+s} & \longrightarrow & A^s & \longrightarrow & 0 \end{array}$$

with exact rows and all three vertical map being surjective. Then apply the Snake Lemma and Exercise 4.16 on page 99.

★

5.4 Complexes

The consept of a complex originated in topology around the end of the 19^{th} century, when the topologists begun the study of so-called *triangulated spaces*. In is simplest form such a space is represented as union (with some conditions) of oriented simplexes, which are continuous images of certain standard simplices. A standard 1-simplex is just a oriented closed interval, a 2-simplex a triangle, a 3-simplex a tetrahedron etc.

Long before complexes appeared in analysis, but without playing such a centre stage rôle as they did in topology. In courses of calculus of several variable we learned relations like $\operatorname{curl}\operatorname{grad} f = 0$ and $\operatorname{div}\operatorname{curl} t = 0$ for a function f and vector field T, both defined and twice continuously differentiable in an open subset of $\mathbb{R}^3$, and the operators grad, curl and div combine to make up a complex, one of those called deRham complexes.

The definition

In traditionally topology it is customary to distinguish between chain complexes and cochain complexes, but they are just two ways of representing the same thing, although the roles they play in topology are different. In many modern presentations of homological algebra, this practice has for the most ceased, and one just speaks about complexes.

Complexes (komplekser)

(5.28) A *complex* (C_i, d_i) of A-module is a sequence of A-modules $\{C_i\}_{i\in\mathbb{Z}}$ indexed by $\mathbb{Z}$ together with a sequence $\{d_i\}_{i\in\mathbb{Z}}$ of A-linear maps $d_i\colon C_i \to C_{i+1}$ subjected to the condition that the composition of two consecutive ones be zero; that is, $d_{i+1}\circ d_i = 0$ for all i. A complex may be displayed as

$$\cdots \longrightarrow C_i \xrightarrow[d_i]{} C_{i+1} \xrightarrow[d_{i+1}]{} C_{i+2} \longrightarrow \cdots$$

The maps d_i are called *differentials*, but they also obey the names *boundary* or *coboundary** maps. The notation $C_\bullet$ (pronounced C-dot) for a complex is standard, the differential being tacitly understood. The symbolic function of the dot is to indicate a placeholder for the index.

*The names stem from topology where one has chains and chain complexes that give homology, cochains and cochain complexes that give cohomology, and this dichotomy persists for the differentials; hence, both boundary maps and coboundary maps.

There is another and more compact way of denoting a complex; one simply sums up the modules and introduces the module $C_\bullet = \bigoplus_i C_i$. It is a *graded module* A-module* together with an A-linear endomorphism which is homogeneous of degree one and is of square zero. The elements of each summand C_i are the elements homogenous of degree i. Summing up the d_i's gives an A-linear map $d\colon C_\bullet \to C_\bullet$ whose square is zero;

*We consider A a graded ring by declaring all elements to of degree zero.

i. e. $d^2 = 0$. It is a homogeneous map of degree one; that is, it sends homogeneous elements to homogeneous elements, but raises their degree by one. So to give a complex of A-modules is equivalent to giving such a graded A-module

(5.29) A *morphism*, or simply a *map*, ψ from one complex $C_\bullet = (C_i, d_i)$ to another $D_\bullet = (D_i, d'_i)$ is a sequence $\{\psi_i\}_{i\in\mathbb{Z}}$ of A-linear map $\psi_i \colon C_i \to D_i$ that complies with the constituting condition that the ψ_i's commute with the differentials. In other words, $\psi_{i+1} \circ d_i = d'_{i+1} \circ \psi_i$ for all i. Displayed, the map presents itself as a diagram

Morphisms of complexes (morfier mellom komplekser)

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & C_i & \xrightarrow{d_i} & C_{i+1} & \xrightarrow{d_{i+1}} & C_{i+2} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle \psi_i} & & \downarrow{\scriptstyle \psi_{i+1}} & & \downarrow{\scriptstyle \psi_{i+2}} & & \\ \cdots & \longrightarrow & D_i & \xrightarrow[d'_i]{} & D_{i+1} & \xrightarrow[d'_{i+1}]{} & D_{i+2} & \longrightarrow & \cdots \end{array}$$

where all squares are commutative. Two composable maps of complexes, *i. e.* maps so that the source of one equals the target of the other, are composed level by level, and with this composition the complexes of A-modules form a category Cplx_A (which in fact, turns out to be abelian).

As common usage is, one says that a map ψ between two complexes is an *isomorphism of complexes* if it has an inverse. This amounts to each ψ_i being an isomorphism since then the inverses automatically commute with differentials.

(5.30) In the compact notation, a complex is a graded module equipped with a differential of degree one, and a map $\psi \colon C_\bullet \to D_\bullet$ between two complexes is just an A-linear map respecting the grading and commuting with the differentials. The kernel $\ker \psi$, the image $\operatorname{im} \phi$ and the cokernel $\operatorname{coker} \phi$ of ϕ are all graded modules in a natural way as described in Paragraph 4.46 on page 104. The kernel and the image are invariant under the differential; indeed, if $\phi(x) = 0$ one has $\psi(d_C(x)) = d_D(\psi(x)) = 0$ and if $y = \phi(x)$, it holds $d_D(y) = d_D(\psi(y)) = \phi(d_C(x)) =$. Thus they are both complexes. It follows that the differential d_D passes to cokernel $\operatorname{coker} \phi$ and consequently also $\operatorname{coker} \psi$ is a complex. One verifies easily that $\ker \phi$ (respectively $\operatorname{im} \phi$ and $\operatorname{coker} \phi$) is the kernel (respectively the image and the cokernel) of ϕ in the category Cplx_A of complexes. Hence one has the notion of exact sequences of complexes; and in particular short exact sequences, these are tailored like

$$0 \longrightarrow C_\bullet \xrightarrow{\alpha} D_\bullet \xrightarrow{\beta} E_\bullet \longrightarrow 0.$$

In the compact notation they are just usual short exact sequences of graded modules, but with the maps α and β homogenous of degree zero and commuting with the differentials—as one says, they are *chain-wise exact*; in each degree i, it holds that. $\ker \beta_i = \operatorname{im} \alpha_i$.

(5.31) The direct sum $C_\bullet \oplus D_\bullet$ of the two complexes has the obvious underlying graded module $C_\bullet \oplus D_\bullet = \bigoplus_i (C_i \oplus D_i)$ and is equipped with the differential defined by $d(x,y) = (d_C(x), d_D(x))$. Both the two projections and the two canonical inclusions are maps of complexes.

The shift operator (skiftoperatoren)

(5.32) There is a *shift operator* acting on complexes which lowers the degrees by one and changes the sign of the differential; that is, the shifted complex $C_\bullet(1)$ satisfies $(C_\bullet(1))_i = C_{i+1}$ for all i, and the differential is given as $d_{C(1)} = -d_C$. The shift also operates on homomorphisms by the the natural rule $\phi(1)_i = \phi_{i+1}$, and thus (1) is an endofunctor $(1)\colon \mathsf{Cplx}_A \to \mathsf{Cplx}_A$. The d-fold iterate of (1) is naturally denoted by (d).

(5.33) There is a variant of the definition of a complex with the differential decreasing the degree by one; it displays as

$$\ldots \longrightarrow C^i \xrightarrow{d^i} C^{i-1} \xrightarrow{d^{i-1}} C^{i-2} \xrightarrow{d^{i-2}} \ldots .$$

There is of course only a notational difference between the two definitions, and one my pass from one to the other by the conventions $(C^\bullet)_i = (C_\bullet)_{-i}$ and $d^i = (-1)^i d_{-i}$; that is rising (or lowering) the indices accompanied with a change of sign.

Exercises

(5.20) Let $\psi\colon C_\bullet \to D_\bullet$ be a map of complexes. Assume that each ψ_i is a bijection and call the inverse ϕ_i. Show that ϕ_i's commute with the differentials.

(5.21) Let $C_\bullet = (C_i, d_i)$ be a complex of A-modules. Let furthermore $\{\alpha_i\}_{i\in\mathbb{Z}}$ be a sequence of units in A Show that $C'_\bullet = (C_i, \alpha_i d_i)$ is a complex isomorphic to $C_\bullet$. In particular if $\{\epsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of signs; that is, each $\epsilon_i \in \{1, -1\}$, the complexes (C_i, d_i) and $(C_i, \epsilon_i d_i)$ are isomorphic.

(5.22) If $\{C^j_\bullet\}_{j\in J}$ is a family of complexes of A-modules define their direct sum $\bigoplus_{j\in J} C^j_\bullet$ and a direct product $\prod_{j\in J} C^j_\bullet$ and verify that they have the approriate universal properties in the category Cplx_A of complexes.

(5.23) Show that the functor sending an A-module M to the complex $M_\bullet$ whose only non-zero term is the module M in degree zero defines an exact functor $\mathsf{Mod}_A \to \mathsf{Cplx}_A$.

★

The homology of a complex

Homology of a complex (homologien til et kompleks)

(5.34) Since $d_{i+1} \circ d_i = 0$, it holds true that $\operatorname{im} d_i \subseteq \ker d_{i+1}$ and one may form the i-th *homology module* $H_i(C_\bullet) = \ker d_i / \operatorname{im} d_{i-1}$ of the complex. This can of course be done at each level i, and a fundamental invariant of a complex is the so-called total homology $H_\bullet(C_\bullet)$, the direct sum of all the homology modules $H_\bullet(C_\bullet) = \bigoplus_i H_i(C_\bullet)$.

In the compact notation with $C_\bullet$ being a graded module equipped with an endomorphism d_C of square zero, the total homology is just given as the quotient $H_\bullet(C_\bullet) = \ker d_C / \operatorname{im} d_C$.

Exact complexes (ekasakte komplekser)

A complex is said to be *exact* or *acyclic* if all the homology modules vanish; that is, it is exact at every stage.

Ascyclic complexes (asykliske komplekser)

The homology is a functorial construction. A map of complexes $\phi\colon C_\bullet \to D_\bullet$ is required to commute with the differentials and therefore it maps $\ker d_C$ into $\ker d_D$; indeed, $d_D(\phi(x)) = \phi(d_C(x)) = 0$ whenever $x \in \ker d_C$. Similarly ϕ sends $\operatorname{im} d_D$ into $\operatorname{im} d_D$ because $\phi(d_C(x)) = d_D(\phi(x))$. Thus ϕ induces an A-linear map $H_\bullet\phi\colon H_\bullet C_\bullet \to H_\bullet D_\bullet$. It as a matter of easy checking that $H_\bullet(\phi\circ\psi) = H_\bullet\phi \circ H_\bullet\psi$, so that the total homology and each H^i are a covariant functors $\mathsf{Cplx}_A \to \mathsf{Mod}_A$.

Long exact sequences and exact triangles

Where there are short exact sequences there must also be long exact sequences, and we have now come to the point when we shall, hopefully comforting any doubters, establish this fundamental dogma on which the whole homological algebra rests: with any short exact sequence of complexes is associated a long exact sequence in a functorial way. The main players in this performance are the so-called *connecting homomorphisms*. (5.35) A short exact sequence of complexes is just a short exact sequence of the underlying graded modules; that is, an exact sequence shaped like

$$0 \longrightarrow B_\bullet \xrightarrow{\alpha} C_\bullet \xrightarrow{\beta} D_\bullet \longrightarrow 0 \tag{5.19}$$

where of course α and β are morphisms belonging to the category Cplx_A; that is, they are A-linear maps homogeneous of degree zero (*i. e.* respect the grading) and commuting with the differentials. Saying (5.19) being exact is to say it is exact in each degree; in other words, each of he sequences

$$0 \longrightarrow B_i \xrightarrow{\alpha_i} C_i \xrightarrow{\beta_i} D_i \longrightarrow 0$$

is exact. The long exact sequence associated with (5.19) is, well, a long sequence that is exact at each term, and it is shaped as follows

$$\begin{aligned} \dots \longrightarrow H_iB_\bullet \xrightarrow{H_\bullet\alpha} H_iC_\bullet \xrightarrow{H_\bullet\beta} H_iD_\bullet \\ \xrightarrow{\ \delta\ } H_{i+1}B_\bullet \xrightarrow{H_\bullet\alpha} H_{i+1}C_\bullet \xrightarrow{H_\bullet\beta} H_{i+1}D_\bullet \longrightarrow \dots \end{aligned} \tag{5.20}$$

The connecting homomorphism (sambandsmorfien)

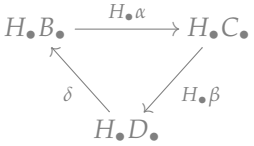
where the newcomer δ—the maps $H_i\alpha$ and $H_i\beta$ are old-timers define above—is the famous *connecting homomorphism* which we are about to construct. Trying to keep the notation as simple and practical as possible we have stripped δ for all sub's and super's, and the dependence on the sequence (5.19) is tacitly understood, as is the degree i. The long sequence extends infinitely in both directions, but of course, if the involved

complexes are concentrated in a certain region; for instance, in positive or negative degrees, the long exact sequence will be confined to the same region.

Long sequences like (5.20) are cumbersome to work with, and a more compact notation has been devised based on so-called *exact triangles**, which compress the long sequence into the compact form

Exact triangles (eksakte triangler)

$$H_\bullet B_\bullet \xrightarrow{H_\bullet\alpha} H_\bullet C_\bullet \xrightarrow{H_\bullet\beta} H_\bullet D_\bullet \xrightarrow{\delta} H_\bullet B_\bullet(1).$$



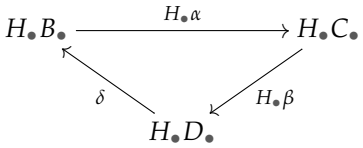
As usual, exactness means that at each module the kernel of the outbound map equals the image of the inbound one, for instance will $\operatorname{im}\delta = \ker H_\bullet(\alpha)$. Rolling out the triangle, degree by degree, one gets back the long sequence (5.20).

PROPOSITION 5.36 *With every short exact sequence of complexes,*

$$0 \longrightarrow B_\bullet \xrightarrow{\alpha} C_\bullet \xrightarrow{\beta} D_\bullet \longrightarrow 0 \tag{5.21}$$

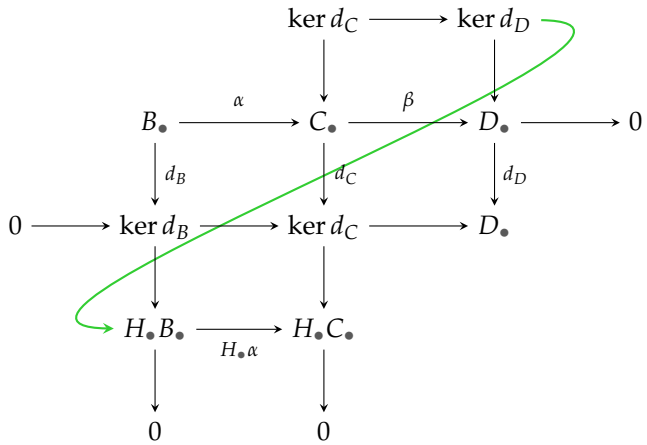
is associated a connecting homomorphism $\delta\colon H_\bullet D_\bullet \to H_\bullet B_\bullet(1)$ giving rise to an exact triangle (and thereby to a long exact sequence)

*One may draw the sequence as a triangle like the one above which justifies the name triangle, but observe that it does not show that the degree of δ equals one.



The connecting homomorphism and the triangle depend functorially on the exact sequence.

PROOF: One may well attack the proof with a direct assault chasing in the diagram, but the chase was already done when proving the snake lemma. The relevant snake diagram is the following one:



where the green snake δ_1 is a precursor for the connecting map δ. Recall the expression $\delta_1(x) = [\alpha^{-1}(d_C(\beta^{-1}(x)))]$ for the snake, where β^{-1} does not stand for a genuine map, but $\beta^{-1}(x)$ denotes any preimage for x, and $[y]$ denotes the homology class of an element y from $B_\bullet$. Homogeneous elements of $D_\bullet$ may be lifted to homogeneous elements of $C_\bullet$ and as the differential d_C is homogeneous of degree one, the snake δ_1 will be homogeneous of degree one too.

To see that the precursor δ_1 passes to the quotient $H_\bullet D_\bullet = \ker d_D / \operatorname{im} d_D$ and yields the desired map δ, we must verify it δ_1 vanishes on $\operatorname{im} d_D$; but if $x = d_D(y)$, it holds that $d_C(z)$ lifts x if z lifts y, and consequently $\delta_1(x) = [\alpha^{-1}(d_C(\beta^{-1}(x)))] = [\alpha^{-1} d_C d_C(z)] = 0$. Hence the snake lemma yields the exact sequence

$$H_\bullet C_\bullet \xrightarrow{H_\bullet \beta} H_\bullet D_\bullet \xrightarrow{\delta} H_\bullet B_\bullet \xrightarrow{H_\bullet \alpha} H_\bullet C_\bullet.$$

This settles the subtler portion of the exactness statement, and the missing piece, that the part

$$H_\bullet B_\bullet \xrightarrow{H_\bullet \alpha} H_\bullet C_\bullet \xrightarrow{H_\bullet \beta} H_\bullet D_\bullet$$

is exact, is not hard and is left to the zealous students.

Finally, we have come to the assertion that the connecting homomorphism δ be functorial. A homomorphism between two exact sequences in Cplx_A is best digested by drawing the diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & B_\bullet & \xrightarrow{\alpha} & C_\bullet & \xrightarrow{\beta} & D_\bullet & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \phi_B} & & \downarrow{\scriptstyle \phi_C} & & \downarrow{\scriptstyle \phi_D} & & \\
0 & \longrightarrow & B'_\bullet & \xrightarrow{\alpha'} & C'_\bullet & \xrightarrow{\beta'} & D'_\bullet & \longrightarrow & 0
\end{array}$$

where all maps are maps of complexes, and the two squares are commutative; the connecting homomorphism being a functorial construct means that the three squares in

$$\begin{array}{ccccccc}
H_\bullet B_\bullet & \xrightarrow{H_\bullet \alpha} & H_\bullet C_\bullet & \xrightarrow{H_\bullet \beta} & H_\bullet D_\bullet & \xrightarrow{\delta} & H_\bullet B_\bullet(1) \\
\downarrow{\scriptstyle H_\bullet \phi_B} & & \downarrow{\scriptstyle H_\bullet \phi_C} & & \downarrow{\scriptstyle H_\bullet \phi_D} & & \downarrow{\scriptstyle H_\bullet \phi_B(1)} \\
H_\bullet B'_\bullet & \xrightarrow{H_\bullet \alpha'} & H_\bullet C'_\bullet & \xrightarrow{H_\bullet \beta'} & H_\bullet D'_\bullet & \xrightarrow{\delta'} & H_\bullet B'_\bullet(1)
\end{array}$$

commute. The only challenges establishing this is to verify that the right square commute, the two others commute by the functoriallity of $H_\bullet \alpha$ and $H_\bullet \beta$. By the quasi-formula $\delta(x) = [\alpha^{-1}(d_C(\beta^{-1}(x)))]$, where x represents a class in $H_\bullet B_\bullet$, we find

$$\begin{aligned}
\phi_B \delta(x) &= [\phi_B \alpha^{-1}(d_C(\beta^{-1}(x)))] = [\alpha'^{-1} \phi_C(d_C((\beta^{-1}(x)))] = \\
&= [\alpha'^{-1} d_C(\phi_C(\beta^{-1}(x)))] = [\alpha'^{-1} d_C(\beta'^{-1}(\phi_D(x)))] = \delta'(\phi_D(x)),
\end{aligned}$$

and we are through. ❏

EXAMPLE 5.9 *Koszul complexes:* The so-called Koszul complexes* form a large collection of complexes. A Koszul complex depends on finite sequences $f_1, \ldots, f_n$ of ring elements. The simplest ones involve merely one element f from the ring and is denoted $K(f)$. It is a two term complex with $K_1 = K_0 = A$, (and $K_i(f) = 0$ for all other i's), and the differential is just multiplication by f:

$$\ldots \longrightarrow 0 \longrightarrow A \xrightarrow{f} A \longrightarrow 0 \longrightarrow 0 \longrightarrow \ldots$$

The homology modules of $K(f)$ that are not automatically equal to zero, are the one in degree zero $H_0K(f) = A/(f)A$ and the one in degree one $H_1K(f) = (0 : f)$, and they enter into the exact sequence

$$0 \longrightarrow (0 : f) \longrightarrow A \xrightarrow{f} A \longrightarrow A/(f)A \longrightarrow 0.$$

When f is a regular element; *i. e.* a non-zero divisor, the Koszul complex $K(f)$ provides a free resolution of $A/(f)A$.

The Koszul complex $K(f_1, f_2)$ on two elements f_1 and f_2 has three non-zero terms; so, displayed it appears as

$$0 \longrightarrow A \xrightarrow{d_2} 2A \xrightarrow{d_1} A \longrightarrow 0,$$

where the repeating zeros are not shown. The first differential is given by the formula $d_1(a_1, a_2) = f_1a_1 + f_2a_2$ and the second by $d_2(a) = (f_2a, -f_1a)$. The homology module $H_2(K(f_1, f_2))$ equals the annihilator of the ideal (f_1, f_2), while $H_0(K(f_1, f_2))$ is the quotient $A/(f, g)$. The homology module $H_1(K(f_1, f_2))$ equals the submodule consisting of pairs (a_1, a_2) such that $f_1a_1 = f_2a_2$ module those of the form $(f_2a, -f_1a)$. In the case when A is a factorial domain, $H_2(K(f_1, f_2)) = 0$ unless $f_1 = f_2 = 0$, and $H_1(K(f_1, f_2))$ vanishes if and only if f_1 and f_2 are without common factors, and in that case, the Koszul complex provides a free resolution of $A/(f_1, f_2)$.

The Koszul complex on three elements is slightly more involved and has four non-zero terms

$$0 \longrightarrow A \xrightarrow{d_2} 3A \xrightarrow{d_1} 3A \xrightarrow{d_0} A \longrightarrow 0$$

where d_1 with respect to the standard basis is given by the matrix

$$\begin{pmatrix} 0 & f_3 & -f_2 \\ -f_3 & 0 & f_1 \\ f_2 & -f_1 & 0 \end{pmatrix}$$

*The Koszul complexes are named after Jean-Louis Koszul who introduced them in the study of Lie algebras; and it seems, however, that Adolf Hurwitz about 50 years before applied several particular cases in algebraic geometry, and some cases appeared already in Hilbert's famous paper xxxx.

Jean-Louis Koszul (1921——2018)

French mathematician

Adolf Hurwitz (1859——1919)

German mathematician

where as the two other maps, d_2 and d_0, have matrices

$$(f_1, f_2, f_3) \text{ and } \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

respectively. For general n the Koszul complexes become much more involved and are best described by the use of exterior powers of maps and modules. ☆

EXERCISE 5.24 Assume that A is a factorial domain and that f and g are two non-zero elements. Show that homology $H_1K(f,g)$ of the Koszul complex is monic submodule of $2A$ generated by the element $(gc^{-1}, -fc^{-1})$ where $c = \gcd(f,g)$. ★

Lecture 6

Tensor products

The term "tensor" appeared for the first time with a meaning resembling the current one in 1898. The German physicists Woldemar Voigt used the word in a paper about crystals. Tensors are these days extensively used in physics, and may be the most prominent example is the so-called "stress-energy-tensor" of Einstein. It governs the general theory of relativity and thereby our lives in the (very) large!

Woldemar Voigt (1850–1919)

German physicist

A slightly less influential occurrence took place in 1938 when the American mathematician Hassler Whitney when working on the universal coefficient theorem in algebraic topology introduced the tensor product of two abelian groups. Certain isolated cases had been known prior to Whitney's work, but Whitneys construct was general, and it is the one we shall give (although subsequently polished by several mathematicians, in particular Nicolas Bourbaki, and generalized to modules). How far apart stress in crystals and the universal coefficient theorem may appear, the concept of tensors is basically the same—the key word being bilinearity.

6.1 Introducing the tensor product

Bilinear maps (bilineære avbildninger)

First of all, let us recall what a *bilinear map* $M \times N \to L$ is, where M, N and L are three modules over a ring A.

It is simply what the name says, a map β which is linear in each of the two variables; that is, when one of the variables is kept fixed, it dependents linearly on the other. For instance, when the second variable is kept constant, it holds true that

$$\beta(ax + by, z) = a\beta(x, z) + b\beta(y, z),$$

where a and b belong to the ring A and x and y are elements in M (and ditto when the first variable is fixed). Frequently, when several rings are around, one says *A-bilinear* to be reminded which ring is considered the base ring.

Hassler Whitney (1907–1989)

American mathematician

A typical example from the world of vector spaces over a field k, would be a scalar product on a vector space V, and within the realm of commutative algebra, the products

of an A-algebra B is a good example; the multiplication map $(a, b) \mapsto ab$ is an A-bilinear map $B \times B \to B$.

Multilinear maps (multilineære avbildninger)

(6.1) There is naturally also the notion of *multilinear maps*, which involves more than two modules. In that case, the source of the map is a product $\prod_i M_i$ of finitely many A-modules M_i and its target is another A-module L. The constituting property is *mutatis mutandis* the same as for bilinear ones: when all but one of the variables are kept constant, the resulting map is A-linear.

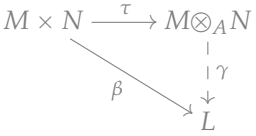
The universal property

(6.2) The tensor product captures in some sense all possible bilinear maps defined on the product of two A-modules M and N, or at least makes them linear. This rather vague formulation becomes precise when phrased as a universal property.

The tensor product (tensorproduktet)

(6.3) The *tensor product* is a pair consisting of an A-module $M \otimes_A N$ together with an A-bilinear map $\tau \colon M \times N \to M \otimes_A N$ that abide by the following rule:

- For each A-bilinear map $\beta \colon M \times N \to L$, there exists a unique A-linear map $\gamma \colon N \otimes M \to L$ suct that $\beta = \gamma \circ \tau$.



In other words, every A-bilinear β factors linearly via τ, as expressed by the commutative diagram in the margin. And as usual with objects satisfying a universal property, the pair τ and $M \otimes_A N$ is unique up to a unique isomorphism.

Exsistence

The construction of the tensor product is rather abstract and serves the sole purpose of establishing the existence. It will seldom be referred to in the sequel, if at all. To ease getting a grasp on the tensor product remember the mantra, so true in modern mathematics: "Judge things by what they do, not by what they are".

(6.4) The construction starts out with the free A module $F = A^{M \times N}$ on the set $M \times N$. The elements of F are finite, formal linear combinations $\sum_i a_i \cdot (x_i, y_i)$ with $x_i \in M$, $y_i \in M$ and $a_i \in A$. In particular, every pair (x, y) is an element of F, and by definition these pairs form a basis for F. We proceed by letting G be the submodule of F generated by all expressions either of the form

$$(ax + a'x', y) - a(x, y) - a'(x', y), \tag{6.1}$$

or of the form

$$(x, ay + a'y') - a(x, y) - a'(x, y'), \tag{6.2}$$

where a and a' are elements from A while x and x' lie in M and y and y' in N.

The tensor product $N \otimes_A M$ is defined as the quotient F/G, and the residue class of a pair (x, y) will be denoted by $x \otimes y$. Having forced the two expressions (6.1) and

(6.2) above to be zero by factoring out the submodule G, we have made $x\otimes y$ a bilinear function of x and y; that is, the following two relations hold true in $M\otimes_A N$:

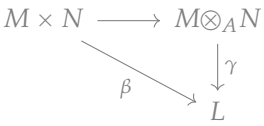
$$\begin{aligned}(ax+a'x')\otimes y &= a(x\otimes y)+a'(x'\otimes y),\\ x\otimes(ay+ay') &= a(x\otimes y)+a'(x\otimes y').\end{aligned} \tag{6.3}$$

In other words, the map $\tau\colon M\times N\to M\otimes_A N$ sending (x,y) to $x\otimes y$ is A-bilinear.

PROPOSITION 6.5 *The pair τ and $M\otimes_A N$ as constructed above satisfy the universal property in paragraph 6.3; in other words, they are the tensor product of M and N.*

PROOF: We already saw that τ is bilinear, so we merely have to check the factorization property. To that end, let $\beta\colon M\times N\to L$ be bilinear. Since $F=A^{M\times N}$ is a free module on $M\times N$, we may , according to the Universal Mapping Principle for free modules (Proposition 4.36 on page 100), define an A-linear map $\bar\beta\colon F\to L$ by sending the basis-elements (x,y) to the values $\beta(x,y)$. Since β is bilinear, this map vanishes on the submodule G. Consequently it factors through the quotient $F/G=M\otimes_A N$ and thus gives the wanted map $\gamma\colon M\otimes_A N\to L$.

Elements shaped like $x\otimes y$ generate the tensor product, and because the value at $x\otimes y$ of any factorization of β is compelled to be $\beta(x,y)$, the uniqueness of γ comes for free. ❑



(6.6) Before leaving the details of the tensor product construction, there is one observation to be made. It will useful to reduce certain questions to questions about finitely generated modules.

Assume given two A-modules N and M a sequence of elements $x_1,\dots,x_r$ from N and a sequence of elements $y_1,\dots,y_r$ from M. Assume further that $\sum_{1\leqslant i\leqslant r}x_i\otimes y_i=0$ in $N\otimes_A M$. We contend that one may then find finitely generated modules $N_0\subseteq N$ containing the x_i's and $M_0\subseteq M$ containing the y_i's, so that the relation $\sum_i x_i\otimes y_i=0$ already holds in $N_0\otimes_A M_0$. Indeed, saying that the relation $\sum_i x_i\otimes y_i=0$ holds in $N\otimes_A M$ is to say that the element $\sum_i(x_i,y_i)$ in $A^{N\times M}$ belongs to the submodule G, and as such it can be expanded as a finite combination of the generators of G; that is, of elements described in (6.1) and (6.2). Letting N_0 (resp. M_0) be the submodule of N (resp. M) generated by the x_i's (resp. the y_i's) and the finite number of x's (resp. y's) that appear in such an expansion, it follows that $\sum_i x_i\otimes y_i=0$ in $N_0\otimes_A M_0$.

6.2 *Basic working formulas*

In this section we present a few principles and properties of the tensor products which together with some basic formulas hopefully should help students grasp "the spirit of the tensor product" and make it easier to work with it. We also discuss some particular

classes of modules, like cyclic modules and free modules, which behave particularly well when exposed to a tensor product.

Decomposable tensors.

Decomposable tensors (dekomponerbare tensorer)

(6.7) For several reasons, tensors of the form $x \otimes y$ deserve a special name; they are dubbed *decomposable tensors*. Only in a very few highly special cases all elements in a tensor product will be decomposable; the usual situation is that most are not (A simple example is discussed in Problem 6.13 on page 160 below. See also Example 6.5 on page 167). A general element in $M \otimes_A N$ may however, be expressed as a finite linear combination $\sum_i a_i \cdot x_i \otimes y_i$ of decomposable tensors since this is already true in the free module $F = A^{M \times N}$.

Consequently, if $\{x_i\}$ is a set of generators of M and $\{y_j\}$ one for N, the decomposable tensors $\{x_i \otimes y_j\}$ form a set of generators for $M \otimes_A N$; in particular, if both factors are finitely generated, the same holds for the tensor product $M \otimes_A N$.

(6.8) To define a map ϕ from $M \otimes_A N$ into any module, it suffices to give the values of ϕ on decomposable tensors $x \otimes y$, provided these values depend bilinearily on x and y. This is an informal and convenient reformulation of the universal property from Paragraph 6.3, certainly more suggestive than working with pairs (x, y).

(6.9) Another useful property of decomposable tensors is subsumed in the slogan "scalars can be moved past the tensor product"; or in precise terms, for every element $a \in A$ it holds true that

$$(ax) \otimes y = x \otimes (ay).$$

This is a simple consquence of the fundamental bilinear relations (6.3) on page 149; with the notation of (6.3), just set $x' = y' = 0$.

Functoriality

Linear maps between A-modules are fundamental tools in algebra, and it comes as no surprise that exploring how maps behave when exposed to tensor products occupies a large part of the theory. As a modest start we shall observe that the tensor product construct is functorial, in the precise meaning that when considered a function of either variable, it gives a functor $\mathsf{Mod}_A \to \mathsf{Mod}_A$; so we have to tell how to tensorize maps.

(6.10) Any A-linear map $\phi\colon M \to M'$ gives rise to an A-linear map $M \otimes_A N \to M' \otimes_A N$ that on decomposable tensors acts as $x \otimes y \mapsto \phi(x) \otimes y$. Since the expression $\phi(x) \otimes y$ depends bilinearily on x and y, this is a viable definition, and the resulting map is naturally babtized $\phi \otimes \mathrm{id}_N$.

It holds true that $\psi \otimes \mathrm{id}_N \circ \phi \otimes \mathrm{id}_N = (\psi \circ \phi) \otimes \mathrm{id}_N$ when ψ and ϕ are two composable maps (the two sides obviously agree on decomposable tensors and thus the identity holds true), and clearly $\mathrm{id}_M \otimes \mathrm{id}_N = \mathrm{id}_{M \otimes_A N}$. Therefore the pair of assignments

$$M \mapsto M \otimes_A N \quad \text{and} \quad \phi \mapsto \phi \otimes \mathrm{id}_N$$

define a functor $-\otimes_A N\colon \mathsf{Mod}_A \to \mathsf{Mod}_A$.

PROPOSITION 6.11 *The functor $-\otimes_A N$ is an A-linear functor. It transforms direct sums into direct sums.*

A formal consequence of a functor being additive is that it preserves finite direct sums (as we established in Proposition 5.7 on page 126), but in general additive functors do not commute with infinite direct sums, so a proof is needed for that case. One is given in Exercise 6.1.

Additive functors (additive funktorer)

A-linear functors (lineære fuktorer)

PROOF: Recall that in Section 5.4 on page 126 we introduced the notion of additive functors: saying the functor is *additive* is saying it transforms sums of maps to sums of maps, and it is *A-linear* if it additionally respects products with scalars; expressed in symbols this reads

$$(a\phi + b\psi)\otimes \mathrm{id}_N = a\cdot\phi\otimes\mathrm{id}_N + b\cdot\psi\otimes\mathrm{id}_N. \tag{6.4}$$

This follows easily from how $-\otimes_A N$ acts on maps together with the basic bilinear relations in (6.3) on page 149. Indeed, one finds

$$\begin{aligned}(a\phi + b\psi)\otimes\mathrm{id}_N\ (x\otimes y) &= ((a\phi(x) + b\psi(x))\otimes y = \\ = a\phi(x)\otimes y + b\psi(x)\otimes y &= \big(a\cdot\phi\otimes\mathrm{id}_N + b\cdot\psi\otimes\mathrm{id}_N\big)\,x\otimes y,\end{aligned}$$

and the two sides of (6.4) agree on decomposable tensors. Hence they are equal since the decomposable tensors generate $M\otimes_A N$. ❑

(6.12) The situation is completely symmetric in the two variables, so if $\psi\colon N\to N'$ is a map, there is a map $\mathrm{id}_M\otimes\psi$ from $M\otimes N$ to $M\otimes N'$ that sends $x\otimes y$ to $x\otimes\psi(y)$, and naturally, one sets $\phi\otimes\psi = (\phi\otimes\mathrm{id}_N)\circ(\mathrm{id}_{M'}\otimes\psi)$.

Some formulas

(6.13) When working with tensor products a series of formulas are invaluable. Here we given the most basic ones revealing the multiplicative nature of the tensor product; together with the direct sum it behaves in a way resembling the product in a ring.

PROPOSITION 6.14 *Suppose that M, N and L are modules over the ring A. Then we have the following four canonical isomorphisms.*

i) Neutrality: $M\otimes_A A \simeq M$;
ii) Symmetry: $M\otimes_A N \simeq N\otimes_A M$;
iii) Associativity: $(M\otimes_A N)\otimes_A L \simeq M\otimes_A(N\otimes_A L)$;
iv) Distributivity: $(M\oplus N)\otimes_A L \simeq (M\otimes_A L)\oplus(N\otimes_A L)$.

There are some comments to be made. Firstly, these isomorphisms are so natural that for all practical purposes they may be consider as identities. Secondly, the general

mechanism that extends associativity from products with three factors to products with arbitrary many factors applies to tensor products, so that any number of parentheses placed in any way in a tensor product with any number of factors can be resolved. And finally, an easy induction establishes the fourth property for any number of summands; with a somehow subtler argument, one may even show it holds for infinitely many.

PROOF: In each case we indicate how a pair of mutual inverses A-linear maps acts on decomposable tensors; this will basically suffice in the two first cases, but in particular the case of associativity, requires some more work.

Neutrality: The product xa is bilinear in x and a and therefore the map $x\otimes a \to xa$ extends to an A-linear map $M\otimes_A A \to M$. The map $M \to M\otimes_A A$ sending x to $x\otimes 1$ is obviously an inverse.

Symmetry: In this case the short hand assignments are $x\otimes y \mapsto y\otimes x$ and $y\otimes x \mapsto x\otimes y$. They are bilinear in view of the fundamental relations (6.3), and hence yield maps between $M\otimes_A N$ and $N\otimes_A M$ which obviously are mutually inverse.

Associativity: This case is more subtle than one should believe at first sight; the (very short) shorthand definition of a map

$$(M\otimes_A N)\otimes_A L \to M\otimes_A (N\otimes_A L)$$

would be $(x\otimes y)\otimes z \mapsto x\otimes(y\otimes z)$, but this is not viable since $x\otimes y$ is not a general member of $M\otimes_A N$. To salvage the situation one introduces some auxiliary maps, one for each $z \in L$.

So, for each element z from L, which we keep fixed, we define an A-linear map

$$\eta_z \colon M\otimes_A N \to M\otimes_A (N\otimes_A L)$$

by the assignment $x\otimes y \mapsto x\otimes(y\otimes z)$; this is legitimate since the expression $x\otimes(y\otimes z)$ is bilinear in x and y (the third variable z is kept fixed).

Obviously the map $\eta_z(t)$ is linear in z and *a priori* being linear in t, it depends bilinearily on t and z. We infer that sending $t\otimes z$ to $\eta_t(z)$ induces a map $(M\otimes_A N)\otimes_A L \to M\otimes_A (N\otimes_A L)$. On decomposable tensors this map behaves as wanted; that is, it sends $(x\otimes y)\otimes z$ to $x\otimes(y\otimes z)$.

A symmetric construction yields a map the other way which sends a decomposable tensor $x\otimes(y\otimes z)$ to $(x\otimes y)\otimes z$. Finally, these two maps are mutually inverses since they act as inverse maps on the decomposables, and the decomposables generate the tensor products.

Distributivity: Another way of phrasing this is to say at the tensor product respects finite directs sums, and this we already established in Proposition 6.11 above. A vague indication of an *ad hoc* proof (in the flavour of the preceding cases) is the short hand

definition $(x,y)\otimes z \mapsto (x\otimes z, y\otimes z)$. The salient point is to extend this to an isomorphism. The detailed proof, formulated for general direct sums, is given in Exercise 6.1 below. Be aware, that contrary to the tensor product $-\otimes_A N$, the hom–functors $\mathrm{Hom}_A(-,N)$, even though being additive, do not commute with infinite direct sums; see Exercise 5.7 on page 128. ❑

Exercises

(6.1) *Tensor products preserve arbitrary direct sums.* Let A be a ring and $\bigoplus_{i\in I} M_i$ a direct sum of A-modules where the index set I is of any cardinality. Let N be another A-module.

a) Define the map $\tau\colon (\bigoplus_{i\in I} M_i)\times N \to \bigoplus_{i\in I}(M_i\otimes_A N)$ by sending $((x_i)_{i\in I}, y)$ to $\sum_i x_i\otimes y$. Show that τ is a well-defined bilinear map. Hint: Merely finitely many of the x_i's are non-zero.

b) Show that τ induces an isomorphism $(\bigoplus_i M_i)\otimes_A N \simeq \bigoplus_{i\in I}(M_i\otimes_A N)$.

(6.2) Consider the four isomorphisms in Proposition 6.14 on page 151. Be explicit about what it means that they are functorial (in every variable involved) and prove your assertions.

★

(6.15) It is worth while to dwell a little on the associativity. Since the parentheses are irrelevant, we may as well skip them and write $M\otimes_A N\otimes_A L$ for $M\otimes(N\otimes_A L)$ (or for that matter for $(M\otimes_A N)\otimes_A L$). According to the universal property of the tensor product bilinearity is the clue for defining maps having source $M\otimes_A N$, and there is a similar trilinearity principle for defining maps sourced in a triple tensor product $M\otimes_A N\otimes_A L$. It suffices to specify ϕ on decomposable tensors $x\otimes y\otimes z$ as long as the specifying expression is trilinear in x, y and z; the precise statement is a as follows:

Lemma 6.16 (Trilinearity principle) *Let A be a ring and M, N, K and L four A-modules. Assume given a map $\phi\colon M\times N\times K\to L$ such that $\phi(x,y,z)$ depends in a trilinear manner on the variables. Then there is unique A-linear map $\phi\colon M\otimes_A N\otimes_A K\to L$ such that $\phi(x\otimes y\otimes z) = \phi(x,y,z)$.*

Proof: The argument is *mutatis mutandis* the same as we gave in the proof of Proposition 6.14 concerning the associative law: Fix an element $z\in K$ and consider $\phi(x,y,z)$; it depends in a bilinear manner on x and y and hence gives rise to a linear map $\eta_z\colon M\otimes_A N\to L$. The dependence of η_z on z obviously being linear, assigning $\eta_z(t)$ to $t\otimes z$ for $t\in M\otimes_A N$ and $z\in K$ is a bilinear in z and t and therefore yields the desired map $(M\otimes_A N)\otimes_A K\to L$.

And again, the map is unambiguously determined since its values are prearranged on the decomposable tensors which generate $(M\otimes_A N)\otimes_A K$. ❑

As a final comment, there is nothing special about the number three in this context. A

similar statement—that is, a *principle of multi-linearity*—holds true for tensor products with any number of factors, but we leave that to the imagination of the reader.

The case of cyclic modules

We now turn to discuss two situation which are frequently met when working with tensor products. Hopefully they will illuminate the working mechanism of the tensor product, but anyhow, they illustrate some of the different phenomena that can occur.

(6.17) Our first example is about the tensor product of two cyclic module and reads as follows:

PROPOSITION 6.18 *Let A be a ring. For any two ideals $\mathfrak{a}$ and $\mathfrak{b}$ in A it holds true that $A/\mathfrak{a}\otimes_A A/\mathfrak{b} \simeq A/(\mathfrak{a}+\mathfrak{b})$. In particular, one has $A/\mathfrak{a}\otimes_A A/\mathfrak{b} = 0$ if and only if the two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are comaximal.*

PROOF: Sending a pair $([x]_\mathfrak{a}, [y]_\mathfrak{b})$ from $A/\mathfrak{a} \times A/\mathfrak{b}$ to the element $[xy]_{\mathfrak{a}+\mathfrak{b}}$ in $A/\mathfrak{a}+\mathfrak{b}$ is well defined and bilinear; indeed, changing x (resp. y) by a member of $\mathfrak{a}$ (resp. $\mathfrak{b}$) changes xy by a member of $\mathfrak{a}$ (resp. $\mathfrak{b}$) as well, so the map is well defined, and as a product clearly depends bilinearily on the factors. This induces a map $A/\mathfrak{a}\otimes_A A/\mathfrak{b} \to A/\mathfrak{a}+\mathfrak{b}$.

One the other hand, the tensor product $A/\mathfrak{a}\otimes_A A/\mathfrak{b}$ is a cyclic A-module generated by $1\otimes 1$ because $[a]_\mathfrak{a}\otimes[b]_\mathfrak{b} = ab\cdot 1\otimes 1$, and clearly elements from both the ideals $\mathfrak{a}$ and $\mathfrak{b}$ kill it; indeed

$$x\cdot 1\otimes 1 = (x\cdot 1)\otimes 1 = 1\otimes(x\cdot 1).$$

So $A/\mathfrak{a}\otimes_A A/\mathfrak{b}$ is a quotient of $A/\mathfrak{a}+\mathfrak{b}$ and is thus squeezed between two copies of $A/\mathfrak{a}+\mathfrak{b}$, by two maps one sending 1 to $1\otimes 1$ and one $1\otimes 1$ to 1. Hence all three must coincide. ❑

The proposition shows that the tensor product of two non-zero modules very well may vanish, and for cyclic modules this happens precisely when the respective annihilators are comaximal. We also observe that an inclusion $\mathfrak{b}\subseteq\mathfrak{a}$ yields an isomorphism $A/\mathfrak{a}\otimes_A A/\mathfrak{b} \simeq A/\mathfrak{a}$, in particular it holds true that $A/\mathfrak{a}\otimes_A A/\mathfrak{a} \simeq A/\mathfrak{a}$.

Finite abelian groups

(6.19) A modest instance of the tensor product being zero, is found among finite abelian groups. Powers of two relatively prime integers p and q are comaximal, for natural numbers μ and ν it holds that $(p^\mu, q^\nu) = \mathbb{Z}$—and we infer that

$$\mathbb{Z}/p^\mu\mathbb{Z}\otimes_\mathbb{Z}\mathbb{Z}/q^\nu\mathbb{Z} = 0.$$

We also infer that if the two natural numbers satify $\mu \leqslant \nu$ it holds that true that

$$\mathbb{Z}/p^\mu\mathbb{Z}\otimes_\mathbb{Z}\mathbb{Z}/p^\nu\mathbb{Z} \simeq \mathbb{Z}/p^\mu\mathbb{Z}$$

for any prime number p since $(p^\mu, p^\nu) = (p^{\min(\nu,\mu)})$. Together with the formulas from Proposition 6.14 and the Fundamental Theorem for Finitely Abelian Groups, these two formulas make it clear how to compute the tensor product of any pair of finite abelian groups.

The case of free modules

(6.20) In the second example we show that the tensor product of two free modules is free.

PROPOSITION 6.21 (TENSOR PRODUCT OF FREE MODULES) *Assume that E and F are free A-modules. Then the tensor product $E\otimes_A F$ is free. More precisely, if $\{e_i\}_{i\in I}$ and $\{f_j\}_{j\in J}$ are bases for respectively E and F, the tensors $e_i\otimes f_j$ with $(i,j) \in I \times J$ form a basis for $E\otimes_A F$.*

This proposition holds true regardless of the cardinalities of I and J, but the case when E and F are of finite rank, warrants to be mentioned specially. One may deduce the finite tank case from Proposition 6.14 by a straightforward induction, however we offer another simple generally valid proof.

COROLLARY 6.22 *If E and F are free A-modules of finite ranks n and m respectively, the tensor product $E\otimes_A F$ is free of rank nm. In particular, for vector spaces V and W of finite dimension over a field k it holds true that $\dim_k V\otimes_k W = \dim_k V \cdot \dim_k W$.*

PROOF OF PROPOSITION 6.21: We contend that the set $\{e_i\otimes f_j\}_{(i,j)\in I\times J}$ is a basis for the tensor product $E\otimes_A F$. As already observed, the elements $e_i\otimes f_j$ form a generating set so we merely have to verify they are linearly independent.

Denote by $a_i(x)$ the i-th coordinate of an element $x \in E$ relative to the basis $\{e_i\}$; that is, one has $x = \sum_i a_i(x)e_i$. Similarly, let $b_j(y)$ be the j-th coordinate of an element $y \in F$. All the $a_i(x)$'s and all the $b_j(y)$'s depend linearly on their arguments.

For each pair of indices $\mu \in I$ and $\nu \in J$ the expression $a_\mu(x)b_\nu(y)$ depends bilinearly on x and y and therefore $x\otimes y \to a_\mu(x)b_\nu(y)$ gives a map $\delta_{\mu\nu}\colon E\otimes F \to A$. This map vanishes on $e_i\otimes f_j$ unless $i = \mu$ and $j = \nu$, and takes the value one on $e_\mu\otimes f_\nu$.

With the $\delta_{\mu\nu}$'s up our sleeve, the rest is a piece of cake: Just apply $\delta_{\mu\nu}$ to a potential dependence relation

$$\sum c_{ij} \cdot e_i\otimes f_j = 0$$

to obtain $c_{\mu\nu} = 0$ for every pair μ and ν of indices. ❑

6.3 *Functorial properties*

(6.23) The tensor product functor $-\otimes_A N$ and the functor $\mathrm{Hom}_A(N,-)$ live in a close partnership; they form what is called *a pair of adjoint functors* in the vernacular of homological algebra; that is, there is an identity as in the following proposition:

PROPOSITION 6.24 (ADJOINTNESS) *There is a functorial isomorphism*

$$\mathrm{Hom}_A(M\otimes_A N, L) \simeq \mathrm{Hom}_A(M, \mathrm{Hom}_A(N, L)).$$

The word "functorial" refers to all three variable. The dependence is covariant in L and contravariant in M and N (a sanity check is that the variances are the same on both sides). The full name of the isomorphism in the theorem would be

$$\theta_{M,N,L}\colon \mathrm{Hom}_A(M\otimes_A N, L) \to \mathrm{Hom}_A(M, \mathrm{Hom}_A(N, L)),$$

but to save the presentation from notational obesity, we shall systematically abbreviate it to θ (think of the parameters always being present but as darkened lights one can turn on, if more clarity is needed).

PROOF: The salient point is that $\mathrm{Hom}_A(M, \mathrm{Hom}_A(N, L))$ is canonically isomorphic to the space of bilinear maps $M \times N \to L$; and once that is realized, the proposition becomes just another reformulation of the universal property of the tensor product.

One may consider members of $\mathrm{Hom}_A(M, \mathrm{Hom}_A(N, L))$ as being maps $\Phi(x, y)$ defined on $M \times N$: Assume $\Phi(x, y)$ is bilinear; when x is a specified member of M, the corresponding map $\Phi(x, -)$ from N to L is given as $y \mapsto \Phi(x, y)$. The other way around, when $\phi\colon M \to \mathrm{Hom}_A(N, l)$ is given, we put $\Phi(x, y) = \phi(x)(y)$. The required linearities and bilineraties of the involved maps are immediate to check.

And that's it: according to the universal property the tensor product enjoys, any such bilinear map Φ can be written unambiguously as $\Phi(x, y) = \phi(x\otimes y)$ with a linear map $\phi\colon M\otimes_A N \to L$.

Then to the functoriality: the hart of the matter is quit prosaic once one sees through the formal underwood. Given $\gamma\colon L \to L'$. Functoriality in this case means that

$$\theta \circ \gamma_* = \gamma_* \circ \theta$$

This equality boils down to the trivial and tautological observation that when evaluated at a pair (x, y), both sides equal $\gamma(\Phi(x, y))$; indeed, the sole difference is that on the left hand side we first apply γ to Φ and then consider the result first a function in x and then in y; whereas on the right hand side order is opposite: we begin with considering Φ first as a function in x and then in y for then to apply γ; of course, they are the same.

Next, suppose that $\alpha\colon M \to M'$ is given; we are then to establish the equality:

$$\theta \circ (\alpha\otimes \mathrm{id}_N)^* = \alpha^* \circ \theta. \tag{6.5}$$

Indeed, both sides equal $\Phi(\alpha(x), y)$, and again this is just an expression for trivial fact that the order of applying α and considering x and y as a first and a second variable,

does not make a difference. Functoriality in N is quite symmetric to functoriality in M; and one has

$$\theta \circ (\mathrm{id}_M \otimes \beta)^* = \beta^* \circ \theta. \tag{6.6}$$

when $\beta\colon N \to N'$ an A-linear map. ❑

Right exactness

In analogy with the notion of left exactness, which we discussed in connection with the hom-functors, a covariant and additive functor F between two module-categories* Mod_A and Mod_B is said to be *right exact* if it transforms exact sequences shaped like

*Or more generally, between two abelian categories.

Right exact functors (høyre eksakte funktorer)

$$M_0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow 0$$

into exact sequences shaped like

$$F(M_0) \longrightarrow F(M_1) \longrightarrow F(M_2) \longrightarrow 0.$$

A fundamental and most useful property of the tensor product is that it is right exact. This section is devoted to giving a proof this, with some easy consequences included at he end.

(6.25) Here it comes:

PROPOSITION 6.26 (RIGHT EXACTNESS) *Given a ring A and an A-module N. The functor $-\otimes_A N$ is a right exact functor.*

Our approach relies on Proposition 6.24 above and illustrates the general fact that adjoint functors tend to share exactness properties; if one is exact in some sense, the other tends to be exact in a related sense. It is possible to give a proof of right exactness based on the construction of the tensor product. This is however tedious, cumbersome and not very enlightening, and according to our mantra should be avoided. Let us also mention that it is common usage to call N a *flat* A-module if the functor $(-)\otimes_A N$ is exact; *i. e.* when it transforms injective maps into injective maps.

Flat modules (Flate moduler)

PROOF: Let the exact sequence

$$M_\bullet : \quad M_0^\alpha \longrightarrow M_1^\beta \longrightarrow M_2 \longrightarrow 0$$

be given; the task is then to show that the sequence

$$M_\bullet \otimes_A N : \quad M_0 \otimes_A N \xrightarrow{\alpha \otimes \mathrm{id}_N} M_1 \otimes_A N \xrightarrow{\beta \otimes \mathrm{id}_N} M_2 \otimes_A N \longrightarrow 0$$

is exact. Our tactics will be to apply the principle of left exactness of the hom-functor as expressed in the proposition LEFT EXACTNESS II (Proposition 5.17 on page 130), and in

fact, we shall do this twice. With that principle in mind, we start out by observing that the sequence $\mathrm{Hom}_A(M_\bullet \otimes_A N, L)$ being exact for every A-module L will be sufficent, and this sequence appears as the upper line in the following grand diagram.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Hom}_A(M_2 \otimes_A N, L) & \xrightarrow{(\beta\otimes \mathrm{id}_N)^*} & \mathrm{Hom}_A(M_1 \otimes_A N, L) & \xrightarrow{(\alpha\otimes \mathrm{id}_N)^*} & \mathrm{Hom}_A(M_0 \otimes_A N, L) \\ & & \wr\downarrow\theta & & \wr\downarrow\theta & & \wr\downarrow\theta \\ 0 & \longrightarrow & \mathrm{Hom}_A(M_2, \mathrm{Hom}_A(N,L)) & \xrightarrow[\beta^*]{} & \mathrm{Hom}_A(M_1, \mathrm{Hom}_A(N,L)) & \xrightarrow[\alpha^*]{} & \mathrm{Hom}_A(M_0, \mathrm{Hom}_A(N,L)) \end{array}$$

The next step is to evoke Proposition 6.24 above and replace $\mathrm{Hom}_A(M_\bullet \otimes_A N, L)$ with the complex $\mathrm{Hom}_A(M_\bullet, \mathrm{Hom}_A(N, L))$; the latter is displayed as the bottom line of the grand diagram. The crux of the proof is that this latter sequence is exact, again by LEFT EXACTNESS II, so once we know that the two rows in the grand diagram are isomorphic (as sequences) we are through. But indeed they are, since with the vertical maps being the canonical isomorphisms from Proposition 6.24 (ADJOINTNESS), the two squares commute according to the functoriality properties stated in Proposition 6.24. ❑

(6.27) Proposition 6.18 on page 154 describes the tensor product of two cyclic module. An analogous result holds true with just one of the modules being cyclic:

PROPOSITION 6.28 *Let $\mathfrak{a} \subseteq A$ be a an ideal and M and A-module. Then one has a canonical isomorphism $M \otimes_A A/\mathfrak{a} \simeq M/\mathfrak{a}M$, which sends $m \otimes [a]$ to $[am]$.*

PROOF: The starting point is the exact sequence

$$0 \longrightarrow \mathfrak{a} \longrightarrow A \longrightarrow A/\mathfrak{a} \longrightarrow 0,$$

which when tensorized by M, yields the exact sequence

$$\mathfrak{a} \otimes_A M \xrightarrow{\alpha} M \longrightarrow M \otimes_A A/\mathfrak{a} \longrightarrow 0,$$

because the tensor product is right exact. The map α sends $a \otimes x$ to ax, hence its image is equal to $\mathfrak{a}M$, and we are done. ❑

EXAMPLE 6.1 Be aware that the tensor product can be a bloodthirsty killer. Injective maps may cease being injective when tensorized, and they can even become zero. The simplest example is multiplication by an integer n, that is; the map $\mathbb{Z} \to \mathbb{Z}$ that sends x to nx. It vanishes when tensorized by $\mathbb{Z}/n\mathbb{Z}$. This also illustrates the fact that the functor $- \otimes_A N$ is not always exact, even though always being right exact. In this example the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow 0,$$

is transformed into the exact sequence (right exactness of $\otimes$)

$$\mathbb{Z}/n\mathbb{Z} \xrightarrow{0} \mathbb{Z}/n\mathbb{Z} \xrightarrow{\beta} \mathbb{Z}/n\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}/n\mathbb{Z} \longrightarrow 0,$$

where β must be an isomorphism—its kernel is zero since the sequence is exact. So part of the conclusion is that $\mathbb{Z}/n\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/n\mathbb{Z}$ which as well ensues from Proposition 6.18 on page 154. ☆

(6.29) Even whole modules may succumb under the action of the tensor product; for instance, we saw that $\mathbb{Z}/p\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}/q\mathbb{Z} = 0$ when p and q are relatively prime integers, which illustrates a general fact. Recall that an A-module M is *divisible* by an element $a \in A$ if the multiplication map $M \to M$ is surjective; in other words every $x \in M$ may be written as ax' for some $x' \in M$.

PROPOSITION 6.30 *Let $a \in A$ and assume that M and N are A-modules such that M is divisible by a and $a \in \operatorname{Ann} N$, then $M\otimes_A N = 0$.*

PROOF: The short argument goes like this. Every $x \in M$ is of the form ax' for some $x' \in M$, so that $x\otimes y = ax'\otimes y = x'\otimes ay = 0$, and as the decomposable tensors generate $M\otimes_A N$, we are through. ❑

Exercises

(6.3) Decompose $\mathbb{Z}/16\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}/36\mathbb{Z}$ and $(\mathbb{Z}/6\mathbb{Z}\oplus\mathbb{Z}/15\mathbb{Z})\otimes_{\mathbb{Z}}(\mathbb{Z}/21\mathbb{Z}\oplus\mathbb{Z}/14\mathbb{Z})$ as direct sums of cyclic groups.

(6.4) Let A be a ring and M an A-module. Show that $M\otimes_A(A/\operatorname{Ann} M) = M$. Show that if N is a second A-module and the annihilators $\operatorname{Ann} M$ and $\operatorname{Ann} N$ are comaximal ideals, then $M\otimes_A N = 0$.

(6.5) Show that $\mathbb{Q}/\mathbb{Z}\otimes_Z\mathbb{Q} = 0$. Show further that it holds true that $\mathbb{Q}\otimes_{\mathbb{Z}}\mathbb{Q} \simeq \mathbb{Q}$, but that one has $\mathbb{Q}/\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Q}/\mathbb{Z} = 0$. HINT: Proposition 6.30.

(6.6) Let $\mathfrak{a}$ be a proper ideal in the ring A. Assume that M is an A-module for which there is surjection $\phi\colon M \to A/\mathfrak{a}$. Show that $\mathfrak{a}M \neq M$. HINT: Tensorize ϕ by $A/\mathfrak{a}$.

(6.7) Let M be finitely generated non-zero module over the ring A.

a) Show that there is a prime ideal $\mathfrak{p}$ in A and a surjective map $M \to A$. HINT: Consider a generating set $\{x_1, \ldots, x_r\}$ with r minimal and use that the quotient module $M/(x_1, \ldots, x_r)$ is cyclic.

b) Show that it holds true that $M^{\otimes n} \neq 0$ for any natural number n. HINT: Exhibit a surjective map $M \to A/\mathfrak{p}$. Proceed by induction on n and right exactness of the tensor product.

(6.8) Assume that G is a finite abelian group. Show that $G \otimes_{\mathbb{Z}} \mathbb{Q} = 0$.

(6.9) Let A be a domain contained in a field K and let M be an A-module. Assume that $\operatorname{Ann} M \neq (0)$. Prove that $K \otimes_A M = 0$.

(6.10) Let $A \to B$ be a surjective ring homomorphism. Show that for any two B-modules M and N (which automatically are A-modules) it holds true that $M \otimes_A N = M \otimes_B N$.

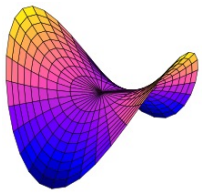
(6.11) Show that $\mathbb{Z}[i] \otimes_{\mathbb{Z}} \mathbb{Z}[i]$ is a free abelian group of rank 4 while $\mathbb{Z}[i] \otimes_{\mathbb{Z}[i]} \mathbb{Z}[i] = \mathbb{Z}[i]$ and is of rank two as an abelian group.

✲ (6.12) Let E and F be vector spaces over the field k, and let $\tau \in E \otimes_k F$ be a tensor. Show tha one may express τ as a finite sum $\tau = \sum_i e_i \otimes f_i$ where the e_i's are linearly independent vectors from E.

(6.13) *Decomposable tensors and a saddle surface.* Only in a very few highly special cases will all elements in a tensor product be decomposable; the usual situation is that most are not. A simple example is $W = V \otimes_k V$ where V is a two-dimensional vector space over k. Let $\{e_1, e_2\}$ be a basis for V. This example also illustrate that beautiful geometry can be unveiled by tensor product considerations.

The tensor product W is of dimension four with a basis $\{e_i \otimes e_j\}$ where $1 \leqslant i, j \leqslant 2$. Let x_{ij} be coordinates relative to this basis; that is, any vector v is expressed as $v = x_{11} \cdot e_1 \otimes e_1 + x_{12} \cdot e_1 \otimes e_2 + x_{21} \cdot e_2 \otimes e_1 + x_{22} \cdot e_2 \otimes e_2$,

a) Establish that he decomposable tensors are shaped like



$$(ue_1 + ve_2) \otimes (se_1 + te_2) = us \cdot e_1 \otimes e_1 + ut \cdot e_1 \otimes e_2 + vs \cdot e_2 \otimes e_1 + vt \cdot e_2 \otimes e_2,$$

with u, v, s and t being scalars.

b) Show that the decomposable tensors are precisely those lying on the subset $x_1 x_4 - x_2 x_3 = 0$.

c) In the real case, that is when $k = \mathbb{R}$, convince yourself that this locus is the cone in $\mathbb{R}^4$ with apex the origin over a saddle-surface in $\mathbb{R}^3$; *i. e.* one given as $z = xy$ (or in our coordinates $x_3 = x_1 x_2$).

(6.14) *The rank stratification.* Let V be a vector space over a field k of finite dimension. The dual space $V^* = \operatorname{Hom}_k(V, k)$ consists of linear functionals on V and is a vector space of the same dimension as V. Chose a basis $\{e_i\}$ for V and let $\{\phi_i\}$ be the *dual* basis for V^*. That is, the functionals ϕ_i are defined as $\phi_i(e_j) = 0$ when $i \neq j$ and $\phi_i(e_i) = 1$.

a) Let W be a second vector space of finite dimension over k with basis $\{f_j\}$. Prove that the assignment $\phi \otimes w$ to $v \mapsto \phi(v)w$ induces an isomorphism $\Gamma \colon V^* \otimes_k W \xrightarrow{\simeq} \operatorname{Hom}_k(V, W)$.

b) Given a linear map $\theta \colon V \to W$ whose matrix relative to the bases $\{e_i\}$ and $\{f_j\}$ is (a_{ij}). Show that the element in $V^* \otimes W$ corresponding to θ equals $\sum_i \phi_i \otimes \theta(e_i)$ and that $\sum_i \phi_i \otimes \theta(e_i) = \sum_{ij} a_{ij} \phi_i \otimes f_j$.

c) Show that the non-zero decomposable tensors in $V^* \otimes_k W$ under the map Γ correspond to the linear maps of rank one. HINT: Chose an appropriate basis for V.

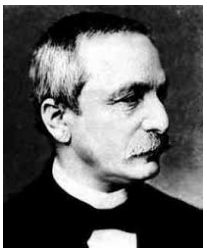
d) Show that a linear map in $\mathrm{Hom}_k(V, W)$ is of at most rank r if and only if the corresponding tensor in $V^* \otimes W$ is the sum of at most r decomposable tensors. HINT: Chose an appropriate basis for V

(6.15) *The Kronecker product.* Let $\phi\colon E \to F$ and $\psi\colon G \to H$ be two A-linear maps between free A-modules of finite rank. Let $\{e_i\}_{i\in I}$, $\{f_j\}_{j\in J}$, $\{g_k\}_{k\in K}$ and $\{h_l\}_{l\in L}$ be bases of E, F, G and H respectively, and let the matrices of ϕ and ψ in the appropriate bases be Φ and Ψ. Show that the matrix of $\phi\otimes\psi$ in the bases $\{e_i\otimes f_j\}_{(i,j)\in I\times J}$ and $\{g_k\otimes h_l\}_{(k,l)\in K\times L}$ is given as the matrix $(\Phi_{ij}\Psi_{kl})$ with rows indexed by pairs $(i,j) \in I \times J$ and colums by $(k,l) \in K \times L$. This matrix is called the *Kronecker product* of Ψ and Φ. ★

The Kronecker product (Kronecker-produktet)

6.4 Change of rings

Working in algebra one frequently finds it necessary, or at least highly desirable, to change the ground ring. For instance, one extends the ground field to have sufficiently many roots of polynomial at hand, or one reduces the ground ring modulo an ideal to make arguments simpler. The tensor product is the perfect tool to carry the modules one studies on to the new ground ring.



Leopold Kronecker (1823–1891)

German mathematician

(6.31) Assume we are given a new ground ring B which is an algebra over the old one A. In particular, B is an A-module. and we may form the tensor product $M\otimes_A B$ of B with any A-module M. This tensor product will be a B-module in a canonical way: multiplication in "in the second variable" of $M\otimes_A B$ produces a B-module structure on $M\otimes_A B$. Indeed, if $b \in B$, the multiplication-by-b-map $x \mapsto bx$ is an A-linear map $[b]\colon B \to B$, and therefore induces a map $\mathrm{id}_M\otimes[b]$. On decomposable tensors it acts as $b \cdot x\otimes c = x\otimes bc$. For a general tensor t we shall, of course, denote the action by $b \cdot t$ or simply by bt.

The module axioms come for free, it ensues directly from functoriality of $-\otimes_A B$, and in view of the tensor product being additive (Proposition 6.11 on page 151) and right exact (Proposition 6.26 on page 157), we arrive at

PROPOSITION 6.32 *Given an A-algebra B. The the change-of-rings-functor, which sends M to $M\otimes_A B$, is a right exact additive functor $-\otimes_A B\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$.*

(6.33) Notice, that elements in B coming from A may be moved past the $\otimes$-sign; *i. e.* if $a \in A$, one has $x\otimes ab = ax\otimes b$. Be aware however, there is a hidden pitfall. For any $b \in B$ the notation $a \cdot b$ is a sloppy version of the correct notation $u(a) \cdot b$, where $u\colon A \to B$ is the structure map and the product is the product in the ring B. Hence $ax\otimes b = x\otimes u(a)b$

would be the correct way of writing. For instance, when A is of positive characteristic p, the map u could be the Frobenius map $a \mapsto a^p$. Then $a \cdot x \otimes b = x \otimes a^p b$

Transitivity and adjointness

(6.34) Sometimes one wants to perform consecutive base changes, and in that respect the tensor product behaves well. It is transitive in the following sense.

PROPOSITION 6.35 (TRANSITIVITY) *Assume that B is an A-algebra and C is a B-algebra. Then there is a canonical isomorphism $(M \otimes_A B) \otimes_B C \simeq M \otimes_A C$ as C-modules.*

PROOF: The short descriptions of the pair of inverse maps are $m \otimes x \otimes y \mapsto m \otimes xy$ and $m \otimes z \mapsto m \otimes 1 \otimes z$. By the principles of bi- and tri-linearity both extend to maps between the tensor products, and acting as mutually inverses on decomposable tensors, they are mutually inverse. ❏

(6.36) Changing the base ring preserves tensor products, but not hom-modules in general. One has the following:

PROPOSITION 6.37 *Let B be an A algebra and M and N two A-modules. Then there is a canonical isomorphism of B-modules*

$$(M \otimes_A N) \otimes_A B \simeq (M \otimes_A B) \otimes_B (N \otimes_A B).$$

In other words, the functor $(-) \otimes_A B$ takes tensor products into tensor products.

PROOF: Two mutually inverse B-module homomorphisms are defined by the assignments

$$x \otimes y \otimes b \mapsto x \otimes 1 \otimes y \otimes b$$
$$x \otimes y \otimes bb' \leftarrow\!\shortmid\, x \otimes b \otimes y \otimes b'$$

of decomposable tensors. The obey respectively a tri-linear and a quadri-linear requirement and thereby define genuine maps between the modules. ❏

(6.38) In addition to the base change functor given by the tensor product

$$(-) \otimes_A B \colon \mathsf{Mod}_A \to \mathsf{Mod}_B,$$

there is a functor going the other way, $(-)_A \colon \mathsf{Mod}_B \to \mathsf{Mod}_A$, whose action is kind of trivial. If N is a B-module, N_A is equal to N but regarded as an A-module—one just forgets the B-module structure. The same happens for maps; they are kept intact, but regarded as being merely A-linear. Such functors that throws away part of a structure, are called *forgetful functors* in the parlance of category theory.

The point is that the tensor product $(-) \otimes_A B$ and the forgetful functor $(-)_A$ are so-called *adjoint functors*:

PROPOSITION 6.39 (ADJOINTNESS) *Given an A-algebra B. Then there is a canonical isomorphism*

$$\mathrm{Hom}_B(M\otimes_A B, N) \simeq \mathrm{Hom}_A(M, N_A).$$

functorial in both M and N.

PROOF: The map from the left hand side to the right hand side is simply a "restriction" map. It sends a given B-linear map $\phi\colon M\otimes_A B \to N$ to $\phi(x\otimes 1)$, which cearly is A-linear in x. To define a map the other way, let $\phi\colon M \to N_A$ be A-linear. The expression $\psi(x\otimes b) = b\phi(x)$ is A-bilinear in b and x and by the universal property enjoyed by the tensor product, it extends to an A-linear map $\psi\colon M\otimes_A B \to N$, which turns out to be B-linear (remember, multiplication by elements from B is performed in the right factor):

$$\psi(b'\cdot x\otimes b) = \psi(x\otimes bb') = bb'\phi(x) = b'\cdot\psi(x\otimes b).$$

Finally, and as usual, the two maps are mutially inverses, agreeing on decomposable tensors. ❑

Maps between free modules and base change

The tensor product being an additive functor, it is clear that the change-of-ring-functors transform free modules to free modules; indeed, if $E \simeq nA$ with a basis $\{e_i\}$ corresponding to the standard basis $\{\epsilon_i\}$ of nA, it holds true that $E\otimes_A B \simeq nB$ with a basis $\{e_i\otimes 1\}$ corresponding to the standard basis $\{\epsilon_i\}$ (as a basis for nB this time).

(6.40) It is of interest to know how changing the base ring affects maps between free modules and how their associated matrices change. So let F be a second free A-module of rank m with a basis $\{f_j\}$ and suppose that $\phi\colon E \to F$ is an A-linear map. Recall that the entires of the matrix $\Phi = (a_{ij})$ associated with ϕ are the coefficients in the developments

$$\phi(e_i) = \sum_j a_{ji} f_j$$

of $\phi(e_i)$ in terms of the basis elements f_j. Applying $\phi\otimes\mathrm{id}_B$ to the basis element $e_i\otimes 1$, yields

$$\phi\otimes\mathrm{id}_B(e_i\otimes 1) = \phi(e_i)\otimes 1 = (\sum_j a_{ji} f_j)\otimes 1 = \sum u(a_{ji})\cdot f_j\otimes 1$$

where $u\colon A \to B$ denotes the structure maps as in paragraph 6.33 above. Hence the matrix of $\phi\otimes\mathrm{id}_B$ is the matrix $(u(a_{ji}))$ obtained by applying the structure map u to the entries of (a_{ji}).

Examples

(6.2) As an example consider the polynomial ring $A = \mathbb{C}[x]$ and let $a \in \mathbb{C}$ be a complex number. Furthermore, let $B = \mathbb{C}[x]/(x-a) \simeq \mathbb{C}$, so that the structure map u is the

evaluation at a; *i. e.* $u(P(x)) = P(a)$. If ϕ is a map between two free $\mathbb{C}[x]$-modules with matrix $P = (P_{ij}(x))$ relative to some bases, the matrix of $\phi\otimes\mathrm{id}[\mathbb{C}]$ is just the matrix P evaluated at a; that is, the matrix $(P_{ij}(a))$.

(6.3) For a second example, take $A = \mathbb{Z}$ and $B = \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ for some prime number p. If ϕ is a map between two free abelian groups whose matrix relative to some bases is (n_{ij}), the matrix of $\phi\otimes\mathrm{id}[\mathbb{F}_p]$ relative to the corresponding bases will be $([n_{ij}])$ (where $[n]$ as usual denotes the congruence class of an integer n modulo p).

On the other hand, changing the ring from $\mathbb{Z}$ to $\mathbb{Q}$; that is, passing to the map $\phi\otimes\mathrm{id}[\mathbb{Q}]$, does not change the matrix. We merely consider the integers n_{ij} as being rational numbers!

(6.4) Our third example will be of a rather different flavour than the two previous ones. This time we let A be a ring of characteristic p. The Frobenius map $a \to a^p$ is then a ring homomorphism $A \to A$ and gives A an alternative A-algebra structure in which $a \cdot x = a^p x$ (the product to the left is the *new* product, whereas the one to the right is the original product in A). In view of the considerations above, a ring-change via the Frobenius map, changes matrices of A-linear maps between free A- modules by rising their entries to the p-th power.

☆

6.5 *Tensor products of algebras*

Setting the stage of this section we let C and introduce two C-algebras A and B. The star of the show will be the tensor product $A\otimes_C B$ and the objective of the play is to give it a ring structure compatible with the underlying C-module structure, thus making it a C-algebra.

(6.41) On decomposable tensors the product ought to abide by the rule

$$a\otimes b \cdot x\otimes y = ax\otimes by, \tag{6.7}$$

and indeed, this extends to a product on $A\otimes_C B$:

PROPOSITION 6.42 *Given a ring C and two C-algebras A and B. Then there is a unique C-algebra structure on $A\otimes_C B$ whose product on decomposable tensors satisfies $a\otimes b \cdot a'\otimes b' = aa'\otimes bb'$.*

Notice that, *a priori*, the C-module structure on $A\otimes_C B$ is in place, so only the ring structure is lacking.

PROOF: To give an argument that the assignment in (6.7) can be extended to give a product of arbitrary tensors, we appeal once more to the principle of bilinearity (at the end of paragraph 6.2 on page 150). In fact, we shall apply it twice, once for each factor,

the basic observation being that the right hand side of (6.7) is C-bilinear both in (a, b) and (x, y)

The first application of the principle shows that multiplication by $a\otimes b$ for a fixed pair (a, b) extends to a C-linear map $A\otimes_C B \to A\otimes_C B$. This yields a map

$$\eta_0 \colon A \times B \to \mathrm{Hom}_C(A\otimes_C B, A\otimes_C B)$$

that sends (a, b) to the multiplication-by-$a\otimes b$-map; in other words, it holds true that $\eta_0(a, b)(\sum_i x_i\otimes y_i) = \sum_i ax_i\otimes by_i$. In its turn, this map depends bilinearily on the pair (a, b), and by a second application of the principle we arrive at a C-linear map

$$\eta \colon A\otimes_C B \to \mathrm{Hom}_C(A\otimes_C B, A\otimes_C B),$$

which on decomposable tensors behave as desired; *i. e.* $\eta(a\otimes b)(x\otimes y) = ax\otimes by$. Subsequently the product of two arbitrary tensors s and t is defined as $s \cdot t = \eta(t)(s)$. This establishes the product in $C\otimes_A B$, but there are of course, verifications to be done.

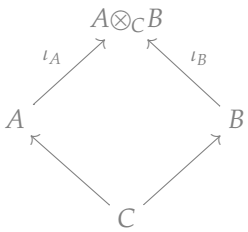
That the ring axioms hold, is a matter of straightforward verifications —they follow by the uniqueness parts of the principles of bilinearity and trilinearity. For example, both expressions $\eta(t)(s)$ and $\eta(s)(t)$ are bilinear in s and t and since they agree on decomposable tensors, they are equal, and hence the product is commutative. One checks associativity in a similar manner, but by using the Trilinearity Principle (Lemma 6.16 on page 153). The two expressions $\eta(tu)(s)$ and $\eta(t)(us)$ are both linear in each of the variables s, t and u, and agreeing on decomposable tensors, they coincide; that is, $(t \cdot u) \cdot s = t \cdot (u \cdot s)$. ❑

EXERCISE 6.16 Check that the distributive law holds in $A\otimes_C B$. ★

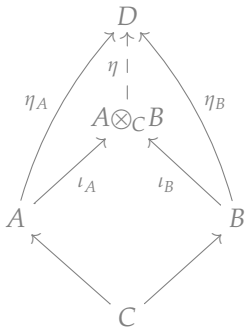
The universal property

The tensor product $A\otimes_C B$ enjoys a universal property that plays a paramount role in algebraic geometry. In algebraic terms it reflects the geometric construction of so called *fibre products*; a simple variant of which are the products $X \times Y$ of two schemes. This is a foundational construct on which the whole theory of algebraic geometry rests.

(6.43) With the setting as in the previous section, there are two canonical C-algebra homomorphisms having target $A\otimes_C B$; one with A as source and the other sourced at B.



The first, call it ι_A, is given as $a \mapsto a\otimes 1$ and the second, call it ι_B, as $b \mapsto 1\otimes b$. Elements from C may be moved past the tensor product sign so that $c\otimes 1 = 1\otimes c$ for $c \in C$; or expressed in terms of a diagrams: The diagram in the margin commutes. (Where the maps $C \to A$ and $C \to B$ are the structure maps defining the C-algebra structures.) Moreover, the tensor product is universal among C-algebras living in such diagrams:



PROPOSITION 6.44 (THE UNIVERSAL PROPERTY) *With the notation as just introduced, assume given a C algebra D and two C-algebra homomorphisms $\eta_A\colon A \to D$ and $\eta_B\colon B \to D$. Then there is a unique map of C-algebras $A\otimes_C B \to D$ such that $\eta\circ\iota_A = \eta_A$ and $\eta\circ\iota_B = \eta_B$.*

PROOF: Indeed, the expression $\eta_A(a)\eta_B(b)$ is C-bilinear, and according to the universal property of the tensor product it gives rise to a C-linear map $A\otimes_C B \to D$ satisfying $\eta(a\otimes b) = \eta_A(a)\eta_B(b)$. This is our desired map, but some checking remains to be done. Let us begin with verifying that η respects products. Since we know that η is linear, it will suffice to do this for decomposable tensors:

$$\eta(aa'\otimes bb') = \eta_A(aa')\eta_B(bb') = \eta_A(a)\eta_A(a')\eta_B(b)\eta_B(b') = \eta(a\otimes b)\eta(a'\otimes b'),$$

where the two extreme equalities hold true by the very definition of η and the middle one because both η_A and η_B are ring maps.

Next, one has $\eta\circ\iota_A = \eta_A$ and $\eta\circ\iota_B = \eta_B$ since $\eta_A(1) = \eta_B(1) = 1$, and finally, that η is unique follows, since it is determined by the values on decomposable tensors, and these satisfy

$$\eta(a\otimes b) = \eta((a\otimes 1)(1\otimes b)) = \eta(a\otimes 1)\eta(1\otimes b) = \eta_A(a)\eta_B(b).$$

❑

Base change of polynomial rings and algebras of finite type

We continue with the stage set as above, with B being an A-algebra through the structure map $u\colon A \to B$.

(6.45) A natural question is how polynomial rings behave under base change, and the answer is they do in the obvious and simplest way.

There is a map of A-algebras $A[x_1,\dots,x_r] \to B[x_1,\dots,x_r]$ sending x_i to x_i and hence a polynomial $\sum_\alpha a_\alpha x^\alpha$ is mapped to $\sum_\alpha u(a_\alpha)x^\alpha$. Together with the inclusion of B as the constants in $B[x_1,\dots,x_r]$, it induces, in view of the universal property of the tensor product, a map of A-algebras $A[x_1,\dots,x_r]\otimes_A B \to B[x_1,\dots,x_r]$.

LEMMA 6.46 *Let A be a ring and B be an A-algebra. Then the following equality holds true $A[x_1,\dots,x_r]\otimes_A B = B[x_1,\dots,x_r]$*

Recall the notation from Paragraph 1.16 on page 21 where the monomial $x_1^{\alpha_1}\cdot\ldots\cdot x_r^{\alpha_r}$ was denoted by x^α, with α being the multi-index $(\alpha_1,\dots,\alpha_r)$.

PROOF: Considered as A-module, the polynomial ring $A[x_1,\dots,x_r]$ is free, and the monomials x^α, with α running through all multi-indices, form a basis. The same holds for $B[x_1,\dots,x_r]$; the monomials x^α form a B-basis. The lemma then follows since these monomials correspond. ❑

LEMMA 6.47 *Let $\mathfrak{a}\subseteq A[x_1,\dots,x_r]$ be an ideal. Then the following equality holds true*

$$A[x_1,\dots,x_r]/\mathfrak{a}\otimes_A B = B[x_1,\dots,x_r]/\mathfrak{a}B[x_1,\dots,x_r].$$

PROOF: Suppose that $\mathfrak{a}$ is generated by a collection $\{f_i\}$ of polynomials. Then the ideal $\mathfrak{a}B[x_1,\ldots,x_r]$ will be generated by the images of f_i's under the natural map $A[x_1,\ldots,x_r] \to B[x_1,\ldots,x_r]$ that is the polynomials obtained by applying the structure map u to the coefficients. Since the tensor product is right exact, the lemma follows. ❑
In several contexts, when *e.g.* u is a canonical inclusion, as $\mathbb{Z}\subseteq\mathbb{Q}$ or $\mathbb{Q}\subseteq\mathbb{C}$, the effect on the generators is nil, one just considers the generators as being members of $B[x_1,\ldots,x_r]$. However, in other situations the effect on the f_i's can be quite dramatic, in the worst case they can even vanish. For examples this occurs if the structure map is the reduction mod p-map $u\colon \mathbb{Z}\to\mathbb{Z}/p\mathbb{Z}=\mathbb{F}_p$, and the coefficients of $f_i\in\mathbb{Z}[x]$ are all divisible by p.

Examples

(6.5) A particular application of the previous lemma is that the tensor product of two polynomial rings (over the base ring) again is a polynomial ring, that is, one has

$$A[x_1,\ldots,x_r]\otimes_A A[y_1,\ldots,y_s] = A[x_1,\ldots,x_r,y_1,\ldots,y_s].$$

Polynomials give a striking example that the decomposable tensors are scarce. In $A[x,y]$ for instance, the decomposable tensors are the polynomials that factor as a product $p(x)q(y)$ of which there are few compared to the total collection of polynomials.

(6.6) Be aware that the tensor product of two integral domains need not be an integral domain. Even if both factors are fields, the tensor product might acquire zero-divisors. A simple example is $\mathbb{C}\otimes_{\mathbb{R}}\mathbb{C}$. The complex numbers $\mathbb{C}$ can be described as the quotient $\mathbb{R}[x]/(x^2+1)$ so by Lemma 6.47 above it holds that $\mathbb{C}\otimes_{\mathbb{R}}\mathbb{C}=\mathbb{C}[x]/(x^2+1)$. This latter ring is isomorphic to the direct product $\mathbb{C}\times\mathbb{C}$, sending the residue class of a polynomial $p(x)$ to the pair $(p(i),p(-i))$ yields an isomorphism, and the direct product has many zero divisors.

☆

Exercises

(6.17) Let k be a field and let $f(x)$ be an irreducible polynomial in $k[x]$ so that $K=k[x]/(f(x))$ is field. Moreover let L be field extension of k in which f splits as a product $f(x)=f_1(x)\ldots f_r(x)$ of irreducible polynomials. Use the Chinese Remainder Theorem to prove that $K\otimes_k L\simeq\prod L_i$ where L_i is the field $L_i=L[x]/(f_i(x))$.

(6.18) Assume that K is an algebraic number field; *i. e.* a finite extension of the field $\mathbb{Q}$ of rational numbers. Show that $K\otimes_{\mathbb{Q}}\mathbb{R}$ is isomorphic to a product of fields each being isomorphic either to $\mathbb{R}$ or $\mathbb{C}$. Show that $[K:\mathbb{Q}]=r_1+2r_2$ where r_1 denotes the number of factors isomorphic to $\mathbb{R}$ and r_2 the number of factors isomorphic to $\mathbb{C}$.
HINT: By the Primitive Element Theorem one may assume that $K=\mathbb{Q}[x]/(f(x))$ where $f(x)\in\mathbb{Q}[x]$ is irreducible polynomial.

(6.19) Show that $A=\mathbb{R}[x,y]/(x^2+y^2)$ is an integral domain, but that $A\otimes_{\mathbb{R}}\mathbb{C}$ is not.

(6.20) Let $\phi\colon A\to B$ be a ring homomorphism and $\mathfrak{p}\subseteq A$ a prime ideal. Show that

the fibre of $\phi^*\colon \operatorname{Spec} B \to \operatorname{Spec} A$ over $\mathfrak{p}$ is naturally homeomorphic to the spectrum $\operatorname{Spec} k(\mathfrak{p})\otimes_A B$ where $k(\mathfrak{p})$ is the fraction field of $A/\mathfrak{p}$; that is, $k(\mathfrak{p}) = A_{\mathfrak{p}}/\mathfrak{p}A_p$

(6.21) Let K be a field of positive characteristic p and let $t \in K$ be an element which is not a p-th power. Show that $L = K[x]/(x^p - t)$ is a field and that $L\otimes_K L$ has non-trivial nilpotent elements.

(6.22) Let k be a field of positive characteristic p, and let $\mathfrak{a}$ be an ideal in $A = k[x_1, \ldots, x_r]$ generated by polynomials $f_i(x) = \sum_\alpha a_{i\alpha}x^\alpha$. Let $F\colon k \to k$ be the Frobenius map $a \mapsto a^p$; and let k_F denote the field k endowed with the k-algebra structure induced by the Frobenius map; that is, members a of k act on k_F as $a \cdot x = a^p x$, Show that $(k[x_1, \ldots, x_r]/\mathfrak{a})\otimes_k k_F \simeq k[x_1, \ldots, x_r]/\mathfrak{a}_F$ where $\mathfrak{a}_F$ is the ideal generated by the polynomials $f_i = \sum_\alpha a_{i\alpha}^p x^\alpha$

(6.23) Let $\bar{\mathbb{C}}$ be $\mathbb{C}$ equipped with the alternative algebra structure induced by complex conjugation; *i. e.* $a \cdot z = \bar{a}z$. Let $f(x) \in \mathbb{C}[x]$ be a polynomial. Describe $\mathbb{C}[x]/(f(x))\otimes_{\mathbb{C}}\bar{\mathbb{C}}$. When are the $\mathbb{C}$-algebras $\mathbb{C}[x]/(f(x))\otimes_{\mathbb{C}}\bar{\mathbb{C}}$ and $\mathbb{C}[x]/(f(x))$ isomorphic?

6.6 *Appendix: Flatness*

Flat modules (flate moduler)

Recall that an A-module M is *flat* if the functor $(-)\otimes_A M$ is exact. This functor is right exact, as we established in Proposition 6.26 on page 157, so it being exact amounts to it sending injections to injections. There is however a small, but important, point: it suffices to consider injections between *finitely generated* modules.

PROPOSITION 6.48 *Let A be a ring and M an A-module. Then the following three statements are equivalent*

i) M is flat;

ii) For all injective maps $\phi\colon N \to N'$ between two A-modules, the induced map $\phi\otimes\mathrm{id}_M\colon N\otimes_A M \to N\otimes_A M$ is injective;

iii) For all injective maps $\phi\colon N \to N'$ where N and N' are finitely generated A-modules, the map $\phi\otimes\mathrm{id}_M\colon N\otimes_A M \to N\otimes_A M$ is injective.

PROOF: We already remarked that *i)* and *ii)* are equivalent, and trivially *ii)* implies *iii)*, and we only need to show that *iii)* implies *ii)*.

So assume that $\phi\colon N \to N'$ is injective and that $x = \sum_{1\leqslant i\leqslant r} x_i\otimes y_i \in N\otimes_A M$ maps to zero in $N'\otimes_A M$; that is, the relation $\sum_i \phi(x_i)\otimes y_i = 0$ holds in $N'\otimes_A M$. To be able to apply *iii)* we shall replace N and N' with appropriate finitely generated modules.

The substitute for N is easy, the submodule generated by the x_i's (which are finite in number) will do, so henceforth we may assume that N is finitely generated. To find a replacement of N' we appeal to Paragraph 6.6 on page 149, which furnishes finitely generated submodules $N'_0 \subseteq N'$ and $M_0 \subseteq M$ such that $\sum_i \phi(x_i)\otimes y_i = 0$ in $N'_0\otimes_A M_0$.

Now, clearly $N_1' = N_0' + \phi(N)$ is a finitely generated submodule of N' in which ϕ takes values, and moreover, by construction, $\sum_i \phi(x_i)\otimes y_i = 0$ in $N_1'\otimes_A M$. By assumption assertion *iii)* holds, and we may infer that $\sum_i x_i\otimes y_i = 0$, and we are done. ❑

(6.49) When using Proposition 6.48 above to check that a module is flat, one may restrict oneself to consider injections of ideals into the ring. Clearly if $\mathfrak{a}\subseteq A$, tensorizing the sequence

$$0 \longrightarrow \mathfrak{a} \xrightarrow{\ \iota\ } A \longrightarrow A/\mathfrak{a} \longrightarrow 0$$

by a module M and remembering that $A/\mathfrak{a}\otimes_A M = M/\mathfrak{a}M$ and $A\otimes_A M = M$, one obtains the exact sequence

$$\mathfrak{a}\otimes_A M \xrightarrow{\iota\otimes \mathrm{id}_M} M \longrightarrow M/\mathfrak{a}M \longrightarrow 0,$$

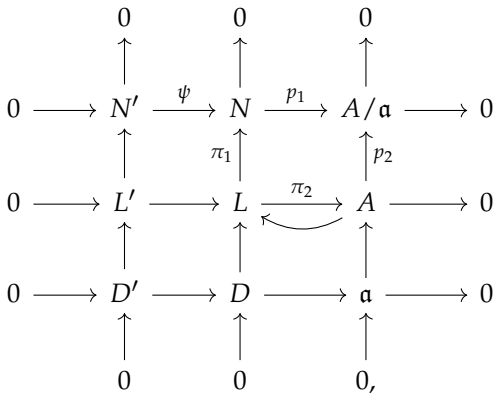
where $\iota\otimes\mathrm{id}_M$ has the effect $a\otimes x \mapsto ax$. It ensues that when M is flat, the map $\iota\otimes\mathrm{id}_M$ will give an isomorphism $\mathfrak{a}\otimes_A M \simeq \mathfrak{a}M$, and the theme of this paragraph is that the converse also holds:

PROPOSITION 6.50 *The A-module M is flat if and only if $\mathfrak{a}\otimes_A M \simeq \mathfrak{a}M$ for all ideals $\mathfrak{a}$. Moreover, it suffices to consider ideals that are finitely generated.*

The modern standard proof of this result uses the derived functors of the tensor product (*i. e.* the functors $\mathrm{Tor}_i^A(-,-)$). However a direct proof is not difficult, it involves merely a diagram-hunt in a grand diagram; and of course, it is nothing but the tiny relevant part extracted from the big machine that makes the Tor-functors function. The reduction to the case of finitely generated ideals follows *mutatis mutandis* the proof of Proposition 6.48 above, and we will not repeat it.

PROOF: Given an injection $\psi\colon N' \to N$ between two finitely generated modules. We may assume that N requires just one more generator than N'; indeed, if this is not the case, we may factor ψ as the composition of two injections $N' \to N'' \to N$ where the numbers of generators of N'' lies strictly between those of N' and N. By functoriallity $\psi\otimes\mathrm{id}_M$ is the composition of the tensorized injections $N'\otimes_A M \to N''\otimes_A M \to N\otimes_A M$, which both may be assumed injective by induction on the difference of numbers of generators. Hence we may concentrate on injections whose cokernel is cyclic, *i. e.* is shaped like $A/\mathfrak{a}$.

Consider the grand diagram

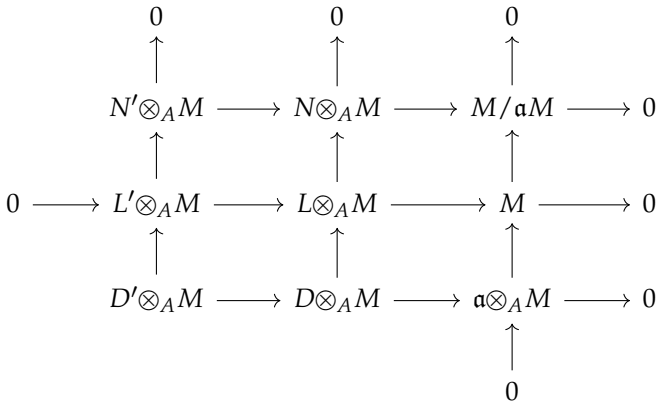


where the upper sequence is the one we started with, and the rightmost one is the canonical one. The hub of the diagram is the central module L, which is the so-called *fibered product* of N and A. It is the submodule of the product $N \times A$ where the two maps p_1 and p_2 coincide; that is, it is given as

Fibered products (fibrerte produkter)

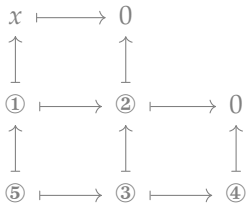
$$L = \{\, (x, y) \in N \times A \mid p_1(x) = p_2(y) \,\}.$$

Filling out the rest of the diagram is easy: we just put $L' = \ker \pi_2$, $D = \ker \pi_1$ and $D' = D \cap L'$. Notice that, and this is the salient point that makes the proof work, because A is free, the middle horizontal sequence is split exact and therefore remains exact after being tensorized by M. The tensorized diagram appears as



where the rightmost sequence is exact by the assumption that $\mathfrak{a} \otimes_A M = \mathfrak{a}M$ and, as observed above, the middle horizontal one is exact. The finish of the proof is either an adaption of the snake lemma (the snake is sneaking to the right and upwards), or a

direct diagram-hunt, staring with an element $x \in N' \otimes_A M$, mapping to zero in $N \otimes_A M$ and following the path indicated in the margin. ❑



EXAMPLE 6.7 *Flat modules over* PID*'s :* The following criterion for when modules over a PID are flat is an illustrating example of the use of the criterion in Proposition 6.50. It applies *e.g.* to $\mathbb{Z}$-modules and modules over the ring $k[x]$ of polynomials with coefficients in field. Recall that module M over a ring A is *torsion free* if $ax = 0$, with $x \in M$ and $a \in A$, implies that either $x = 0$ or a is a zero divisor (that is, $a = 0$ if A is a domain).

PROPOSITION 6.51 *A module M over a principal ideal domain A is flat if and only if it is torsion free.*

PROOF: An arbitrary ideal $\mathfrak{a}$ in A will be principal, say generated by t. All elements of the tensor product $\mathfrak{a} \otimes_A M$ are then of the form $t \otimes x$, and the map $\mathfrak{a} \otimes_A M \to M$ acts by sending $t \otimes x$ to tx. This map is obviously injective if and only if t does not kill any non-zero element from M, and this holds for all ideals (t) if and only if M is torsion free. ❑

☆

Exercises

(6.24) Show that the direct sum $\bigoplus_{i \in I} M_i$ of a familliy $\{M_i\}_{i \in I}$ of A modules is flat if and only if all the M_i's are flat. Conclude that free modules are flat, and hence projective modules will be flat as well.

(6.25) Show that polynomial ring $A[t]$ is a flat algebra over A.

(6.26) Let A be a ring and assume given a short exact sequence

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

of A-modules where M'' is flat. Prove that if one of M or M' is flat, then the other one is also flat. Give an example where M' and M are flat, but M'' is not.

(6.27) Show that any direct product $\prod_{i \in I} \mathbb{Z}$ of copies of $\mathbb{Z}$, is flat over $\mathbb{Z}$.

(6.28) Let $\mathfrak{a} \subseteq A$ be an ideal. Show that $A/\mathfrak{a}$ is flat over A if and only if $\mathfrak{a}^2 = \mathfrak{a}$.

★

Lecture 7

Localization

Very early in our mathematical carrier, if not in our lives, we were introduced to *fractions*, so we should be well familiar with their construction and have their properties in the backbone. Anyhow, recall that to every pair of integers m and n with $n \neq 0$, one forms the 'fraction' m/n. Two such fractions m'/n' and m/n are considered equal—that is, have the same numerical value—precisely when $nm' = n'm$. The fractions, or the rational numbers as we call them, are entities *per se* and not only results of division: formally, they are equivalence classes of pairs (m, n) with respect to the equivalence relation above. The fractions obey the familiar rules for adding and multiplying we learned in school, and they form a field, the field $\mathbb{Q}$ of rational numbers.

There is a simple and very general version of this construction. It gives us the freedom to pass to rings were *a priori* specified elements become invertible. Virtually any set of elements can be inverted; there is merely one natural constraint. If s and t occur as denominators, their product st will as well; indeed, one has $s^{-1}t^{-1} = (st)^{-1}$. Hence the natural notion is the concept of *multiplicatively closed sets*.

The process is indeed very general. It even accepts zero divisors as denominators, but it will then be murderous: if a is a zero divisor, say $a \cdot b = 0$ with $b \neq 0$, and a becomes inverted, b gets killed; indeed, it will follow that $b = a^{-1} \cdot a \cdot b = a^{-1} \cdot 0 = 0$. In principle, one can even push this as far as inverting 0, which however will be devastating: the entire ring collapses to the null-ring.

There are several ways of defining these localized rings. We shall follow most text books and mimic the way one constructs the rational numbers. This is a direct and intuitive construction which does not require much machinery.

The name "localization" has its origin in geometry where one considers rings of functions, say continuous functions on a topological space X. If $U \subseteq X$ is an open set, every function whose zeros all lie in the complement $X \backslash U$ of U, becomes invertible when restricted to U; hence one obtains many functions on U by inverting certain functions on X. In general, far from all are shaped like that, but in special situations,

important in algebraic geometry, all algebraic functions on U arise in this way.

7.1 *Localization of rings*

We start out with introducing the notion of multiplicatively closed sets, and proceed to construction the localized rings. They will be characterized by a universal property. Core examples will be given, and their basic properties will be established, notably the relation between ideals in the ring A and ideals in the localizations $S^{-1}A$.

Multiplicatively closed sets

Multiplicatively closed sets (multiplikativt lukkede mengder)

(7.1) The notion of a multiplicatively closed set were already introduced in the formulation of the Fundamental Existence Theorem (Theorem 2.49 on page 49), but we remind you that a subset S of a ring A is a *multiplicatively closed set*, or for short also called a *multiplicative set*, if it contains the unit element, and the product of every two elements from S belongs to S. That is, the following two conditions are satisfied

- $1 \in S$;
- If $s, t \in S$, then $st \in S$.

Examples

Examples of multiplicative sets abound, but for the moment we only mention a few of the more important ones.

(7.1) The set of all powers a of an element in A; that is, the set $S = \{\, a^n \mid n \in \mathbb{N}_0 \,\}$, obviously is multiplicatively closed.

(7.2) The complement $S = A\backslash\mathfrak{p}$ of any *prime ideal* $\mathfrak{p}$ in A is multiplicatively closed; indeed, from $st \notin S$ ensues that $st \in \mathfrak{p}$ and at least one of the factors s or t belongs to $\mathfrak{p}$; that is, it does not lie in S. In fact an ideal $\mathfrak{a}$ is prime if and only if the complement $A\backslash\mathfrak{a}$ is multiplicatively closed. This argument generalizes immediately and shows that the complement $S = A\backslash\bigcup_{i\in I}\mathfrak{p}_i$ of a union of prime ideals (finite or not) will be a multiplicative set. Indeed, if $st \notin S$, there is at least one index i so that $st \in \mathfrak{p}_i$; consequently either s or t belongs to $\mathfrak{p}_i$ and hence not to S.

(7.3) It is fairly clear that the intersection of any family of multiplicatively closed sets is multiplicatively closed, and one may therefore speak about the *the multiplicative set generated* by a subset T of A. It equals the intersection of all multiplicatively closed sets containing T, and one convinces oneself on the spot that its elements are all finite products of elements from T. So for example, the multiplicative set in $\mathbb{Z}$ generated by 2 and 5 consists of all numbers of the form 2^a5^b, with $a, b \in \mathbb{N}_0$.

(7.4) An occasionally useful multiplicatively set is the set $S = 1 + \mathfrak{a}$ where $\mathfrak{a}$ is an ideal in A.

☆

The construction of the localization $S^{-1}A$

The construction of the localized ring $S^{-1}A$ follows *grosso modo* the same lines as the construction of the rational numbers from the integers, but with a necessary twist due to the possible presence of zero divisors in S, which has the serious consequence that the cancellation law does not hold, and this complicates the matter.

(7.2) A fraction has an enumerator and a denominator, and in our context the latter will be confined to S. A natural starting point is therefore the Cartesian product $A \times S$ with the first factor representing all possible enumerators and the second all possible denominators. The next step is to introduce an equivalence relation on $A \times S$ telling when two fractions are to be considered equal, and inspired by the case of rational numbers, we declare the pairs (a, s) and (b, t) to be equivalent when for some $u \in S$, it holds true that $u(at - bs) = 0$; the factor u is necessary to resolve problems possible zero divisors would cause.

The salient point of the construction is that this is an equivalence relation. We shall (temporarily) write $(a, s) \sim (b, t)$ when (a, s) and (b, t) are equivalent. The relation is obviously reflexive and symmetric. To see it is transitive assume given three pairs (a, s), (b, t) and (c, u) such that $(a, s) \sim (b, t)$ and $(b, t) \sim (c, u)$; transcribing the equivalences into equalities in A, we find that $v(at - bs) = 0$ and $w(bu - ct) = 0$ for some elements $v, w \in S$. Since

$$t(au - cs) = u(at - bs) + s(bu - ct),$$

we infer that

$$vwt(au - cs) = wu(v(at - bs)) + sv(w(bu - ct)) = 0.$$

From S being multiplicatively closed it ensues that $vwt \in S$, and so $(a, s) \sim (c, u)$. We let the *localization* $S^{-1}A$ of A in S be the set of equivalence classes $A \times S/\sim$, and denote by a/s or as^{-1} the class of the pair (a, s).

Localization (lokalisering)

The next task is to give a ring structure to $S^{-1}A$, and there is no hocus-pocus about that, it is done by the familiar formulas for adding and multiplying fractions we know from school:

$$a/s + b/t = (at + bs)/st \qquad a/s \cdot b/t = ab/st. \tag{7.1}$$

However, some checking is necessary. First of all, the definitions in (7.1) are expressed in terms of representatives of equivalence classes, and it is paramount that they do not dependent on which representatives are used. Secondly, the ring axioms must be verified. Once we know the definitions are legitimate, this is just straightforward high school algebra, safely left to volunteering students.

Let us check that the sum is well defined, leaving the product to the eager students. Notice that it will suffices to vary the representatives of one of the addends at the time;

so assume that $(a,s) \sim (a',s')$; *i. e.* it holds that $u(as' - a's) = 0$ for some $u \in S$. We find

$$s't(at + bs) - st(a't + bs') = t^2(as' - a's)$$

which is killed by u. Therefore the sum does not depend on the representative of the first addend, and by symmetry, neither on the representative of the second. Consequently the sum is well defined.

EXERCISE 7.1 Show that the product is well defined. On a rainy day when all your friends are away, verify the ring axioms for $S^{-1}A$. ★

localization map (lokaliseringsavbildning)

(7.3) There is a canonical map $\iota_S \colon A \to S^{-1}A$, which is called the *localization map*. It is nothing but the map that sends an element a in A to the class of the pair $(a,1)$; that is, a is mapped to the fraction $a/1$. By the very definition in (7.1) of the sum and the product in the localization $S^{-1}A$, this a ring homomorphism. Seemingly, this map does nothing to a, but kill it if necessary, and our next proposition details this murderous behaviour of ι_S.

PROPOSITION 7.4 *All elements in $S^{-1}A$ are of the form a/s. It holds true that $\iota_S(a) = 0$ if and only if a is killed by some element from S; i. e. if and only if there is an $s \in S$ such that $sa = 0$.*

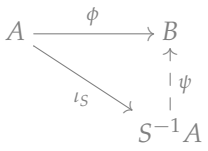
PROOF: By definition, every element in $S^{-1}A$ is an equivalence class a/s. The zero element in $S^{-1}A$ is represented by the pair $(0,1)$ and $\iota_S(a)$ by the pair $(a,1)$. Hence $\iota_S(a) = 0$ if and only if $s \cdot (a \cdot 1 - 0 \cdot 1) = 0$ for an $s \in S$; that is, if and only if $s \cdot a = 0$ for an $s \in S$. ❑

(7.5) When S contains no zero divisors, the map ι_S will be injective, and we shall identify A with its image in $S^{-1}A$. This simplifies the notation significantly. We may safely write a instead of $\iota_S(a)$ or $a/1$, and the inverse image $\iota_S^{-1}(\mathfrak{b})$ of an ideal (or any subset for that matter) will simply be the intersection $A \cap \mathfrak{b}$.

The universal property

(7.6) The localized ring $S^{-1}A$ together with the localization map ι_s are characterized by a universal property, which loosely may by phrased by saying that any ring map with source A that maps elements from S to units, extends* to a ring map from $S^{-1}A$.

**The verb extend is suggestive, but not entirely pertinent when ι_S is not injective.*



PROPOSITION 7.7 *Let A be a ring and $S \subseteq A$ a multiplicative subset. Assume given a ring map $\phi \colon A \to B$ that sends S into the group of units in B. Then there is a unique map of rings $\psi \colon S^{-1}A \to B$ so that $\phi = \psi \circ \iota_S$.*

PROOF: The sole way of realizing a ring map $\psi \colon S^{-1}A \to B$ extending ϕ is to put

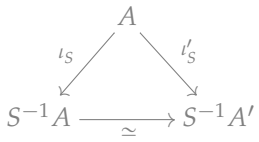
$$\psi(a/s) = \phi(a) \cdot \phi(s)^{-1}, \tag{7.2}$$

and the gist of the proof is that this is a legitimate definition, *i. e.* that $\phi(a) \cdot \phi(s)^{-1}$ is independent of the chosen representative (a,s). But from $a/s = a'/s'$ ensues that

$t \cdot (as' - sa') = 0$ for an element $t \in S$, and hence, since ϕ is map of rings, that

$$\phi(t) \cdot \big(\phi(a) \cdot \phi(s') - \phi(a') \cdot \phi(s)\big) = 0.$$

The element $\phi(t)$ is invertible by assumption, and we conclude that $\phi(a) \cdot \phi(s)^{-1} = \phi(a') \cdot \phi(s')^{-1}$. That $\phi = \psi \circ \iota_S$, is trivial, and that ψ is a homomorphism follows directly from (7.2) and the usual formulas for products and sums of fractions (the formulas in (7.1)). ❑



As any other object characterized by a universal property, the pair ι_S and $S^{-1}A$ is unique up to an unambiguous isomorphism: if $\iota'_S \colon A \to S^{-1}A'$ solves the same universal problem, one has $\iota'_S = \psi \circ \iota_S$ for a unique isomorphism $\psi \colon S^{-1}A \to S^{-1}A'$.

Examples

(7.5) We have not excluded that 0 lies in S. In this case however, the localized ring will be the null-ring since 0 becomes invertible. This situation occurs *e.g.* when S has nilpotent members.

(7.6) An simple situation to have in mind is when A is contained in a field K. The localized ring $S^{-1}A$ is then just the subring $A[s^{-1}|s \in S]$ of K generated by the inverses of members of S. The elements of this ring are all shaped like as^{-1}; indeed, every sum $\sum_i a_i s_i^{-1}$ can be rendered on this form with s a common denominator of the terms. The universal property of the localized ring $S^{-1}A$ then immediately gives a map of rings $S^{-1}A \to A[s^{-1}|s \in S]$ which one easily checks is an isomorphism using the description of $S^{-1}A$ in Proposition 7.4 on page 176.

(7.7) The ring of integers $\mathbb{Z}$ within the field rationals $\mathbb{Q}$ is a particular instance of the situation in the previous example. When S is the set of all powers of a given number p, that is, $S = \{\, p^n \mid n \in \mathbb{N}_0 \,\}$, the ring $S^{-1}\mathbb{Z} = \mathbb{Z}[1/p] = \{\, a/p^n \mid a \in \mathbb{Z}, n \in \mathbb{N}_0 \,\}$ will be the ring of rational numbers whose denominators are powers of p (see also Example 1.12 on page 18).

In a similar vein, when p is a prime and S is the complement of the principal ideal $p\mathbb{Z}$, the localization $S^{-1}\mathbb{Z}$ will be the ring $\mathbb{Z}_{(p)} = \{\, a/b \mid a, b \in \mathbb{Z}, (p, b) = 1 \,\}$ of rational numbers whose denominator is prime to p. We have already met these rings in Example 2.23 on page 55.

(7.8) Consider the polynomial ring $k[x, y, z]$ in three variables over the field k. Here we aim at describing the localization of $k[x, y, z]$ in the prime ideal (z) and showing that $k[x, y, z]_{(z)} = k(x, y)[z]_{(z)}$; that is, the polynomial ring over the fraction field $k(x, y)$ localized at the prime ideal (z). The principle from Example 7.6 takes effect, and we can work with subrings of the rational function field $k(x, y, z)$.

Let S be the subset of $k[x, y, z]$ whose members are the non-zero polynomials involving only the variables x and y. It is obviously multiplicatively closed, and just

as obviously, it holds that $S^{-1}k[x,y,z] = k(x,y)[z]$. Localizing both rings in the ideals generated by z we obtain the desired equality, noticing that $S \subseteq k[x,y]\backslash(z)$, so elements in S are already invertible in $k[x,y,z]_{(z)}$ and thus $(S^{-1}k[x,y,z])_{(z)} = k[x,y,z]_{(z)}$. ☆

Functoriality

(7.8) The ring $S^{-1}A$ depends of course on both A and S, so functoriality is naturally formulated in terms of the pair (A,S). Assume given another pair (B,T) and a map of rings $\phi\colon A \to B$ such that ϕ sends elements of S into T. Then there is induced a map of rings $\phi_{S,T}\colon S^{-1}A \to T^{-1}B$ satisfying $\phi_{S,T} \circ \iota_S = \iota_T \circ \phi$: since ϕ takes S into T, the elements $\phi(s)$ become invertible in $T^{-1}B$, and the universal property of $S^{-1}A$ guarantees that $\iota_T \circ \phi$ extends to a uniquely defined map $\phi_{S,T}\colon S^{-1}A \to T^{-1}A$. This map simply sends a/s to $\phi(a)/\phi(s)$.

$$\begin{array}{ccc} A & \xrightarrow{\phi} & B \\ {\scriptstyle \iota_S}\downarrow & & \downarrow{\scriptstyle \iota_T} \\ S^{-1}A & \xrightarrow[\phi_{S,T}]{} & T^{-1}B \end{array}$$

(7.9) A particular case to notice is when $A = B$ and $\phi = \mathrm{id}_A$. If $S \subseteq T$, there is a canonical map $S^{-1}A \to T^{-1}A$ which just interprets fractions a/s in $S^{-1}A$ as a fractions in $T^{-1}A$. This map might appear very much like doing nothing; but be aware, it can have a non-trivial kernel. When some member of T kills elements in A not killed by anyone in S, there will non-zero members of the kernel.

EXERCISE 7.2 Let S and T be two multiplicatively closed subsets of A. Let $T' = \iota_S(T)$. Prove that $\iota_S(T)$ is a multiplicatively closed subset of $S^{-1}A$ and that $\iota_S(T)^{-1}S^{-1}A$ is canonically isomorphic with $(ST)^{-1}A$, where ST is the multiplicatively closed set whose elements are products of elements from S and T. ★

The field of fraction of a domain

The field of fractions of a domain (kvotientkroppen til et område)

(7.10) Every domain is contained in a field $K(A)$ canonically attached to A, which in some sense is the smallest field containing A. It is called the *field of fractions of* A and is constructed as the localisation $\Sigma^{-1}A$ of A in the set Σ of non-zero elements from A; the set Σ is multiplicatively closed since A is a domain. Elements of $K(A)$ are all of the form ab^{-1} with a and b elements from A and $b \neq 0$, and since zero divisors are absent from A, it holds true that $a/b = a'/b'$ if and only if $ab' = a'b$. In particular, we see that $K(A)$ is a field: that $a/b \neq 0$, means that $a \neq 0$ and then b/a is defined and serves as the inverse of a/b.

The field $K(A)$ is the smallest field containing A in the sense that if $A \subseteq L$ with L a field, the universal property of a localization furnishes an injectiv map from $K(A)$ into L whose image is a copy of $K(A)$ lying between A and L.

EXAMPLE 7.9 A familiar example of fraction fields is the field $\mathbb{Q}$ of rational numbers, the fraction field of $\mathbb{Z}$. Another is the field $\mathbb{C}(x_1,\ldots,x_r)$ of rational functions in the variables $x_1,\ldots,x_r$, which is the fraction field of the polynomial ring $\mathbb{C}[x_1,\ldots,x_r]$. ☆

(7.11) Every multiplicatively closed set S not containing 0 is contained in Σ. Hence

there is a canonical map $S^{-1}A \to K(A)$, and since there are no zero divisors around, it is an embedding. This map is as canonical as can be, simply, it sending as^{-1} to as^{-1}, but there is in the outset a subtle distinction between the two localizations as^{-1} and as^{-1}; they live formally in the distinct rings $S^{-1}A$ and $K(A) = \Sigma^{-1}A$. However, in the sequel we gladly ignore these subtleties and consider the two rings to be equal: we shall (when A is a domain) tacitly identify $S^{-1}A$ with its *alter ego* in $K(A)$. Notice that the maps $\phi_{S,T}$ from Paragraph **??**, where T is another multiplicatively closed subset T containing S, then become inclusions.

Exercises

✲ (7.3) Let A be a domain with fraction field K, and let $\Sigma = A\backslash\{0\}$. Let $t_1, \dots, t_n$ be variables. Show that $\Sigma^{-1}A[t_1, \dots, t_n] = K[t_1, \dots, t_n]$.

(7.4) Show that the field of fractions of the formal power series ring $k[\![t]\!]$ is the ring $k((t))$ of formal Taylor series in t. That is, the ring whose elements are series of the form $\sum_{i \geqslant -n} a_i t^i$ with $a_i \in k$ and where addition and multiplication are the natural ones. The addition is performed termwise and the multiplcation is the usual Cauchy product: $\sum_{\mu \geqslant -m} a_\mu t^\mu \cdot \sum_{\nu \geqslant -n} b_\nu t^\nu = \sum_{i \geqslant -\max(n,m)} \sum_{\mu+\nu=i} a_\mu b_\nu t^i$.

★

Saturation and equality

A very natural question is when will two multiplicative sets S and T give rise to the same localization? As usual, care must be taken when saying that things are equal. In the present context, the precise meaning of $S^{-1}A$ and $T^{-1}A$ being the same, is that there is an isomorphism $\theta\colon S^{-1}A \to T^{-1}A$ compatible with the localization maps; *i. e.* it satisfies $\theta \circ \iota_S = \iota_T$. However, when A is a domain and all localizations are considered to be subrings of the fraction field $K(A)$, there is no abracadabra, and equal means equal.

(7.12) Recall that a multiplicatively closed set is said to be *saturated* if with s every factor of s belongs to S; that is, if $s = uv$ lies in S, so does u (and hence by symmetry v). Every multiplicative S has a *saturation*, a smallest saturated multiplicative set $\widehat{S}$ containing S. It consists of all factors of elements from S, or when described in symbols, it appears as

The saturation of a multiplicative set (metningen til en multiplikativ mengde)

$$\widehat{S} = \{\, u \in A \mid uv \in S \text{ for some } v \in A \,\}.$$

This set is multiplicatively closed: $uv \in S$ and $u'v' \in S$ implies $uu' \cdot vv' \in S$ since S is multiplicatively closed, and it is evidently saturated since a factor of a factor is a factor!

Example 7.10 Let a be a natural number and let S be the multiplicatively closed set $S = \{\, a^n \mid n \in \mathbb{N}_0 \,\}$ of powers of a. Moreover, let $p_1, \dots, p_r$ be the different prime factors

of a. Then the elements of the saturation $\widehat{S}$ are all integers of the form $\pm p_1^{\nu_1} \cdot \ldots \cdot p_r^{\nu_r}$ with the exponents ν_i being arbitrary non-negative integers. ☆

In the context of the Fundamental Existence Theorem (Theorem 2.49 on page 49) you were asked (Problem 2.23 on page 53) to show that a multiplicative set S is saturated if and only if the complement is a union of prime ideals. Connecting up with that, we have the following description of the saturation $\widehat{S}$, or rather of its complement:

PROPOSITION 7.13 *Let S be a multiplicatively closed set in A. The complement of the saturation $\widehat{S}$ is the union of the prime ideals maximal subjected to not meeting S.*

PROOF: A member x of A does not belong to $\widehat{S}$ if and only if the the principal ideal (x) does not meet S. Hence, according to the Fundamental Existence Theorem, $x \in A \backslash \widehat{S}$ if and only if there is a prime ideal maximal subjected containing (x) and to not meeting S. ❑

(7.14) The next lemma answers the retoric question at the top of this paragraph:

LEMMA 7.15 *The three following assertions hold true for multiplicatively closed subsets S and T of a ring A:*

i) The canonical map $S^{-1}A \to \widehat{S}^{-1}A$ is an isomorphism;
ii) If $S \subseteq T$ and the canonical map $S^{-1}A \to T^{-1}A$ is an isomorphism, then $T \subseteq \widehat{S}$;
iii) There is an isomorphism between $S^{-1}A$ and $T^{-1}A$ compatible with the localization maps if and only if $\widehat{S} = \widehat{T}$.

PROOF: We begin with proving *i)*: Take an element from $\widehat{S}^{-1}A$. It is shaped like au^{-1} with $u \in \widehat{S}$, so that $uv \in S$ for some $v \in A$. Hence $au^{-1} = av(uv)^{-1}$, and the map is surjective. That an element $as^{-1} \in S^{-1}A$ maps to zero, means that a is killed by some u in $\widehat{S}$, but $u \in \widehat{S}$ means that $uv \in S$ for some v. Hence $(uv)a = 0$ and $a = 0$ in $S^{-1}A$.

To establish *ii)* we first observe that the canonical map being injective means that an element $a \in A$ satisfying $ta = 0$ for some $t \in T$, also satisfies $sa = 0$ for some $s \in S$. Now, let $t \in T$. The canonical map being surjective entails that t^{-1} lies in its image, *i. e.* for an $s \in S$ it holds that $t^{-1} = as^{-1}$ in $T^{-1}A$. This means that $u(at - s) = 0$ for some $u \in T$. But by the initial observation, it ensues that $v(at - s) = 0$ for some $v \in S$, and t is a factor of the element vs which lies in S.

The last assertion *iii)* is a direct consequence of the two others. ❑

Exercises

⁂ (7.5) Show that $\mathbb{Z}[1/10] = \mathbb{Z}[1/2, 1/5]$. Generalize.

⁂ (7.6) Prove that any intermediate ring $\mathbb{Z} \subseteq A \subseteq \mathbb{Q}$ is a localization of $\mathbb{Z}$ in a multiplicative set S.

⁂ (7.7) Prove that the group of units A^* in A is a multiplicative set. Show that the

localization maps ι_S is an isomorphism if and only if S is a subset of A^*.

⁂ (7.8) Consider be the polynomial ring $A = k[x_1, \ldots, x_n]$ over the field k. Let S be the set of non-zero polynomials in A that depend only on the first r variables; *i. e.* those on the form $p(x_1, \ldots, x_r)$. Show that S is multiplicatively closed and that $S^{-1}A = K[x_{r+1}, \ldots, x_n]$ where K is the field $k(x_1, \ldots, x_r)$ of rational functions.

⁂ (7.9) Let A be any ring. Describe the saturation of the set $\{1\}$.

⁂ (7.10) Both the set of even and the set of odd (non-zero) numbers are multiplicatively closed . What are their saturations? Let p be a prime. Verify that the set of integers congruent one mod p constitute a multiplicative set. What is the saturation?

⁂ (7.11) Describe all the saturated multiplicative sets in $\mathbb{Z}$. Generalise to any factorial domain A.

⁂ (7.12) Given an ideal $\mathfrak{a}$ in A. Describe the saturation of the multiplicative set $1 + \mathfrak{a}$.

(7.13) Let M be an A-module. A ring element $x \in A$ is called a *zero divisor* on M if $xm = 0$ for some non-zero $m \in M$. Show that the set S of non-zero divisors on M form a saturated multiplicatively closed set. Hence the set of zero divisors $\mathcal{Z}(M)$ on M is the union of prime ideals. Show that $\mathcal{Z}(M)$ is the union of the ideals maximal among the prime ideals not meeting S. These are called the *maximal associated ideals* to M.

(7.14) When L is a set of primes in $\mathbb{Z}$ (finite or infinite), let $\mathbb{Z}_{(L)}$ denote the localization in all primes outside L; that is, we put $\mathbb{Z}_{(L)} = \bigcap_{p \in L} \mathbb{Z}_{(p)}$. Show that there is a natural isomorphism $\mathbb{Z}_{(L)} \otimes_{\mathbb{Z}} \mathbb{Z}_{(L')} \simeq \mathbb{Z}_{(L \cap L')}$.

★

Ideals and localization

There is a strong relationship between ideals in A and ideals in $S^{-1}A$. It relies on the two functorial ways of transporting ideals to and fro along ring maps (as explained in Section 2.9 on page 32). On the one hand any ideal $\mathfrak{b}$ in $S^{-1}A$ may be pulled back to give an ideal $\iota_S^{-1}(\mathfrak{b})$ in A (when ι_S is injective and A is considered to be contained in $S^{-1}A$, this is just the intersection $A \cap \mathfrak{b}$). On the other hand, one may extend ideals in A to ideals in $S^{-1}A$: the extension of $\mathfrak{a} \subseteq A$ to $S^{-1}A$ is the ideal $\mathfrak{a}S^{-1}A$ in $S^{-1}A$ generated by $\iota_S(\mathfrak{a})$. To simplify the notation we shall write $S^{-1}\mathfrak{a}$ for it.

(7.16) The extension map $\mathcal{I}(A) \to \mathcal{I}(S^{-1}A)$ preserves inclusions (but not strict inclusions) and as extension maps always do, it preserves products and sums of ideals (see Proposition 2.13 on page 33).

In general the extension map from $\mathcal{I}(A)$ to $\mathcal{I}(S^{-1}A)$ is not injective. For instance, it may happen that $S \cap \mathfrak{a} \neq \varnothing$, in which case the extension $S^{-1}\mathfrak{a}$ will contain an element invertible in $S^{-1}A$ and consequently be equal to whole ring $S^{-1}A$; and of course, this may be the case for several different ideals. In quite another corner, ideals $\mathfrak{a}$ contained in the kernel of ι_S reduce to the zero ideal in $S^{-1}A$. So, some ideals are blown up to $S^{-1}A$ (those meeting S) and some collapsed to zero (those contained in

$\ker \iota_S$). See Exercises 7.17 and 7.18 below for a discussion of when ideals have coinciding localizations.

EXAMPLE 7.11 A simple instance of the extension map not being injective is the case when $A = \mathbb{Z}$ and $S = \{ p^n \mid n \in \mathbb{Z} \}$ for some prime p. All the ideals $\mathfrak{a} = (p^m)$ extend to the entire ring $S^{-1}\mathbb{Z}$. This also illustrates that forming extensions does not commute with forming infinite intersections; indeed, one has $\bigcap_m (p^m) = 0$ whereas $\bigcap_m p^m S^{-1}\mathbb{Z} = S^{-1}\mathbb{Z}$. ☆

(7.17) The extension map is however surjective. Any ideal $\mathfrak{b} \subseteq S^{-1}A$ equals $S^{-1}\iota_S^{-1}\mathfrak{b}$; that is, when pulling an ideal back to A and subsequently extending the result, one recovers the original ideal. To see this, notice that if $b = a/s$ belongs to $\mathfrak{b}$, the element a belongs to $\iota_S^{-1}(\mathfrak{b})$ as $\iota_S(a) = b \cdot s$, and therefore b lies in the extension $S^{-1}\iota_S^{-1}(\mathfrak{b})$.

PROPOSITION 7.18 (IDEALS IN LOCALIZATIONS) *The extension map from the lattice $\mathcal{I}(A)$ to the lattice $\mathcal{I}(S^{-1}A)$ given by $\mathfrak{a} \to S^{-1}\mathfrak{a}$ is surjective. It preserves inclusions, products, sums and finite intersections. One has $\iota_S^{-1}(S^{-1}\mathfrak{a}) = \{ a \in A \mid sa \in \mathfrak{a} \text{ for some } s \in S \}$, and for ideals $\mathfrak{b} \subseteq S^{-1}A$ it holds true that $\mathfrak{b} = S^{-1}(\iota_S^{-1}\mathfrak{b})$.*

PROOF: We have already proved most of the proposition, only the assertions about sums, products and intersections remain unproven. It is a general feature of extension of ideals that products and sums are preserved, so we concentrate on the finite intersections; and of course, the case of two ideals will suffice.

Clearly $S^{-1}(\mathfrak{a} \cap \mathfrak{a}') \subseteq S^{-1}\mathfrak{a} \cap S^{-1}\mathfrak{a}'$. So assume that $b \in S^{-1}\mathfrak{a} \cap S^{-1}\mathfrak{a}'$. One may then express b as $b = a/s = a'/s'$ with elements a and a' from respectively $\mathfrak{a}$ and $\mathfrak{a}'$. This yields $t(a \cdot s' - a' \cdot s) = 0$ for some $t \in S$. But then $tsa' = ts'a \in \mathfrak{a} \cap \mathfrak{a}'$ and consequently $b = tsa'/tss'$ lies in $S^{-1}(\mathfrak{a} \cap \mathfrak{a}')$. ❑

(7.19) Prime ideals behave more lucidly under localization than general ideals. Either they blow up and become equal to the entire localized ring $S^{-1}A$, or they persist being prime. Moreover, every prime ideal in $S^{-1}A$ is of the shape $\mathfrak{p}S^{-1}A$ for an unambiguous prime ideal $\mathfrak{p}$ of A, so that two different prime ideals persist being different unless both blow up. One has:

PROPOSITION 7.20 (PRIME IDEALS IN LOCALIZATIONS) *Assume $\mathfrak{p}$ is a prime ideal in the ring A and S is a multiplicative subset of A. Extending $\mathfrak{p}$ to the localization $S^{-1}A$ has two possible outcomes. Either $S^{-1}\mathfrak{p} = S^{-1}A$, and this occurs if and only if $\mathfrak{p} \cap S \neq \varnothing$, or otherwise $S^{-1}\mathfrak{p}$ is a prime ideal and $\iota_S^{-1}(S^{-1}\mathfrak{p}) = \mathfrak{p}$.*

PROOF: If $S \cap \mathfrak{p} \neq \varnothing$ the ideal $\mathfrak{p}$ blows up to the entire ring in $S^{-1}A$; that is, it holds that $S^{-1}\mathfrak{p} = S^{-1}A$. If not, $S^{-1}\mathfrak{p}$ is a prime ideal; indeed, suppose that $bb' \in S^{-1}\mathfrak{p}$ and that $b = a/s$ and $b' = a'/s'$ with $a, a' \in A$ and $s, s' \in S$. We infer that $tss'bb' = taa' \in \mathfrak{p}$ for some $t \in S$, and hence either a or a' lies in $\mathfrak{p}$ since t does not. Moreover, if $\iota_S(a) = a's^{-1}$ for some $a' \in \mathfrak{p}$, if follows that $sta = ta' \in \mathfrak{p}$, hence $a \in \mathfrak{p}$ since $\mathfrak{p}$ is a prime ideal. ❑

PROPOSITION 7.21 (PRIME IDEALS IN LOCALIZATIONS II) *The prime ideals in the localization $S^{-1}A$ are precisely the ideals of the form $S^{-1}\mathfrak{p}$ for $\mathfrak{p}$ a prime ideal in A not meeting S. The prime ideal $\mathfrak{p}$ is uniquely defined.*

In other words, extension and contraction of ideals are mutually inverse maps between the sets of prime ideals in $S^{-1}A$ and of prime ideals in A not meeting S.

PROOF: By the previous proposition, the ideals $S^{-1}\mathfrak{p}$ are all prime, so let $\mathfrak{q}$ be a prime ideal in $S^{-1}A$. Then $\mathfrak{p} = \iota_S^{-1}\mathfrak{q}$ is prime, and by the last sentence in proposition 7.18 above it holds that $\mathfrak{q} = S^{-1}\mathfrak{p}$. ❑

(7.22) Localization commutes as we have seen, with several processes ideals can be exposed to, like forming products, intersections and sums. In this paragraph we treat the case of radicals, and as a further example, transporters are covered in Problem 7.23 below.

PROPOSITION 7.23 (RADICALS LOCALIZE) *Let $\mathfrak{a}$ be an ideal in A and $S \subseteq A$ a multiplicative set. Then it holds true that*

$$\sqrt{S^{-1}\mathfrak{a}} = S^{-1}\sqrt{\mathfrak{a}}.$$

PROOF: If $xs^{-1} \in S^{-1}\sqrt{\mathfrak{a}}$ with $x^\nu \in \mathfrak{a}$ and $s \in S$, it holds that $(xs^{-1})^\nu = x^\nu s^{-\nu} \in S^{-1}\mathfrak{a}$, and so xs^{-1} lies in $\sqrt{S^{-1}\mathfrak{a}}$. For the other inclusion, assume that $xs^{-1} \in \sqrt{S^{-1}\mathfrak{a}}$, which means that for some natural number ν it holds that $(xs^{-1})^\nu = at^{-1}$ with $t \in S$ and $a \in \mathfrak{a}$. Hence $(xt)^\nu = s^\nu a t^{\nu-1} \in \mathfrak{a}$, and consequently $x \in \sqrt{\mathfrak{a}}$ and $xs^{-1} \in S^{-1}\sqrt{\mathfrak{a}}$. ❑

Exercises

(7.15) Show that if $S^{-1}\mathfrak{a} = S^{-1}A$, then the same holds for all powers $\mathfrak{a}^\nu$ of $\mathfrak{a}$.

(7.16) Let p and q be different prime numbers and let S be the multiplicative set $S = \{\, p^n \mid n \in \mathbb{N}_0 \,\}$. Describe $\mathbb{Z} \cap (pq)S^{-1}\mathbb{Z}$.

(7.17) Let S be multiplicatively closed in the ring A and let $\mathfrak{a} \subseteq A$ be an ideal.

a) Show that the ideals $(\mathfrak{a} : s)$ when s runs through S form a directed family of ideals; hence their union is an ideal;

b) Show that $\bigcup_{s\in S}(\mathfrak{a} : s) = \iota_S^{-1}(S^{-1}\mathfrak{a})$;

c) Show that $\bigcup_{s\in S}(\mathfrak{a} : s)$ is maximal among the idelas $\mathfrak{b}$ such that $S_{-1}\mathfrak{b} = S^{-1}\mathfrak{a}$.

(7.18) Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals and S a multiplicative set in the ring A. Show that $S^{-1}\mathfrak{a} = S^{-1}\mathfrak{b}$ if and only if for each pair of elements $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ there are elements s and t in S such that $sa = tb$.

(7.19) Suppose that $\mathfrak{p}$ is a prime ideal and that $\mathfrak{a}$ an ideal contained in $\mathfrak{p}$. Show that $\iota_S^{-1}(S^{-1}\mathfrak{a}) \subseteq \mathfrak{p}$.

(7.20) Let $j\colon \operatorname{Spec} S^{-1}A \to \operatorname{Spec} A$ be the map induced from the localization map $\iota_S\colon A \to S^{-1}A$.

a) Show by an example that j is not necessarily an open embedding. HINT: Let *e.g.* $A = \mathbb{Z}$ and S the multiplicative subset generated by every second prime.

b) Show that j is a homeomorphism onto its image (when the image is endowed with induced topology). Show that the image $j(\operatorname{Spec} A)$ equals the intersection of all the open sets containing it.

(7.22) Given an example of a ring A and a non-zero prime ideal $\mathfrak{p}$ such that $A_{\mathfrak{p}} = A/\mathfrak{p}$.
HINT: Let A be the product of two fields.

(7.23) Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in A, Prove that $S^{-1}(\mathfrak{a} : \mathfrak{b}) = (S^{-1}\mathfrak{a} : S^{-1}\mathfrak{b})$.

★

The local ring at a prime ideal.

A few ways of forming rings of fraction are omnipresent in algebra and algebraic geometry and they are used over and over again. The most prominent one is the localization $A_{\mathfrak{p}}$ of A at a prime ideal $\mathfrak{p}$.

The complement $S = A\backslash\mathfrak{p}$ of a prime ideal $\mathfrak{p}$ is as we have seen multiplicatively closed, and the corresponding localized ring is written as $A_{\mathfrak{p}}$. The elements are fractions a/b with $b \notin \mathfrak{p}$. The ring $A_{\mathfrak{p}}$ will be a local ring whose only maximal ideal is the expansion of $\mathfrak{p}$; that is, the ideal $\mathfrak{p}A_{\mathfrak{p}}$:

PROPOSITION 7.24 *The localisation $A_{\mathfrak{p}}$ is a local ring with maximal ideal $\mathfrak{p}A_{\mathfrak{p}}$. The assignment $\mathfrak{q} \mapsto \mathfrak{q}A_{\mathfrak{p}}$ is a one-to-one correspondence between prime ideals in $A_{\mathfrak{p}}$ and prime ideals $\mathfrak{q}$ in A contained in $\mathfrak{p}$. The inverse corresprondence is the pull-back $\mathfrak{p} \mapsto \iota_S^{-1}(\mathfrak{p})$.*

Notice, that when all zero-divisors of A lie in $\mathfrak{p}$, so that the localization map is injective, and we identify A with its image in $A_{\mathfrak{p}}$, the inverse correspondence will just be $\mathfrak{p} \mapsto \mathfrak{p} \cap A$.

PROOF: This is nothing but Proposition 7.21 on the previous page, according to which the prime ideals in $A_{\mathfrak{p}}$ are precisely the ideals in $A_{\mathfrak{p}}$ of the form $\mathfrak{q}A_{\mathfrak{p}}$ where $\mathfrak{q}$ is a prime ideal in A not meeting S, that is, contained in $\mathfrak{p}$. Moreover 7.21 also tells us that $\mathfrak{q} = \iota^{-1}(\mathfrak{q}A_{\mathfrak{p}})$. ❑

Examples

(7.12) If p is a prime number, the localized ring $\mathbb{Z}_{(p)}$ at the maximal ideal generated by p consists of the rational numbers which when written in lowest terms, have a denominator relatively prime to p. The maximal ideal in $\mathbb{Z}_{(p)}$ is generated by p, and the residue field is the field $\mathbb{F}_p$ with p elements.

(7.13) If $p(x)$ is an irreducible polynomial in the polynomial ring $k[x]$ over a field k, the ring $k[x]_{(p(x))}$ is the subring of $k(x)$ consisting of the rational functions whose denominator when written in lowest terms does not have $p(x)$ as factor. The maximal ideal is generated by $p(x)$, and the residue field will be the field obtained by adjoining a root of $p(x)$ to k. In particular, if $p(x)$ is linear, say $p(x) = x - a$, the elements of

$k[x]_{(p(x))}$ are the rational functions whose denominator does not vanish at a. The residue field will be k itself.

(7.14) We continue working with a prime ideal $\mathfrak{p}$ in a ring A. The quotient $A/\mathfrak{p}$ is naturally contained in the residue field $k(\mathfrak{p}) = A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$, as the inverse image of $\mathfrak{p}A_\mathfrak{p}$ equals $\mathfrak{p}$, and since elements in the latter are all classes of the form $[as^{-1}]$ with $a \in A$ and $s \in A\backslash\mathfrak{p}$, it ensues that the residue field $k(\mathfrak{p})$ equals the fraction field of $A/\mathfrak{p}$.

☆

(7.25) Although the prime ideal $\mathfrak{p}A_\mathfrak{p}$ pulls back to the prime ideal $\mathfrak{p}$, powers of $\mathfrak{p}A_\mathfrak{p}$ do not always pull back to powers of $\mathfrak{p}$. However, there is always a ring map

$$A/\mathfrak{p}^r \to A_\mathfrak{p}/\mathfrak{p}^r A_\mathfrak{p} \tag{7.3}$$

for the simple reason that $\mathfrak{p}^r$ maps into $\mathfrak{p}^r A_\mathfrak{p}$, but it might very well fail both to be injective and surjective. That surjectivity may fail is not unexpected, since, for instance, $A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$ will be a field whereas $A/\mathfrak{p}$ is not unless $\mathfrak{p}$ is maximal. That injectivity may fail, is certainly more subtle. It leads to the introduction of the so called *symbolic powers* $\mathfrak{p}^{(r)} = \mathfrak{p}^r A_\mathfrak{p} \cap A$ of $\mathfrak{p}$ (they will be treated more thoroughly in Exercise 10.11 on page 275). The kernel of the map in (7.3) equals the quotient $\mathfrak{p}^{(r)}/\mathfrak{p}^r$, so the map is not injective precisely when the two differ. Below (Example 7.15) we shall give an example of a symbolic square $\mathfrak{p}^{(2)}$ being different from the plain square $\mathfrak{p}^2$. When $\mathfrak{p}$ is a maximal ideal, however, the map in (7.3) will always be an isomorphism.

Symbolic powers (symbolske potenser)

LEMMA 7.26 *Let $\mathfrak{m}$ be a maximal ideal in A. Then $\mathfrak{m}^r A_\mathfrak{m} \cap A = \mathfrak{m}^r$. Moreover, the canonical map $A/\mathfrak{m}^r \to A_\mathfrak{m}/\mathfrak{m}^r A_\mathfrak{m}$ is an isomorphism.*

PROOF: This hinges on a classic from algebra, namely the formula

$$(1-x)(1+x+\ldots+x^{r-1}) + x^r = 1,$$

valid for $r \geqslant 1$ and in any ring. It implies that elements in $A/\mathfrak{m}^r$ not lying in the maximal ideal are invertible. Indeed, if $s \in A\backslash\mathfrak{m}$, we may find an element $t \in A\backslash\mathfrak{m}$ such that $x = 1 - st \in \mathfrak{m}$ (simply because $A/\mathfrak{m}$ is a field: lift the inverse of $[s]$ in $A/\mathfrak{m}$ to A). With that x the formula above gives

$$st \cdot \sum_{0 \leqslant i \leqslant r-1} (1-st)^i - 1 \in \mathfrak{m}^r.$$

Consequently the class of st in $A/\mathfrak{m}^r$, and hence *a fortiori* the class of s, is invertible.

Now, take any element $a \in \mathfrak{m}^r A_\mathfrak{m} \cap A$; it may be written as $a = xs^{-1}$ with $x \in \mathfrak{m}^r$ and $s \notin \mathfrak{m}$. Hence $sa = x \in \mathfrak{m}^r$, but since s is invertible mod $\mathfrak{m}^r$, it ensues that $a \in \mathfrak{m}^r$. Indeed, $a = (1-ts)a + tx \in \mathfrak{m}^r$ where t is an element in A such that $1 - ts \in \mathfrak{m}^r$.

In the same vein, the map in the lemma is surjective because with t as above, we find $as^{-1} - at \in \mathfrak{m}^r A_{\mathfrak{m}}$, and the element at of A maps to $as^{-1} \bmod \mathfrak{m}^r$. ❑

EXAMPLE 7.15 *A symbolic square that differs from the plain square:* Lemma 7.26 does not always hold for prime ideals that are not maximal; the ideal of a line in the cone over a plane quadric is among the simplest examples: Let k be a field and let $A = k[x, y, z]$ with constituting relation $z^2 = xy$. The ideal $\mathfrak{p} = (z, x)$ is a prime ideal, since putting x and z to zero induces a map $A \to k[y]$ whose kernel is $\mathfrak{p}$. Neither $x \notin \mathfrak{p}$ nor $y \notin \mathfrak{p}$, since in the polynomial ring $k[X, Y, Z]$ no non-zero linear form lies in $(XY, XZ, X^2, Z^2 - XY)$, simply for degree reasons, so in particular the map just defined is surjective.

In the local ring $A_{\mathfrak{p}}$ the element y is invertible and we therefore have

$$\mathfrak{p}^2 A_{\mathfrak{p}} = (z^2, zx, x^2)A_{\mathfrak{p}} = (xy, zx, x^2)A_{\mathfrak{p}} = (x)A_{\mathfrak{p}},$$

whereas in A we have $\mathfrak{p}^2 = (xy, zx, x^2)$. We already observed that $x \notin (xy, zx, x^2)$ so $\mathfrak{p}^2 \subsetneq \mathfrak{p}^2 A_{\mathfrak{p}} \cap A$. ☆

Exercises

(7.24) Let $\mathfrak{m}$ be a maximal ideal in the ring A and let r be a natural number. Show that the localization map $A \to A_{\mathfrak{m}}$ induces an isomorphism between $\mathfrak{m}^r/\mathfrak{m}^{r+1}$ and $\mathfrak{m}^r A_{\mathfrak{m}}/\mathfrak{m}^{r+1} A_{\mathfrak{m}}$ as vector spaces over $A/\mathfrak{m}$.

(7.25) Let A be a domain with fraction field K. Show that $A = \bigcap_{\mathfrak{m}} A_{\mathfrak{m}}$. HINT: For any $x \in K$ not in A prove that the ideal $\{ y \in A \mid yx \in A \}$ can not be a proper ideal.

★

Inverting powers of a single element.

Given an element $f \in A$. The set $S = \{ f^n \mid n \in \mathbb{N}_0 \}$ of all powers of f is obviously multiplicatively closed, and the corresponding ring of fractions $S^{-1}A$ is denoted A_f. The prime ideals in A_f are exactly those on the form $\mathfrak{p}A_f$ for $\mathfrak{p}$ a prime ideal in A with $f \notin \mathfrak{p}$; that is, for the members of the distinguished open subset $D(f)$ of Spec A.

There is a natural isomorphism between $A[x]/(xf - 1)$ and A_f that sends x to f^{-1}. By the universal mapping property of the polynomial ring the map is well defined, and f being invertibel in $A[x]/(xf - 1)$, the unversal property of A_f furnishes an inverse. This makes the notation $A[f^{-1}]$ for A_f legitimate; the usage is however poisonous when f is a zero-divisor. Adding f^{-1} kills, and in case f is nilpotent, the intoxication is lethal; everything is killed and $A[f^{-1}] = 0$.

EXAMPLE 7.16 It is worthwhile mentioning a concrete example. Consider the ring $A = \mathbb{C}[z]$ of complex polynomials in the variable z and let $f = z - a$. The localized ring A_f consists of those rational functions that are regular away from a; that is, they have at most a pole at a. This generalizes to the ring $\mathcal{O}(\Omega)$ of functions holomorphic in any domain Ω of the complex plane containing a. The localized ring $\mathcal{O}(\Omega)_{z-a}$ has the functions meromorphic in Ω with at most a pole at a as elements. ☆

The total ring of fractions

The set S of non-zero divisors in ring A is closed under multiplication; indeed, if s and t are non-zero divisors and $st \cdot a = 0$ with $a \neq 0$, it would follow that $ta \neq 0$ which would contradict that s is a non-zero divisor. The set S is even a saturated multiplicative set since a zero-divisor can not be factor of a non-zero-divisor.

The total ring of fractions of a ring (kvotientringen til en ring)

The localization of A in S is denoted by $K(A)$ and is called the *total ring of fractions* of A. When A is an integral domain, $S = A\backslash\{0\}$, and all non-zero elements become invertible in $K(A)$. Consequently $K(A)$ is a field; it is called the *field of fractions* of A, a construct we already met in Paragraph 7.10 on page 178.

The field of fractions (kvotientkroppen)

The ring $K(A)$ is in general not a field, but by definition has the property that all non-zero divisors are invertible. In any case, the canonical map $A \to K(A)$ is injective since by their very nature non-zero divisors do not kill non-zero elements.

PROPOSITION 7.27 *The total ring of fractions $K(A)$ of a ring A has the property that every non-zero divisor is invertible. The natural map A to $K(A)$ is injective. Moreover, $K(A)$ is a field if and only if A is an integral domain.*

(7.28) The total rings of fractions of a certain class of reduced rings—recall that A being reduced means it is without non-zero nilpotent elements—has a closer description. The class of rings we have in mind are the reduced rings with a finite number of minimal primes. Since the radical $\sqrt{(0)}$ equals the intersection of the minimal prime ideals, in these rings the zero ideal is the intersection of finitely many prime ideals; that is, one has

$$(0) = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_r, \tag{7.4}$$

where the $\mathfrak{p}_i$ are distinct prime ideals. This is a large class of rings encompassing all reduced Noetherian rings (a class of rings soon to be introduced). The $\mathfrak{p}_i$'s occurring in (7.4) being minimal the intersection is irredundant; *i. e.* the intersection of all but one of the $\mathfrak{p}_i$'s is never zero. An important observation is that the set of zero-divisors in A is precisely the union $\mathfrak{p}_1 \cup \ldots \cup \mathfrak{p}_r$ of the $\mathfrak{p}_i$'s (see Exercise 7.30 below).

PROPOSITION 7.29 (TOTAL RING OF FRACTIONS OF REDUCED RINGS) *Let A be a ring without non-zero nilpotent elements.*

i) The local ring $A_\mathfrak{p}$ at a minimal prime $\mathfrak{p}$ is a field;

ii) When A has only finitely many minimal prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$, the total ring of fraction is a product of fields; i. e. $K(A) = A_{\mathfrak{p}_1} \times \ldots \times A_{\mathfrak{p}_r}$.

Let us first remark that the statements are not generally true for non-reduces rings see Example 7.17 below of what may happen if A is not reduced.

The closed subsets Spec $A/\mathfrak{p}i$ are the irreducible components of Spec A; that is, Spec $A =$ Spec $A/\mathfrak{p}_1 \cup \ldots \cup$ Spec $A/\mathfrak{p}_r$, and the proposition says that the total ring of

fractions of Spec A equals the product of the fraction fields of the irreducible components of Spec A.

PROOF: We begin with proving *i)* which is the easier: If $\mathfrak{p}$ is a minimal prime, the local ring $A_\mathfrak{p}$ has $\mathfrak{p}$ as its sole prime ideal simply because prime ideals in $A_\mathfrak{p}$ correspond to prime ideals A contained in $\mathfrak{p}$. Radicals localize well (7.23 on page 183), and A being reduced, we infer that $A_\mathfrak{p}$ is reduced as well. Because the radical of $A_\mathfrak{p}$ equals $\mathfrak{p}A_\mathfrak{p}$, it follows that $\mathfrak{p}A_\mathfrak{p} = 0$, and consequently, $A_\mathfrak{p}$ is a field.

Statement *ii)* is a little more elaborate. For each index i the localization map $A \to A_{\mathfrak{p}_i}$ extends to a map $K(A) \to A_{\mathfrak{p}_i}$ because no non-zero divisor lies in $\mathfrak{p}_i$, and recollecting these maps we obtain a map

$$\theta\colon K(A) \to A_{\mathfrak{p}_1} \times \ldots \times A_{\mathfrak{p}_r}.$$

It sends as^{-1} to the string $(as^{-1}, \ldots, as^{-1})$. An element a from A maps to zero in $A_{\mathfrak{p}_i}$ precisely when a is killed by an element not in $\mathfrak{p}_i$, and thence $a \in \mathfrak{p}_i$. If this occurs for all indices i, the element a lies in the intersection of the $\mathfrak{p}_i$'s and is therefore equal to zero (the intersection of the $\mathfrak{p}_i$'s vanishes as A is reduced). This proves that θ is injective.

To see that θ is surjective requires some further effort. We begin with choosing an element s_i for each i such that $s_i \notin \mathfrak{p}_i$, but $s_i \in \mathfrak{p}_j$ when $j \neq i$. Then s_i becomes invertible in $A_{\mathfrak{p}_i}$, but maps to zero in $A_{\mathfrak{p}_j}$ when $j \neq i$; indeed, each s_i is killed by any non-zero element in $\bigcap_{j\neq i} \mathfrak{p}_j$, and there are such since the intersection in (7.4) is irredundant.

Now, we come to the salient point of the proof: For any choice of elements $c_1, \ldots, c_r$ with each c_j not in $\mathfrak{p}_j$, the combination $x = \sum_j c_j s_j$ is a non-zero divisor. Indeed, if x belonged to $\mathfrak{p}_i$, it would ensue from $s_j \in \mathfrak{p}_i$ when $j \neq i$ that $c_i s_i = x - \sum_{j\neq i} c_j s_j$ belonged to $\mathfrak{p}_i$, which is absurd since neither c_i nor s_i lies there. Consequently $\sum_j c_j s_j$ is invertible in $K(A)$, and in view of s_i mapping to zero in $A_{\mathfrak{p}_j}$ when $j \neq i$, one finds

$$\theta\big((\sum_j c_j s_j)^{-1}\big) = (c_1^{-1}s_1^{-1}, \ldots, c_r^{-1}s_r^{-1}).$$

Finally, if $a_1, \ldots, a_r$ are are arbitrary elements A one arrive at

$$\theta\big((\sum_j s_i a_i)(\sum_j c_j s_j)^{-1}\big) = (a_1 c_1^{-1}, \ldots, a_r c_r^{-1}),$$

showing that θ is surjective. ❏

To describe the total quotient rings of of rings that are not reduced is in general much more involved, as indicated by the following example. For a certain class of rings—those "without embedded components"—a description similar to the one in the proposition holds. An indication is given in Exercise 7.26 (7.31) below.

EXAMPLE 7.17 Let $B = k[X, Y]/(X^2, XY)$, and as usual, let the lower case versions x and y denote the classes of X and Y in B. We contend that $\mathfrak{m} = (x, y)$ consists of all the

zero-divisors in B, and the total quotient ring $K(B)$ is therefore given as the local ring $K(B) = B_{\mathfrak{m}}$.

Clearly both x and y are zero divisors, so all members in $\mathfrak{m}$ will be too. Assume then that $ab = 0$ with neither a nor b being zero. Consider the classes $[a]$ and $[b]$ in $B/\mathfrak{m} = k$. Their product is zero, hence at least one of them vanishes, say $[a]$, which means that $a \in \mathfrak{m}$. Assuming that $b \notin \mathfrak{m}$, after possibly rescaling b, one may write $a = cx + dy$ and $b = 1 + ex + fy$, where d and f do not belong to the ideal $(x)B$. Then, using the relations $x^2 = xy = 0$ which hold in B, one finds

$$0 = ab = (cx + dy) \cdot (1 + ex + fy) = cx + dy + fdy^2 = cx + yd(1 + yf),$$

and hence $yd \in (x)B$. But because $B/(x)B = k[X, Y]/(X^2, XY, X) = k[Y]$, the ideal $(x)B$ is a prime ideal, and one infers that either $y \in (x)$ or $d \in (x)$, which is a contradiction.

The ring $B_{\mathfrak{m}}$ has two prime ideals, the maximal ideal (x, y) whose elements constitute all zero-divisors, and a sole minimal prime ideal (x) whose elements are all the nilpotents of B. ☆

EXERCISE 7.26 With reference to the example, show that $B_{(x)}$ equals the rational function field $k(Y)$. ★

Exercises

(7.27) Let n be a natural number. Determine the total quotient ring of $\mathbb{Z}/n\mathbb{Z}$.

(7.28) Let A be any ring. Show that the nil-radical of $K(A)$ is equal to the extension of the nil-radical of A.

(7.29) Let A be a ring.

a) Show that if the elements of A are either zero divisors or invertible, then $A = K(A)$.

b) If A has only one prime ideal, prove that $K(A) = A$.

c) Let A be a direct product (indexed by a set of any cardinality) of rings each having only one prime ideals. Prove that $K(A) = A$.

(7.30) Assume that A is a reduced ring so that $(0) = \bigcap_{i \in I} \mathfrak{p}_i$ where the intersection extends over the minimal prime ideals of a ring A. Show that the union $\bigcup_{i \in I} \mathfrak{p}_i$ equals the set of zero divisors in A. HINT: Observe that $\mathfrak{p}_i$ kills $\bigcap_{j \neq i} \mathfrak{p}_j$.

(7.31) Let t A be a ring. Assume that $\sqrt{(0)} = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_r$ is an irredundant intersection of prime ideals. Assume further that the set of non-zero divisors of A equals the union $\mathfrak{p}_1 \cup \ldots \cup \mathfrak{p}_r$. Show that the total ring of fractions $K(A)$ decomposes as the direct product $K(A) = A_{\mathfrak{p}_1} \times \ldots \times A_{\mathfrak{p}_r}$. HINT: Be inspired by the proof of the second assertion in Proposition 7.29.

(7.32) Let k be a field and consider the polynomial ring $A = k[x_1, \ldots, x_n]$. Let $r < n$ be a natural number. Let S be the subset of A of polynomials in the variable $x_{r+1}, \ldots, x_n$.

Show that S is multiplicatively closed and that $S^{-1}k[x_1,\dots,x_n] = K[x_1,\dots,x_r]$ where $K = k(x_{r+1},\dots,x_n)$ is the field of rational functions in the variables $x_{r+1},\dots,x_n$.

(7.33) Let A be a domain with quotient field K. Denote by S the multiplicative set $A\backslash 0$ of non-zero elements in A. Show that $S^{-1}A[T] = K[T]$.

★

7.2 *Localization of modules*

There is also a procedure to localize an A-module M in a multiplicatively closed set S closely resembling the way the fraction ring $S^{-1}A$ was constructed, and the localized module will be denoted by $S^{-1}M$. The construction of $S^{-1}M$ is functorial in M and gives a functor $\mathsf{Mod}_A \to \mathsf{Mod}_{S^{-1}A}$ with the important properties of being additive and exact. Moreover, it preserves tensor products and hom-sets between finitely presented modules. It takes submodules to submodules and respects most of the standard operations on submodules. The localisation functor turns out to coincide with the base change functor $M \mapsto M \otimes_A S^{-1}A$.

Just as with rings, one writes $M_{\mathfrak{p}}$ and M_f for $S^{-1}M$ when S is respectively the complement of a prime ideal $\mathfrak{p}$ *i. e.* $S = A\backslash\mathfrak{p}$ and the set $S = \{f^n\}$ of non-negtative powers of an element.

(7.30) To construction the localized module $S^{-1}M$ we mimick the way $S^{-1}A$ was fabricated. Details will be skipped, but they may be verified *mutatis mutandis* as in the case of ring.

To begin with one introduces an equivalence relation on the Cartesian product $M \times S$ by declaring two pairs (m,s) and (m',s') to be equivalent if

$$t(ms' - m's) = 0 \tag{7.5}$$

for some $t \in S$. One checks that this is an equivalence relation (transitivity is the only challenge) and defines $S^{-1}M$ to be the set of equivalence classes $S^{-1}M = M \times S/\sim$. The equivalence class of a pair (m,s) will be designated either by m/s or by ms^{-1}. The additive group structure of $S^{-1}M$ is introduced in analogy with the usual way of adding fraction, namely as $m/s + n/t = (mt + sn)/st$. And the action of an element $a/s \in S^{-1}A$ is given in the straightforward way: $a/s \cdot m/t = am/st$. There is canonical map $\iota_S \colon M \to S^{-1}M$ that sends m to the class of $(m,1)$.

Naturally, there is a lot of checking to be done; that definitions are legitimate and that axioms are satisfied. Every single step is straightforward, and we leave these soporific verifications to the students for a misty day. Summing up, one has:

PROPOSITION 7.31 *The localization $S^{-1}M$ is an $S^{-1}A$-module, and the canonical localization map $\iota_S \colon M \to S^{-1}M$ is A linear. Every element of $S^{-1}M$ is of the form m/s and two such, m/s*

and m'/s', are equal precisely when $ts'm = tsm'$ for some $t \in S$. The kernel of ι_S consists of the elements m in M killed by some member of S; that is, $\ker \iota_S = \{\, m \in M \mid tm = 0 \text{ for some } t \in S \,\}$.

To simplify the notion, just as with rings, one soon drops the reference to the map ι_S and writes x or $x/1$ for $\iota_S(x)$, but with some cautiousness since the image very well can be zero.

(7.32) When the module M is finitely generated, say by members $m_1, \dots, m_r$, the images $\iota_S(m_i)$ of the m_i's will obviously generate $S^{-1}M$. Indeed, pick a member xs^{-1} from $S^{-1}M$ and write $x = \sum a_i m_i$ then of course $xs^{-1} = \sum a_i s^{-1} m_i$.

Functoriality

(7.33) Given two A-modules M and N and an A-linear map $\phi\colon M \to N$. Sending xs^{-1} to $\phi(x)s^{-1}$ gives an $S^{-1}A$-linear map between the localized modules $S^{-1}M$ and $S^{-1}N$; that is, a map $S^{-1}\phi\colon S^{-1}M \to S^{-1}N$.

A formal definition starts with the map $(x, s) \to (\phi(x), s)$ between the Cartesian products $M \times S$ and $N \times S$, and the salient point is that this respects the equivalence relations from (7.5). Indeed, a relation like $t(xs' - x's) = 0$ leads to the relation $t(\phi(x)s' - \phi(x')s) = 0$ because ϕ is A-linear. Thus $\phi(x)s^{-1}$ does not depend on the choice of representatives, and $xs^{-1} \mapsto \phi(x)s^{-1}$ is a legitimate definition.

(7.34) From the definition of $S^{-1}\phi$ we infer immediately that a linear combination of maps between N and M localizes to the corresponding linear combination; that is, one has

$$S^{-1}(a\phi + b\psi) = aS^{-1}\phi + bS^{-1}\psi,$$

where ϕ and ψ are A-linear maps from M to N and a and b ring elements. And it is equally clear that the association is functorial; it holds true that

$$S^{-1}(\psi \circ \phi) = S^{-1}\psi \circ S^{-1}\phi$$

whenever ϕ and ψ are composable, since it holds true already at the level of the Cartesian products—and of course, $S^{-1}(\mathrm{id}_M) = \mathrm{id}_{S^{-1}M}$.

PROPOSITION 7.35 *Let A be a ring and S a multiplicative subset of A. The localization functor $\mathsf{Mod}_A \to \mathsf{Mod}_{S^{-1}A}$ is additive and exact.*

PROOF: The only subtle point is that the functor is exact. In other words that it brings an exact sequence

$$N \xrightarrow{\psi} M \xrightarrow{\phi} L \tag{7.6}$$

to an exact sequence. So our sole task is to verify that the sequence

$$S^{-1}N \xrightarrow{S^{-1}\psi} S^{-1}M \xrightarrow{S^{-1}\phi} S^{-1}L$$

is exact, which amounts to checking that $\ker S^{-1}\phi = \operatorname{im} S^{-1}\psi$. To that end, pick an element xs^{-1} in the kernel of $S^{-1}\phi$. This means that $\phi(x)s^{-1} = 0$, hence $t\phi(x) = 0$ for some $t \in S$. But then $tx \in \ker\phi$, and since the sequence (7.6) is exact, there is an element y in N such that $\psi(y) = tx$. But then we have $S^{-1}\psi(ys^{-1}t^{-1}) = \psi(y)s^{-1}t^{-1} = xs^{-1}$, and we are through. ❑

Submodules

Given a submodule $N \subseteq M$. The localized module $S^{-1}N$ can be considered to be a submodule of $S^{-1}M$. The inclusion map localizes to an injection whose image consists of elements shaped like fractions ns^{-1} with $n \in N$ and $s \in S$, and thus it can naturally be identified with $S^{-1}N$. Notice that since localization is an exact operation, there is a canonical isomorphism $S^{-1}(M/N) \simeq S^{-1}M/S^{-1}N$ sending a class $[m]s^{-1}$ to the class $[ms^{-1}]$, and certainly, we shall not refrain from the slight abuse of language it is to consider the two to be equal.

(7.36) Localization behaves nicely with respect to many of the standard operations one may perform on submodules like taking sums and finite intersections and forming annihilators and transporters. However, localization does not commute with infinite intersections as we saw in example 7.11 on page 182, nor does it commute with forming annihilators of infinitely generated modules (Exercises 7.34 and 7.35 below), but the formation of arbitrary direct sums commute with localizations. We summarize some of these properties in the following proposition:

PROPOSITION 7.37 *Let A be a ring and S a multiplicative set in A. Let N, N' and $\{N_i\}$ be submodules of the A-module M. Then the following four properties hold true:*

i) $S^{-1}\sum_i N_i = \sum_i S^{-1}N_i$;
ii) $S^{-1}(N \cap N') = S^{-1}N \cap S^{-1}N'$;
iii) $S^{-1}\bigoplus_{i\in I} M_i \simeq \bigoplus_{i\in I} S^{-1}M_i$;
iv) Assume that N is finitely generated, then $(N' : N)S^{-1}A = (S^{-1}N' : S^{-1}N)$.

PROOF: To establish the first equality, observe that from the inclusion $N_i \subseteq \sum_i N_i$ ensues that $S^{-1}N_i \subseteq S^{-1}\sum_i N_i$, hence one has $\sum_i S^{-1}N_i \subseteq S^{-1}\sum_i N_i$. Any element in $S^{-1}\sum_i N_i$ is of the form $(\sum_i x_i)s^{-1}$ with merely finitely many of the x_i's being non-zero, and therefore lies in $\sum_i S^{-1}N_i$.

It holds that $S^{-1}(N \cap N') \subseteq S^{-1}N \cap S^{-1}N'$ as localization respects inclusions, and the second assertion follows because if $y \in S^{-1}N \cap S^{-1}N'$ we have $y = n/s = n'/s'$ with $n \in N$ and $n' \in N'$ and $s, s' \in S$, which means that $ts'n = tsn'$ for some $t \in S$. Putting $x = ts'n = tsn'$, we infer that $x \in N \cap N'$ and $y = x/tss' \in S^{-1}(N \cap N')$.

The third assertion is a direct consequence of the two first for a direct sum of two modules; hence it holds for a finite sum by an obvious induction argument. Finally, the case of an infinite sum follows from the finite case since each element in $\bigoplus_{i\in I} M_i$ and

each in $\bigoplus_i S^{-1}M_i$ lies in a submodule which is a direct sum of a finite number of the M_i's.

In the fourth and last assertion the module N is assumed to be finitely generated. Let $m_1, \dots, m_r$ be generators. These also generate the localized module $S^{-1}N$ over $S^{-1}A$. The trick is to consider the A-linear mapping

$$\mu\colon A \xrightarrow{\mu} \bigoplus_i M/N'$$

that sends a ring element a to the sequence $([am_1], \dots, [am_r])$ where classes are taken mod N'. The transporter $(N' : N) = \{\, a \mid aN \subseteq N' \,\}$ satisfies $(N' : N) = \bigcap_i (N' : Am_i)$ and appears as the kernel of this map and thus lives in the exact sequence

$$0 \longrightarrow (N' : N) \longrightarrow A \xrightarrow{\mu} \bigoplus_{1 \leqslant i \leqslant r} M/N'.$$

Because localization is an exact operation which commutes with direct sums, when localized in S, this sequence becomes

$$0 \longrightarrow S^{-1}(N' : N) \longrightarrow S^{-1}A \xrightarrow{S^{-1}\mu} \bigoplus_{1 \leqslant i \leqslant r} S^{-1}M/S^{-1}N' \,,$$

where $S^{-1}\mu(a) = ([am_1], \dots, [am_r])$ with $a \in S^{-1}A$ and the classes being taken mod $S^{-1}N'$. Since the m_i's generate $S^{-1}N$, it holds that $\ker \mu = (S^{-1}N' : S^{-1}N)$, and the equality $S^{-1}(N' : N) = (S^{-1}N' : S^{-1}N)$ follows. ❑

It is worthwhile mentioning two particular cases of the fourth assertion, namely when $N = M$ and $N' = (0)$, in which case $(N' : N) = \operatorname{Ann} M$, and the case when $N' = 0$ and N is generated by a single element m. Then $(N' : N) = (0 : Am) = \operatorname{Ann} m$. In short, for finitely generated modules forming annihilators commute with localization.

COROLLARY 7.38 *Assume that A is a ring with a multiplicative set S and that M is an A-module. Then*

i) For any element $m \in M$ it holds true that $S^{-1}(0 : Am) = (0 : S^{-1}Am)$;
ii) If M is finitely generated, one has $S^{-1} \operatorname{Ann} M = \operatorname{Ann} S^{-1}M$.

Exercises

(7.34) Localization does not commute with infinite direct products in general. Let $p \in \mathbb{Z}$ be a number and denote by S the multiplicative set $S = \{\, p^n \mid n \in \mathbb{N}_0 \,\}$ in $\mathbb{Z}$. Show that there is a natural inclusion

$$S^{-1} \prod_{i \in \mathbb{N}} \mathbb{Z} \subseteq \prod_{i \in \mathbb{N}} S^{-1}\mathbb{Z},$$

but that the inclusion is strict. HINT: Strings shaped like $(a_i p^{-n_i})_{i \in \mathbb{N}}$ will not lie in the image when n_i tends to infinity with i.

(7.35) Let p be a prime and let S be the multiplicative set $\{\, p^n \mid n \in \mathbb{N}_0 \,\}$ in $\mathbb{Z}$ of all non-negative powers of p. Consider the abelian group $\bigoplus_i \mathbb{Z}/p^i\mathbb{Z}$. Show that

$S^{-1}(\bigoplus_i \mathbb{Z}/p^i\mathbb{Z}) = 0$. Show that $\operatorname{Ann}(\bigoplus_i \mathbb{Z}/p^i\mathbb{Z}) = (0)$. But of course it holds true that $\operatorname{Ann} S^{-1}(\bigoplus_i \mathbb{Z}/p^i\mathbb{Z}) = S^{-1}\mathbb{Z} = \mathbb{Z}[p^{-1}]$, hence localization and forming annihilators do not always commute.

★

Relation with the tensor product

(7.39) Given an A-module M, the "action" of $S^{-1}A$ on M is expressed by the map $M \times S^{-1}A \to S^{-1}M$ that sends (m, as^{-1}) to $am \cdot s^{-1}$. This is obviously A-bilinear, and in view of the universal property enjoyed by the tensor product, induces an A-linear map $\Psi\colon M \otimes_A S^{-1}A \to S^{-1}M$, which on decomposable tensors acts by sending $m \otimes as^{-1}$ to ams^{-1}. This map turns out to be an isomorphism:

PROPOSITION 7.40 *The map Ψ is an $M \otimes_A S^{-1}A \simeq S^{-1}M$ of A-modules.*

PROOF: The crux of the poof is that the tensor product $M \otimes_A S^{-1}A$ is one of the rare instances that all elements are decomposable; that is, they are all of the form $m \otimes s^{-1}$ with $m \in M$ and $s \in S$. Granted this, if $m \otimes s^{-1}$ is mapped to zero, the element m is annihilated by some t from S. But then $m \otimes s^{-1} = tm \otimes s^{-1}t^{-1} = 0$. So the map Ψ is injective, and it obviously is also surjective.

A priori an element from $M \otimes S^{-1}A$ is of the shape $\sum_{1 \leqslant i \leqslant r} m_i \otimes a_i s_i^{-1}$ with $a_i \in A$ and $s_i \in S$. Moving the a_i through the tensor product, we we may bring it on the form $\sum_i m_i \otimes s_i^{-1}$. The trick is now to let $s = s_1 \cdots s_r$ and $t_i = ss_i^{-1} = s_1 \ldots \widehat{s_i} \ldots s_r$, and with this we find

$$x = \sum_i m_i t_i \otimes s^{-1} = (\sum_i m_i t_i) \otimes s^{-1} = m \otimes s^{-1},$$

with $m = \sum_i m_i t_i$. ❑

(7.41) Base change functors preserve tensor products (Proposition 6.37 on page 162) which combined with Proposition 7.40 above, yields that the localization process preserves tensor product:

PROPOSITION 7.42 *Let M and N be two A-modules and S a multiplicative set in A. Then there is a canonical isomorphism*

$$S^{-1}(M \otimes_A N) \simeq S^{-1}M \otimes_{S^{-1}A} S^{-1}N.$$

(7.43) When it comes to hom-sets, the behaviour is rather nice, at least for modules of finite presentation. In general, sending an A-linear map ϕ between two A-modules M and N to the localized map $S^{-1}\phi$ is an A-linear map $\operatorname{Hom}_A(M, N) \to \operatorname{Hom}_{S^{-1}A}(S^{-1}M, S^{-1}N)$. By the universal property of localization it extends to a map $S^{-1}\operatorname{Hom}_A(M, N) \to \operatorname{Hom}_{S^{-1}A}(S^{-1}M, N)$; and in case M is of finite presentation, this map is an isomorphism:

PROPOSITION 7.44 *Let M and N be two A modules and S a multiplicative set in A. Assume that M is of finite presentation. Then the canonical map*

$$S^{-1}\mathrm{Hom}_A(M,N) \xrightarrow{\simeq} \mathrm{Hom}_{S^{-1}A}(S^{-1}M, S^{-1}N)$$

induced by sending ϕ to $S^{-1}\phi$ is an isomorphism.

PROOF: Recall that both localization and the hom-functors are additive functors, hence the proposition holds true whenever M is a free A module of finite rank n; indeed, one finds

$$S^{-1}\mathrm{Hom}_A(nA,N) \simeq S^{-1}nN \simeq nS^{-1}N \simeq \mathrm{Hom}_{S^{-1}A}(nS^{-1}A, S^{-1}N), \tag{7.7}$$

where the isomorphisms are the natural ones (the one in the middle is an isomorphism since localization is additive, and the two others because hom-functors are additive). Since M is assumed to be of finite presentation, it lives in an exact sequence

$$mA \xrightarrow{\psi} nA \xrightarrow{\pi} M \longrightarrow 0, \tag{7.8}$$

with $m, n \in \mathbb{N}$ and where ψ and π are A-linear maps. Consider the diagram

$$\begin{array}{ccccccc}
0 & \longrightarrow & S^{-1}\mathrm{Hom}_A(M,N) & \longrightarrow & S^{-1}\mathrm{Hom}_A(nA,N) & \longrightarrow & S^{-1}\mathrm{Hom}_A(mA,N) \\
 & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & \mathrm{Hom}_{S^{-1}A}(S^{-1}M, S^{-1}N) & \longrightarrow & \mathrm{Hom}_{S^{-1}A}(nS^{-1}A, S^{-1}N) & \longrightarrow & \mathrm{Hom}_{S^{-1}A}(mS^{-1}A, S^{-1}N).
\end{array}$$

The upper sequence is obtain from (7.8) by applying the functor $S^{-1}\mathrm{Hom}_A(-,N)$ to it, and it is therefore exact by left exactness of hom-functors and exactness of the localization functor. The bottom sequence comes from (7.8) with the functor $\mathrm{Hom}_{S^{-1}A}(S^{-1}(-), S^{-1}N)$ applied to it, and is exact for the same reasons. The vertical maps are the canonical maps induced by sending maps ϕ to $S^{-1}\phi$, and it is a matter of simple verification that the squares commute.

Now, the final point is that the two rightmost maps are isomorphisms by the beginning of the proof, and then the Five Lemma tells us that the third map is an isomorphisms as well, which is precisely what we aim at proving. ❑

The first part of the proof fails when M is not of finite presentation; for instance, when $M = \bigoplus_{i\in\mathbb{N}} A$, one has

$$\mathrm{Hom}_A(\bigoplus_{i\in\mathbb{N}} A, A) = \prod_{i\in\mathbb{N}} \mathrm{Hom}_A(A,A) = \prod_{i\in\mathbb{N}} A,$$

and infinite products do not in general commute with localization as we saw in Exercise 7.34. The proof only relies on the two facts that the hom-functor's are left exact and

that localization is exact, and thus remains valid for any *flat* base change; or for that matter, the hom's may be replaced by any additive left exact functor that sends free modules to free modules of the same rank.

7.3 Local properties

Based on the belief that modules over local rings are simpler than others, a general technique is to try to pass from local knowledge—that is, knowledge of the localized modules $M_\mathfrak{p}$—to global knowledge. One envisages to infer properties of the module M itself from properties of the localized modules $M_\mathfrak{p}$.

Local properties (lokale egenskaper)

In this context it is quite natural to introduce the notion of a *local property*. A property of modules, call it P, is said to be a *local property* if all the localizations $M_\mathfrak{p}$ at prime ideals have P if and only if that the module M itself has P. Equally well, one may speak about local properties of homomorphisms of modules: Such a property P is local, if a map $\phi\colon M \to N$ has P whenever all localizations $\phi_\mathfrak{p}$ have P.

The localness of being zero

We shall see several instances of the local to global principle, but begin with the simplest of all properties, namely that of being zero! Applied to kernels and cokernels this leads to local criteria for homomorphisms to be injective or surjective.

(7.45) The point of departure is the following easy lemma which describes when elements remain non-zero in a localization.

LEMMA 7.46 *Let M be an A-module and x and element in M. Assume that $\mathfrak{p}$ is a prime ideal in A. Then x does not map to zero in $M_\mathfrak{p}$ if and only if* $\operatorname{Ann} x \subseteq \mathfrak{p}$.

PROOF: The module $M_\mathfrak{p}$ is M localized in the multiplicative set $S = A\backslash\mathfrak{p}$. Recall from Lemma 7.31 on page 190 that the image of x in $M_\mathfrak{p}$ being zero is equivalent to x being killed by an element in S; that is, by an element belonging to $\operatorname{Ann} x$ but not to $\mathfrak{p}$. ❑

This lemma immediately translates into the following fundamental principle:

PROPOSITION 7.47 (BEING ZERO IS A LOCAL PROPERTY) *An A-module M is equal to zero if and only if either of the two following assertions holds.*

i) $M_\mathfrak{m} = 0$ *for all maximal ideals* $\mathfrak{m}$ *in* A*;*

ii) $M_\mathfrak{p} = 0$ *for all prime ideals* $\mathfrak{p}$ *in* A.

PROOF: Two of the implications are obvious; localizing the zero module yields the zero module. For the rest, it suffices to show that the weaker condition in assertion *i)* implies that $M = 0$. To this end, assume that M is non-zero and let x be a non-zero element in M. The annihilator $\operatorname{Ann} x$ of x is then a proper ideal and is contained in a maximal

ideal $\mathfrak{m}$. By the simple lemma above, the image of x in $M_\mathfrak{m}$ is non-zero and *a fortiori* $M_\mathfrak{m}$ is non-zero. ❑

We have seen that localization is an exact operation and therefore it commutes with the formation kernels and cokernels of homomorphisms. In combination with the localness of being zero, this yields the following important local criterion for a map to be injective or surjective.

COROLLARY 7.48 (BEING INJECTIVE OR SURJECTIVE IS A LOCAL PROPERTY) *Assume M and N be two A-modules. Any A-linear map $\phi\colon M \to N$ is injective (respectively surjective) if and only if either of the two following equivalent conditions is satisfied.*

i) The localization $\phi_\mathfrak{m}\colon M_\mathfrak{m} \to N_\mathfrak{m}$ is injective (respectively surjective) for all maximal ideals $\mathfrak{m}$ in A;

ii) The localization $\phi_\mathfrak{p}\colon M_\mathfrak{p} \to N_\mathfrak{p}$ is injective (respectively surjective) for all prime ideals $\mathfrak{p}$ in A.

PROOF: Localization is an exact functor, so $(\ker\phi)_\mathfrak{m} = \ker(\phi_\mathfrak{m})$ for every maximal (respectively prime) ideal $\mathfrak{m}$, and Proposition 7.47 above tells us that $\ker\phi = 0$ if and only if $\ker\phi_\mathfrak{m} = 0$ for all $\mathfrak{m}$. This takes care of the part about injectivity, for the surjectivity part, one replaces $\ker\phi$ by $\operatorname{coker}\phi$. ❑

COROLLARY 7.49 (BEING AN ISOMORPHISM IS A LOCAL PROPERTY) *Let M and N be two A-modules. An A-linear mapping $\phi\colon M \to N$ is an isomorphism if and only if the localized map $\phi_\mathfrak{m}\colon M_\mathfrak{m} \to N_\mathfrak{m}$ is an isomorphism for all maximal ideals $\mathfrak{m}$, or equivalently, if and only if $\phi_\mathfrak{p}\colon M_\mathfrak{p} \to N_\mathfrak{p}$ is an isomorphism for all prime ideals $\mathfrak{p}$.*

(7.50) There are many other instances of local properties, but let us mention two, namely flatness and projectivity.

PROPOSITION 7.51 (FLATNESS IS A LOCAL PROPERTY) *An A-module M is flat over A if and only if $M_\mathfrak{p}$ is flat over $A_\mathfrak{p}$ for every prime ideal $\mathfrak{p} \in \operatorname{Spec} A$.*

PROOF: We are to check that an A-module M is flat over A if $M_\mathfrak{p}$ is flat over $A_\mathfrak{p}$ for all $\mathfrak{p} \in \operatorname{Spec} A$. So let

$$0 \longrightarrow N \xrightarrow{\phi} N' \tag{7.9}$$

be an injection. We are to prove that $\phi\otimes_A \mathrm{id}_M$ is injective. But for any A-module L and any prime ideal $\mathfrak{p}$ it holds $(L\otimes M)_\mathfrak{p} = L_\mathfrak{p}\otimes_{A_\mathfrak{p}} M_\mathfrak{p}$ and ditto, for any module homomorphism ψ one has $(\phi\otimes\mathrm{id}_M)_\mathfrak{p} = \phi_\mathfrak{p}\otimes_{A_\mathfrak{p}}\mathrm{id}_{M_\mathfrak{p}}$. Hence the maps in the two sequences

$$0 \longrightarrow (N\otimes_A M)_\mathfrak{p} \xrightarrow{(\phi\otimes\mathrm{id}_M)_\mathfrak{p}} (N'\otimes_A M)_\mathfrak{p} \tag{7.10}$$

and

$$0 \longrightarrow N_{\mathfrak{p}} \otimes_{A_{\mathfrak{p}}} M_{\mathfrak{p}} \xrightarrow{\phi_{\mathfrak{p}} \otimes_{A_{\mathfrak{p}}} \mathrm{id}_{M_{\mathfrak{p}}}} N'_{\mathfrak{p}} \otimes_{A_{\mathfrak{p}}} M_{\mathfrak{p}}$$

coincide. The latter is obtained from (7.9) by the two step process of first localizing in $\mathfrak{p}$, which is exact, and subsequently tensorizing by $M_{\mathfrak{p}}$ which also is exact since $M_{\mathfrak{p}}$ is assumed to be flat. The map in (7.10) is therefore injective, and citing Corollary 7.48 that injectivity is a local property of homomorphisms, we are through. ❑

Along the same lines one may show that being projective is a local property for finitely generated modules; it is true without the finiteness limitation, but we content ourselves with the case of finitely generated modules, and we also content ourselves with announcing the result, leaving the proof as an exercise.

PROPOSITION 7.52 (LOCALNESS OF BEING PROJECTIVE) *Let M be a finitely generated module. Then M is a projective A-module if and only if $M_{\mathfrak{p}}$ is a projective $A_{\mathfrak{p}}$-module for all prime ideals $\mathfrak{p}$.*

Exercises

(7.36) Prove Proposition 7.52. HINT: Follow the lines of the proof of Proposition 7.51. The isomorphism in Proposition 7.44 might be useful.

(7.37) Let $C_\bullet$ be a complex of A-modules. Show that localization commutes with taking homology; that is, show that for every prime ideal $\mathfrak{p} \in \operatorname{Spec} A$ one has canonical isomorphisms $H_i(C_\bullet) \otimes_A A_{\mathfrak{p}} \simeq H_i(C_\bullet \otimes_A A_{\mathfrak{p}})$. Conclude that being exact is a local property of complexes.

★

7.4 *An extended version of Nakayama's lemma*

We are prepared to revisit Nakayama's lemma as announced, and give a version whose proof relies on a localization technique. The extended version is valid for all ideals not only those lying in the Jacobson radical; but of course, when weakening the hypothesis, you get a weaker conclusion.

PROPOSITION 7.53 (NAKAYAMA EXTENDED) *Let $\mathfrak{a}$ be an ideal in the ring A, and assume that M is a finitely generated A-module satisfying $\mathfrak{a}M = M$. Then M is killed by an element of the form $1 + a$ with $a \in \mathfrak{a}$; that is, there is an $a \in \mathfrak{a}$ so that $(1 + a)M = 0$.*

PROOF: Let S be the multiplicative set $\{\, 1 + a \mid a \in \mathfrak{a} \,\}$. The ideal $S^{-1}\mathfrak{a}$ is contained in the Jacobson radical of $S^{-1}A$ (by Proposition 2.66 on page 54): indeed, we are to check that for each $a \in \mathfrak{a}$ and each $x \in A$ the element $1 + s^{-1}xa$ is invertible in $S^{-1}A$ whatever $s \in S$ is. But $1 + s^{-1}xa = s^{-1}(s + xa) = s^{-1}(1 + x'a' + xa)$ with $a' \in \mathfrak{a}$ and $x' \in A$, and both factors are invertible in $S^{-1}A$. By Nakayama classic (Proposition 4.52 on page 106)

we may then conclude that $S^{-1}M = 0$. Thence there is for each generator x_i of M, an element $s_i \in S$ killing x_i. The x_i's being finite in number we may form their product, which obviously kills M and is of the required form. ❑

7.5 *The support of a module*

Since prime ideals $\mathfrak{p}$ such that $M_\mathfrak{p} = 0$ are insignificant in the local–to–global process we just described, it is very natural to introduce the subset $\operatorname{Supp} M$ of $\operatorname{Spec} A$ consisting of the prime ideals $\mathfrak{p}$ so that $M_\mathfrak{p} \neq 0$; that is, we define

$$\operatorname{Supp} M = \{\, \mathfrak{p} \in \operatorname{Spec} A \mid M_\mathfrak{p} \neq 0 \,\}.$$

The support of a module (støtten til en modul)

It is called the *support* of M. In many cases, *e.g.* when M is finitely generated, the support of M is a closed subset of $\operatorname{Spec} A$, but whether M is finitely generated or not, it has the weaker property of being *closed under specialization*; that is, with every $\mathfrak{p}$ lying in $\operatorname{Supp} M$ the closed set $V(\mathfrak{p})$ lies there.

When M and N are finitely generated, the support of the direct sum* $M \oplus N$ equals the union $\operatorname{Supp} M \cup \operatorname{Supp} N$ and the support of the tensor product $N \otimes_A M$ equals the intersection $\operatorname{Supp} N \cap \operatorname{Supp} M$. It is worth while observing that the support takes the two "ring-like-operations" direct sum and tensor product in the category of finitely generated A-modules into the two operations union and intersection of the boolean ring of closed subsets of $\operatorname{Spec} A$.

*More generally this is the case for any extension of M by N.

Closure properties of the support

(7.54) For a cyclic module $M = A/\mathfrak{a}$ the support coincides with the closed set $V(\mathfrak{a})$ associated with the ideal $\mathfrak{a}$ since a prime ideal $\mathfrak{p}$ belongs to $V(\mathfrak{a})$ precisely when $(A/\mathfrak{a})_\mathfrak{p} \neq 0$, just apply the simple lemma 7.46 to the element 1 in $A/\mathfrak{a}$. This observation may be generalized to finitely generated modules. Any such module has a support which is a closed subset of $\operatorname{Spec} A$:

PROPOSITION 7.55 *If M is finitely generated A-module, the support* $\operatorname{Supp} M$ *equals the closed subset* $V(\operatorname{Ann} M)$ *of* $\operatorname{Spec} A$*; that is, it consists of the prime ideals* $\mathfrak{p}$ *containing* $\operatorname{Ann} M$*.*

PROOF: Our task is to show that $M_\mathfrak{p} \neq 0$ if and only if $\operatorname{Ann} M \subseteq \mathfrak{p}$, or equivalently, that $M_\mathfrak{p} = 0$ if and only if $\operatorname{Ann} M \nsubseteq \mathfrak{p}$. In case an element $a \in A$ not belonging to $\mathfrak{p}$ kills M, it holds that $M_\mathfrak{p} = 0$; indeed, a becomes invertible in $A_\mathfrak{p}$. This takes care of the if part of the proof. To attack the only if part, assume that $M_\mathfrak{p} = 0$, and let $x_1, \ldots, x_r$ be generators of M. By Lemma 7.46 above, there is for each of the x_i's an element s_i not in $\mathfrak{p}$ killing x_i. The product of the s_i's clearly kills M and does not belong to the prime ideal $\mathfrak{p}$ since none of the s_i's does; hence $\operatorname{Ann} M$ is not contained in $\mathfrak{p}$. ❑

(7.56) The hypothesis that M be finitely generated was used only in the last part of the proof above, and for a general module M it holds true that $\operatorname{Supp} M \subseteq V(\operatorname{Ann} M)$:

from $M_\mathfrak{p} \neq 0$ follows that $M_\mathfrak{q} \neq 0$ whenever $\mathfrak{q} \supseteq \mathfrak{p}$. Indeed, if $\operatorname{Ann} x \subseteq \mathfrak{p}$, it obviously holds that $\operatorname{Ann} x \subseteq \mathfrak{q}$, and an element x mapping to a non-zero element in $M_\mathfrak{p}$ maps to a non-zero element in $M_\mathfrak{q}$. We have thus established

PROPOSITION 7.57 *Let M be an A-module. The support of M is closed under specialization; that is, for each prime ideal $\mathfrak{p} \in \operatorname{Supp} M$ it holds that $V(\mathfrak{p}) \subseteq \operatorname{Supp} M$.*

Examples

(7.18) As already observed, the support of a cyclic module $A/\mathfrak{a}$ equals the closed set $V(\mathfrak{a})$.

(7.19) One has $\operatorname{Supp} \mathbb{Q} = \operatorname{Spec} \mathbb{Z}$ since $S^{-1}\mathbb{Q} = \mathbb{Q}$ for any multiplicative set S in $\mathbb{Z}$. More generally, for the fraction field K of any domain A is of global support; that is, $\operatorname{Supp} K = \operatorname{Spec} A$.

(7.20) An example of the failure for "large modules" of the support being the closed set defined by the annihilator, can be the $\mathbb{Z}$-module $\mathbb{Z}_{p^\infty} = \mathbb{Z}[p^{-1}]/\mathbb{Z}$, where p is a prime.

Each element of $\mathbb{Z}_{p^\infty}$ is the class of a rational number of the form $x = a/p^r$ with a prime to p. Since $yx \in \mathbb{Z}$ if and only if y is divisible by p^r, one has $\operatorname{Ann} x = (p^r)$, and from Lemma 7.46 above it follows that $\operatorname{Supp} \mathbb{Z}_{p^\infty} = \{(p)\}$.

Even though every element of $\mathbb{Z}_{p^\infty}$ is killed by a power of p, the annihilator of $\mathbb{Z}_{p^\infty}$ reduces to the zero ideal because no power of p kills the entire module $\mathbb{Z}_{p^\infty}$ (a power p^r kills the class of p^{-n} only if $n \leqslant r$). This shows that $\operatorname{Supp} \mathbb{Z}_{p^\infty}$, although being closed, differs from $V(\operatorname{Ann} \mathbb{Z}_{p^\infty})$.

(7.21) The support of a module is not always a closed subset of $\operatorname{Spec} A$. Take any infinite sequence of primes p_i which does not including all primes—for instance, every second prime—and consider the module $M = \bigoplus_i \mathbb{Z}/p_i\mathbb{Z}$. The support of M is the infinite subset $\{(p_i)\}$. The only infinite closed subset of $\operatorname{Spec} \mathbb{Z}$ being the entire spectrum, this set is not closed.

☆

The support of extensions

(7.58) Since localization is an additive functor so that $(M \oplus N)_\mathfrak{p} \simeq M_\mathfrak{p} \oplus N_\mathfrak{p}$ it is obvious that the support of a direct sum of two A-modules is the union of their supports. This generalizes to so-called *extensions*; that is, modules in the midst of an exact sequence (which is not necessarily split exact).

Extensions of modules (utvidelser av moduler)

PROPOSITION 7.59 *Assume that*

$$0 \longrightarrow N \longrightarrow M \longrightarrow L \longrightarrow 0$$

is an exact sequence of A-modules. Then $\operatorname{Supp} M = \operatorname{Supp} N \cup \operatorname{Supp} L$.

Proof: The proposition follows immediately from the localization functor being exact. For each prime $\mathfrak{p}$ the localized sequence

$$0 \longrightarrow N_{\mathfrak{p}} \longrightarrow M_{\mathfrak{p}} \longrightarrow L_{\mathfrak{p}} \longrightarrow 0$$

is exact, and the middle module vanishes if and only if the two extreme ones do. ❑

The support of a tensor product

(7.60) The aim of this paragraph is the prove that the support of a tensor product is the intersection of the supports of the two factors, at least when the involved modules are finitely generated.

Proposition 7.61 *Let M and N be two finitely generated A-modules. Then the following equality holds true* $\operatorname{Supp} M\otimes_A N = \operatorname{Supp} N \cap \operatorname{Supp} M$.

The basic argument takes place over a local ring:

Lemma 7.62 *Let A be a local ring with maximal ideal $\mathfrak{m}$ and let M and N be two finitely generated A-modules. Then $M\otimes_A N = 0$ if and only if either $N = 0$ or $M = 0$.*

Proof: The proof is an application of Nakayama's lemma. Let $k = A/\mathfrak{m}$ be the residue class field of A. Assume that both N and M are non-zero. Nakayama's lemma then ensures that both $N\otimes_A k$ and $M\otimes_A k$ are non-zero, and since base change respects tensor products (Proposition 6.37 on page 162), one has

$$(M\otimes_A N)\otimes_A k = (M\otimes_A k)\otimes_k (N\otimes_A k).$$

The tensor product of two non-zero vector spaces being non-zero (*e.g.* Proposition 6.21 on page 155), we infer that $(M\otimes_A N)\otimes_A k \neq 0$, and hence $N\otimes_A M \neq 0$ *a fortiori*. ❑

Proof of Proposition 7.61: The localized modules $N_{\mathfrak{p}}$ and $M_{\mathfrak{p}}$ are finitely generated over $A_{\mathfrak{p}}$ whenever M and N are finitely generated over A, and in view of the isomorphism

$$(M\otimes_A N)_{\mathfrak{p}} \simeq M_{\mathfrak{p}}\otimes_{A_{\mathfrak{p}}} N_{\mathfrak{p}},$$

the proposition then follows from the lemma. ❑

Example 7.22 Proposition Proposition 7.61 may fail when one of the factors is not finitely generated. For instance, if one factor equals the fraction field K of a domain A and the other is of the form $A/\mathfrak{a}$ where $\mathfrak{a}$ is a non-trivial proper ideal, it holds true that $A/\mathfrak{a}\otimes_A K = 0$; hence $\operatorname{Supp} A/\mathfrak{a}\otimes_A K = \varnothing$, but the fraction field K is of global support (one has $K_{\mathfrak{p}} = K$ for all $\mathfrak{p} \in \operatorname{Spec} A$) so that $\operatorname{Supp} K \cap \sup A/\mathfrak{a} = V(\mathfrak{a})$, which is non-empty. ☆

Exercises

(7.38) Let M be a finitely generated A-module. Prove that if $M \otimes_A A/\mathfrak{m} = 0$ for all maximal ideals $\mathfrak{m}$ in A, then $M = 0$. HINT: Combine Nakayama's lemma with Proposition 7.47 on page 196.

Groups of bounded exponent (grupper med begenset eksponent)

(7.39) An abelian group M is said to be of *bounded exponent* if some power p^n of a prime p kills every element of M. Give an example of a group of bounded exponent that is not finitely generated. Prove that if M is of bounded exponent, then $\operatorname{Supp} M = V(\operatorname{Ann} M)$.

(7.40) Let p be a prime number and let M be the abelian group $M = \bigoplus_{i \in \mathbb{N}_0} \mathbb{Z}/p^i\mathbb{Z}$. Determine the annihilator $\operatorname{Ann} M$ and the support $\operatorname{Supp} M$.

★

7.6 The rank of a module

In this section M is a finitely presented module over a ring A. Recall that this means that M lives in a short exact sequence

$$F \longrightarrow E \xrightarrow{\phi} M \longrightarrow 0$$

where E and F are free A-modules of finite rank. Let furthermore $\mathfrak{p}$ be a prime ideal in A and $k(\mathfrak{p})$ is the fraction field of $A/\mathfrak{p}A$; *i. e.* $k(\mathfrak{p}) = A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$. One defines the local rank $\operatorname{rk}_\mathfrak{p} M$ of M at $\mathfrak{p}$ as the dimension $\operatorname{rk}_\mathfrak{p} M = \dim_{k(\mathfrak{p})} M \otimes_A k(\mathfrak{p})$.

We conclude this chapter by taking a closer look at the special case when A is a domain and the prime ideal is the zero ideal. That is $A_\mathfrak{p}$ is the fraction field K of A. Then $M \otimes_A K$ will be a vector space over K, and the dimension $\dim_K M \otimes_A K$ is called *the rank of* M. The properties of the localization functor translates into properties of the rank, it will have the nice properties of being both additive and multiplicative:

The rank of a module (rangen til en modul)

THEOREM 7.63 *Let A be a domain with fraction field K and let M and N be two A-modules. It then holds true that*

i) $\operatorname{rk} M = 0$ *if and only if M is a torsion module;*
ii) $\operatorname{rk} M = \operatorname{rk} N + \operatorname{rk} M/N$ *when* $N \subseteq M$;
iii) $\operatorname{rk} M \otimes_A N = \operatorname{rk} M \cdot \operatorname{rk} N$.
iv) If M is finitely presented then $\operatorname{rk} \operatorname{Hom}_A(M, N) = \operatorname{rk} M \cdot \operatorname{rk} N$.

PROOF: As already said, this follows from the properties of the localization functor combined with the appropriate properties of the dimension of a vector space. ❑

Lecture 8

Projective modules

In paragraph 5.18 we introduced the projective modules as those A-modules P such that $\mathrm{Hom}_A(P, -)$ is an exact functor, and we showed that P is projective if and only if it is the direct summand in a free module. The projective modules play a special and important role both in number theory and algebraic geometry. Certainly more involved than free modules, but still to a great extent maniable, they enjoy a series of good properties. In many cases they furnish important invariants of rings, and they are the supermen of homological algebra, where they among other things serve to define the so-called Ext- and Tor-groups, which describe respectively the 'missing cokernels" and the "missing kernels" we mentioned in Paragraph 5.12 on page 129.

Contrary to what usually is the case, infinitely generated projective modules turn out to be much simpler than the finitely generated ones. A famous result of Hyman Bass's ([**?**]) states that over Noetherian rings with connected* spectra projective modules requiring infinitely many generators are in fact free. We shall not treat Bass's theorem, but it justifies largely that we mostly work with finitely generated projectives.

*The exceptions will be free over each connected component of Spec A, but can of course have bases with different cardinality on different components (and can even be zero on some of them).

Locally free modules (lokalt frie moduler)

There is also a result of Kaplansky's pointing in the same direction as Bass's, but without the Noetherian hypothesis. It asserts that over a local rings all projective modules are free. A projective module P thus has the virtue of being *locally free*; that is, the localization $P_\mathfrak{p}$ is a free $A_\mathfrak{p}$-module for all primes $\mathfrak{p} \in \mathrm{Spec}\ A$. We shall give a proof of this when P is finitely generated, which is a classical application of Nakayama's lemma.

The geometrical counterpart to the locally free modules are the so-called *vector bundles*. In topology these are continuous maps $E \to X$ with all fibres being vector spaces (either real or complex) which are locally trivial; that is, over a suitable open cover $\{U_i\}$ of X it appears (in the complex case) as the projection $\mathbb{C}^n \times U_i \to U_i$ (additionally, there is also an important requirement that the transition functions arising over the intersections $U_i \cap U_j$ all be linear).

8.1 *Projective and locally free modules*

Being projective is a local property of a module as we established in Proposition 7.52 on page 198, but being free is not local—in Examples 5.5, 5.6 and 5.8 on page 133 we exhibited projective modules that are not free, but they are easily checked to be locally free. This illustrates the general fact that the locally free modules, at least among those that are finitely generated are exactly the projective ones. To establish this as a generally valid principle one merely needs to show that over local rings finitely generated projective modules are free:

PROPOSITION 8.1 *Let A be a local ring and P a finitely generated projective module. Then P is free.*

PROOF: This is a classical application of Nakayama's lemma. Let k be the residue field of A and consider $P \otimes_A k$. It is a finite vector space over k and has a basis, say with r elements. Lifting the basis elements to elements in A we obtain a map $\phi\colon rA \to P$ such that $\phi \otimes \mathrm{id}_k$ is an isomorphism, and Nakayama's lemma yields that ϕ is surjective. The kernel of ϕ lives in the short exact sequence

$$0 \longrightarrow \ker\phi \longrightarrow rA \xrightarrow{\phi} P \longrightarrow 0,$$

and the module P being projective the sequence is split and hence stays exact when tensorized by k. Again since $\phi \otimes \mathrm{id}_k$ is an isomorphism, it follows that $\ker\phi \otimes_A \mathrm{id}_k = 0$. Now, any direct summand in a finitely generated module is finitely generated. Therefore Nakayma's lemma applies to $\ker\phi$, and we may infer that $\ker\phi = 0$, which is exactly what we need to conclude that $P \simeq rA$; hence P is free. ❏

As announced, the proposition gives the following corollary:

COROLLARY 8.2 *Let A be a ring and P a finitely generated A-module. Then P is projective if and only it is locally free; that is, if and only if $P_\mathfrak{p}$ is a free $A_\mathfrak{p}$-module for all $\mathfrak{p} \in \operatorname{Spec} A$.*

(8.3) A local basis for P at a prime ideal $\mathfrak{p}$ can be extended to a basis for P over an open and distinguished neighbourhood of $\mathfrak{p}$ in Spec A, and we have the following slightly stronger result than Proposition 8.1.

PROPOSITION 8.4 *Let P be a finitely generated projective module over the ring A. Then there exist a finite set $\{f_i\}_{i\in I}$ of elements from A so that the distinguished open sets $D(f_i)$ cover* Spec *A and such that each localization P_{f_i} is a free module over A_{f_i}.*

PROOF: Since Spec A is quasi-compact (see Exercise 2.58 on page 66) it suffices to find a distinguished neighbourhood round each point $\mathfrak{p}$ over which P is free; in clear text, given a prime ideal $\mathfrak{p}$ we search for an element $f \notin \mathfrak{p}$ such that P_f is free as an A_f-module.

To find such an element, begin with a basis $\{a_i\}$, with say r elements, for the localized module $P_\mathfrak{p}$ over $A_\mathfrak{p}$ whose elements belong to P. Such a basis defines a map $\phi\colon rA \to P$, and it lives in the exact sequence

$$0 \longrightarrow \ker\phi \longrightarrow rA \xrightarrow{\phi} P \longrightarrow \operatorname{coker}\phi \longrightarrow 0.$$

Now, $\operatorname{coker}\phi$ is finitely generated (since P is); its support equals $V(\operatorname{Ann}\operatorname{coker}\phi)$ and does not contain $\mathfrak{p}$. Hence the annihilator ideal $\operatorname{Ann}\operatorname{coker}\phi$ is non-zero, and we let g be any one of its non-zero elements. Over the localization A_g the map ϕ_g is surjective so that the kernel $\ker\phi_g$ is a direct summand in rA_g and thus it is finitely generated. The support equals $V(\operatorname{Ann}\ker\phi_g)$, which is not the entire $\operatorname{Spec} A_g$ (it does not contain $\mathfrak{p}$), and we may find a non-zero element $h \in \operatorname{Ann}\operatorname{coker}\phi_g \cap A$. Then $f = gh$ is your man. ❑

The rank of projective modules

(8.5) Suppose we are given a finitely generated projective module P. At any point $\mathfrak{p}$ in the spectrum $\operatorname{Spec} A$ the module P being locally free has a *local rank* $r_\mathfrak{p}(P)$, namely the non-negative integer r so that $P_\mathfrak{p} \simeq rA_\mathfrak{p}$. This local rank may vary, it can assume different values along different connected components of the spectrum $\operatorname{Spec} A$ (see Example 8.1), but when it is constant, it is simply called the *rank* of P. This is *e.g.* the case whenever $\operatorname{Spec} A$ is connected. In fact, Proposition 8.4 above, about extension of local bases, yields that the rank is a locally constant function on $\operatorname{Spec} A$:

The local rank (den lokale rangen)

The rank of projective modules (rangen til projektive moduler)

PROPOSITION 8.6 (THE RANK IS LOCALLY CONSTANT) *Assume that P is a finitely generated projective module over A. Then the rank $\operatorname{rk}_\mathfrak{p}(P)$ is locally constant; that is, for each r the set $U_r = \{\,\mathfrak{p} \in \operatorname{Spec} A \mid \operatorname{rk}_\mathfrak{p} P = r\,\}$ is both open and closed. In particular, if $\operatorname{Spec} A$ is connected, the local rank is constant.*

PROOF: That U_r is open for all r ensues from 8.4, and hence the complement $U_r^c = \bigcup_{s\neq r} U_s$ is open as well. ❑

EXAMPLE 8.1 When the spectrum is not connected, it is easy to find projective modules whose local rank takes on different values on different connected components. These will also be examples of projective modules that are not free. The simplest example is a direct product $A \times B$ of two non-null rings A and B (the most minimalistic example was already given in Example 5.5). The spectrum $\operatorname{Spec}(A \times B)$ equals the disjoint union $\operatorname{Spec} A \cup \operatorname{Spec} B$. Both A and B are natural $A \times B$-modules—realized as $A \times (0)$ and $(0) \times B$—and as such are direct summands in $A \times B$, thus they are projective. But for instance, $r_\mathfrak{p}(A) = 1$ for $\mathfrak{p} \in \operatorname{Spec} A$ and $r_\mathfrak{p}(A) = 0$ for $\mathfrak{p} \in \operatorname{Spec} B$. ☆

Modules of constant rank

(8.7) A projective module P is, as we just saw, free over each of the local rings $A_\mathfrak{p}$, and of course, it stays free when tensored with the quotient $A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$: When P is of rank

r, it holds true that $P \otimes_A A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p} \simeq rA_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$ for each prime ideal $\mathfrak{p}$; hence "the fibre dimensions" $\dim_{k(\mathfrak{p})} P \otimes_A k(\mathfrak{p})$ will all be the same and equal to r (where, we remind you, $k(\mathfrak{p})$ denotes the fraction field of $A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$). When A is reduced, the converse holds true too; that is, if a finitely generated module has constant "fibre dimenison", it will be projective. Actually, only the maximal and the minimal primes come into play. Recall that when A is reduced, the local rings $A_\mathfrak{p}$ at the minimal primes $\mathfrak{p}$ of A are fields (Proposition 7.29 on page 187), and moreover he set of zero divisors in A is precisely the union of the minimal prime ideals.

PROPOSITION 8.8 *Assume that A is a reduced ring with finitely many minimal prime ideals and let P be a finitely generated A-module. Let r be a natural number. Assume that $\dim_{A/\mathfrak{m}} P \otimes_A A/\mathfrak{m} = \dim_{A_\mathfrak{p}} P \otimes_A A_\mathfrak{p} = r$ for all maximal ideals $\mathfrak{m}$ and all minimal prime ideals $\mathfrak{p}$ in A. Then P is projective of rank r.*

PROOF: Since being projective is a local property, it suffices to show the proposition when A is local. Let $\mathfrak{m}$ be the maximal ideal of A and $k = A/\mathfrak{m}$ the residue field. The k-vector space $P \otimes_A k$ has a basis of r element, which may be lifted to elements in P. By Nakayama's lemma these elements generate P, so that P lives in a short exact sequence shape liked

$$0 \longrightarrow M \longrightarrow rA \longrightarrow P \longrightarrow 0,$$

which when localized at a minimal prime $\mathfrak{p}$, becomes the short exact sequence

$$0 \longrightarrow M_\mathfrak{p} \longrightarrow rA_\mathfrak{p} \longrightarrow P_\mathfrak{p} \longrightarrow 0.$$

Now, $A_\mathfrak{p}$ is a field over which $P_\mathfrak{p}$ is assumed to be a vector space of dimension r, and therefore $M_\mathfrak{p} = 0$. We contend that this implis that $M = 0$; aiming at a contradiction, assume the contrary and pick a non-zero element $x \in M$.

Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_t$ be the minimal prime ideals in A. Since $M_{\mathfrak{p}_i} = 0$, it holds that $\operatorname{Ann} x \nsubseteq \mathfrak{p}_i$ and from Prime Avoidance it ensues that $\operatorname{Ann} x \nsubseteq \mathfrak{p}_1 \cup \ldots \cup \mathfrak{p}_r$, so there is a non-zero divisor killing x. But M is contained in rA and having non-zero element killed by a non-zero divisor, would be absurd, consequently $M = 0$. ❑

EXAMPLE 8.2 The proposition does not hold for modules over the simplest non-reduced ring $\mathbb{Z}/4\mathbb{Z}$. The quotient $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ is not projective (no submodule of a free $\mathbb{Z}/4\mathbb{Z}$-module is killed by 2), but the rank is one everywhere (well, everywhere is not very widespread; the ideal (2) is the only prime ideal). A similar example would be $A = k[x]/(x^2)$ where k is a field. The residue field $k = A/(x)$ is not projective, but of rank one everywhere. ☆

Exercises

(8.1) Let $A = \prod_i k_i$ be a finite product of fields. Show that any A-module is projective. Single out those which are free.

(8.2) Let I be a finite set and for each $i \in I$ let A_i be a local ring with residue field k_i. Describe the projective modules over $\prod_i A_i$.

★

8.2 *Working formulas*

The day-to-day working formulas for projective modules basically say that tensor products and direct sums of projective modules are projective. One might be tempted to say that the category of projective A-modules is "ring–like" a point of view pursued in Exercise 8.6 below. Since probably most projective modules one would meet practising algebraic geometry or commutative algebra, are finitely generated (and in fact, in view of Bass' result, the only non-trivial case), we just treat those.

For free modules the formulas are already established (trivial for sums, see Paragraph 6.20 for products), and bearing in mind that tensor products and hom's in good cases commute with localization, the formulas follow easily from the fact that being locally free is equivalent to being projective when the involved modules are finitely generated.

The dual module (den dual modulen)

(8.9) Before proving the working formulas, we need two new notions. The *dual* of a module M is denoted M^* and is defined as $M^* = \mathrm{Hom}_A(M, A)$. Sending M to M^* is an additive and contravariant functor from Mod_A to itself, and the double dual M^{**} will be a covariant and additive endofunctor of Mod_A. For each M there is a canonical evaluation map $\gamma_M \colon M \to M^{**} = \mathrm{Hom}_A(\mathrm{Hom}_A(M, A), A)$ defined by the assignment $x \mapsto (\phi \mapsto \phi(x))$.

Dual basis (dual basis)

If E is a finitely generated free module with basis $\{e_i\}_{1 \leqslant i \leqslant r}$, the dual module E^* is free with basis the so-called *dual basis* $\{\hat{e}_i\}$. It is defined by $\hat{e}_i(e_j) = \delta_{ij}$, and the easily verified formula $\phi = \sum_i \phi(e_i)\hat{e}_i$, shows that it indeed is a basis.

Furthermore, for every pair of A-modules M and N, there is a canonical map $\rho_{M,N} \colon M^* \otimes_A N \to \mathrm{Hom}_A(M, N)$, which on decomposable tensors is defined by the assignment $\phi \otimes x \mapsto (y \mapsto \phi(y)x)$.

LEMMA 8.10 (DUAL OF A FREE MODULE) *For a finitely generated free module E, the map γ_E is an isomorphism. Moreover, if F is another finitely generated free module, the map $\rho_{E,F}$ is an isomorphism.*

PROOF: The map γ_E sends the basis element e_i to the map $\phi \mapsto \phi(e_i)$. Hence, one has the formula $\phi = \sum_i \gamma_E(e_i)(\phi)\hat{e}_i$, which shows that γ_E is an isomorphism.

Letting $\{f_j\}$ be a basis for F, it holds that $\rho_{E,F}$ sends $\hat{e}_i \otimes f_j$ to the map $e_v \mapsto \delta_{v,i} \cdot f_j$. Hence, if $\phi\colon E \to F$ has matrix (a_{ij}), it holds that $\phi = \sum_{i,j} a_{ij} \cdot \rho_{E,F}(\hat{e}_i \otimes f_j)$, from which we deduce that $\rho_{E,F}$ is an isomorphism. ❑

(8.11) We are now well prepared for the working formulas for finitely generated projective modules:

PROPOSITION 8.16 *Let P and Q be two finitely generated projective modules over the ring A. Moreover, let $\mathfrak{p}$ denote a prime ideal in A.*

i) The direct sum $P \oplus Q$ is projective and $r_{\mathfrak{p}}(P \oplus Q) = r_{\mathfrak{p}}(P) + r_{\mathfrak{p}}(Q)$;

ii) The tensor product $P \otimes_A Q$ is projective, and $r_{\mathfrak{p}}(P \otimes_A Q) = r_{\mathfrak{p}}(P) r_{\mathfrak{p}}(Q)$;

iii) The dual module $P^* = \operatorname{Hom}_A(P, A)$ is projective and $r_{\mathfrak{p}}(P) = r_{\mathfrak{p}}(P^*)$. The canonical evaluation map γ_P yields an isomorphism $P \simeq P^{**}$;

iv) The module $\operatorname{Hom}_A(P, Q)$ is projective, and the canonical map $\rho_{P,Q}$is an isomorphism $P^* \otimes_A Q \simeq \operatorname{Hom}_A(P, Q)$.

PROOF: As noted above, these statements follows from the facts that a module is projective if and only if it is locally free (*i. e.* $P_{\mathfrak{p}}$ is free over $A_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{Spec} A$) and that the corresponding statements hold for free modules, together with the good behaviour of tensor products and hom-modules with respect to localization.

The first statement *i)* is clear.

To prove statement *ii)* recall that base change respects tensor productucts (Proposition 6.37 on page 162), so it holds that $(P \otimes_A Q)_{\mathfrak{p}} = P_{\mathfrak{p}} \otimes_{A_{\mathfrak{p}}} Q_{\mathfrak{p}}$. And when $P_{\mathfrak{p}}$ and $Q_{\mathfrak{p}}$ both are free, it follows from Corollary 6.22 on page 155 that $(P \otimes_A Q)_{\mathfrak{p}}$ is free of rank $r_{\mathfrak{p}}(P) r_{\mathfrak{p}}(Q)$

Statement *iii)* follows since by Proposition 7.44 on page 195 it holds that $(P^*)_{\mathfrak{p}} = (P_{\mathfrak{p}})^*$ and the latter is free of rank $r_{\mathfrak{p}}(P)$. Moreover, Lemma 8.10 gives that $(\gamma_P)_{\mathfrak{p}} = \gamma_{P_{\mathfrak{p}}}$ is an isomorphism for all $\mathfrak{p}$, and hence $\gamma_{\mathfrak{p}}$ is an isomorphism because being an isomorphism is a local property.

Finally, statement *iv)* is a consequence of Lemma 8.10 and that forming tensor products and homomorphism modules commute with localization. ❑

EXAMPLE 8.3 Recall Exercise 4.27 on page 103 where you were asked to prove that the direct product $\prod_{i \in \mathbb{N}} \mathbb{Z}$ of countably many copies of $\mathbb{Z}$ is not a free $\mathbb{Z}$-module, thus giving an example of an infinite product of free modules which is not free. A slight extension of the proof indicated there shows that neither is $\prod_{i \in \mathbb{N}} \mathbb{Z}$ a projective $\mathbb{Z}$-module (check it!), so infinite direct products of projective modules are not always projective. ☆

Exercises

(8.3) Show that $P \to P^{**}$ is always injective, but that it is not necessarily an isomorphism when P is not finitely generated, even when the base ring is a field. HINT: Consider the

$\mathbb{Q}$-vector space $P = \bigoplus_{i\in\mathbb{N}} \mathbb{Q}$ and show that P and P^{**} are not of the same cardinality.

(8.4) Show that the tensor product $P \otimes_A Q$ of two projective modules is projective wether they are finitely generated or not. HINT: Use adjointness, Proposition 6.24 on page 156.

(8.5) *The Eilenberg swindle.* The simplistic behaviour of infinitely generated projectives, is to a large extent rooted in the so called *Eilenberg swindle*. Let F be a free A-module that is not finitely generated and assume that P is a direct summand in F. The swindle is the assertion that $P \oplus F \simeq F$. Let Q be a complement of P in F so that $F = Q \oplus P$.

a) Show that F is isomorphic to the direct sum of countably many copies of itself; that is, $F \simeq F \oplus F \oplus \dots$;
b) Show that $P \oplus F \simeq P \oplus Q \oplus P \oplus Q \oplus P \oplus \dots$;
c) Conclude that $P \oplus F \simeq F$. HINT: Swap parantheses.

(8.6) *K^0 of a ring.* At the top of the present subsection we alluded to the category of projective A-modules being "ring-like". It possesses a sum operation and a product operation and these satisfy formulas closely resembling the ring axioms, but notably only up to isomorphism and not up to equality; and of course, there is no subtraction. Passing to isomorphism classes repairs the first fault, and for the second, there is general technique to extend monoids and introduce a subtraction. One passes to so-called *virtual projective modules*; formally one introduces the ring $K^0(A)$ whose elements are finite linear combinations $\sum_i a_i[P_i]$ of isomorphism classes of finitely generated projective modules with the a_i's being integers (allowed to be negative). The ring operations comply to the rules $[P \oplus Q] = [P] + [Q]$ and $[P \otimes_A Q] = [P] \cdot [Q]$.

Virtual projective modules (øyensynlige projektive moduler)

The construction goes as follows: One begins with the free abelian group G with a basis the set of all isomorphism classes of finitely generated projective A-modules. Next, one considers the subgroup H of G generated by expressions $\overline{P \oplus Q} - \bar{P} - \bar{Q}$ where a bar indicates an isomorphism class, and introduces the underlying abelian group of $K^0(A)$ as the quotient $K^0(A) = G/H$. The class of a module P in $K^0(A)$ is designated by $[P]$. Then by construction $[P \oplus Q] = [P] + [Q]$.

a) Show that the asigment $[P] \cdot [Q] = [P \otimes_A Q]$ extends bilinearly to the entiere $K^0(A)$ and makes $K^0(A)$ a commutative ring with $[A]$ as unit element;
b) If $A \to B$ is a ring homomorphism, show that base change functor $P \mapsto P \otimes_A B$ induces a ring homomorphism $K^0(A) \to K^0(B)$. Show that this makes K^0 a functor;
c) Show that the local rank at a prime ideal $\mathfrak{p}$, is a ring homomorphism $r_{\mathfrak{p}} : K^0(A) \to \mathbb{Z}$.

(8.7) If one in the definition of $K^0(A)$ had allowed all projective modules, and not only the finitely generated ones, show that $K^0(A)$ would have been the zero ring. HINT: Consider the countable direct sum $P = \bigoplus_{i\in\mathbb{N}} A$ and show that $A \oplus P \simeq P$.

8.3 The Picard group

One of the most basic invariant of a variety is the so-called *Picard group*. The elements are isomorphism classes of creatures called invertible sheaves and whose algebraic avatars are the invertible modules which we are about to define. Among other things—and which might be their most important role—they govern maps from the variety to projective spaces. The analogue of the Picard groups in algebraic number theory are the so-called *ideal class groups*, a notion which actually predates the notion of the Picard group by about a century. It is of fundamental importance for the study of number fields and measures how far the ring of integers in a number field is from being factorial.

Ideal class groups (idealklassegrupper)

Invertible modules (invertible moduler)

(8.17) The tensor product $M\otimes_A N$ induces a binary operation on the set of isomorphism classes of A-modules, and the basic working formulas in Proposition 6.14 show it is associative and commutative and has the class of A as a neutral element. Modules do not in general have inverses, but there are good hopes that the tensor product will give a group law on the set of those that have. This motivates the notion of *invertible modules*: an A-module M is invertible if there is an A-module N such that $N\otimes_A M \simeq A$.

Émil Picard (1856–1941)

French mathematician

The invertible modules turn out to coincide with the finitely generated projective modules of rank one, or in view of Corollary 8.2 on page 204 the finitely generated modules which are locally free of rank one.

PROPOSITION 8.18 *Let A be a ring and M an A-module. The following three statements are equivalent:*

i) M is an invertible module;
ii) M is finitely generated and projective of rank one;
iii) M is finitely generated and locally free of rank one.

Moreover, if M is invertible, it holds true that the evaluation map gives an isomorphism $M\otimes_A M^* \simeq A$ so that the dual M^* serves as an inverse for M.

PROOF: The third assertion *iii)* is included only for completeness; its equivalence with *ii)* was established already in Corollary 8.2 on page 204.

We begin with proving that *i)* implies *ii)*; so assume that M is invertible and let N be such that $M\otimes_A N \simeq A$. Let us first establish that M is finitely generated. To that end, identify $M\otimes_A N$ and A and write $1 = \sum x_i\otimes y_i$ with $x_i \in M$ and $y_i \in N$. We contend that the x_i's generate M. Indeed, let M' be the submodule of M generated by the x_i's and consider the quotient M/M'. Since the inclusion $M' \to M$ induces a surjection $M'\otimes_A N \to M\otimes_A N$, it holds that $M/M'\otimes_A N = 0$, and consequently

$$M/M' \simeq M/M'\otimes_A A \simeq M/M'\otimes_A (N\otimes_A M) \simeq (M/M'\otimes_A N)\otimes_A M = 0.$$

We proceed to show that M is projective. Since being projective is a local affair, we may as well suppose that A is a local ring. Let $\mathfrak{m}$ be the maximal ideal and $k = A/\mathfrak{m}$ the residue field. Using that base change preserves tensor products (Proposition 6.37 on page 162) we find

$$(M\otimes_A N)\otimes_A k \simeq (M\otimes_A k)\otimes_k (N\otimes_A k) \simeq k\otimes_k k \simeq k.$$

We conclude that $\dim_k M\otimes_A k = 1$, and thus $M\otimes_A k \simeq k$. Nakayama's lemma then gives that M is monogenic so that we may write $M \simeq A/\mathfrak{a}$ for some ideal $\mathfrak{a}$. Then $\mathfrak{a}$ kills M, and therefore also A as $A \simeq M\otimes_A N$, which is absurd unless $\mathfrak{a} = (0)$.

We assume next that M is finitely generated and projective of rank one and aim at showing the isomorphism $M\otimes_A M^* \simeq A$. Being an isomorphism is a local property (Corollary 7.49 on page 197) and localization commutes with forming hom's (Proposition 7.44 on page 195) so we may certainly assume that A is local and that $M = A$ (M is locally free of rank one). In that setting the evaluation map appears as a map $A\otimes_A \mathrm{Hom}_A(A,A) \to A$, which in view of the harmless identity $A = \mathrm{Hom}_A(A,A)$, is nothing but the map $A\otimes_A A \to A$ that sends $a\otimes b$ to ab; and that is surely an isomorphism. This also proves the final statement in the proposition. ❑

(8.19) According to the proposition the inverse of an invertible module is well defined and hence the set $\mathrm{Pic}\, A$ of their isomorphism classes is an abelian group when equipped with the tensor product as a group law. It is called the *Picard group*. Summed up we have:

The Picard group (Picard-gruppen)

PROPOSITION 8.20 *The set* $\mathrm{Pic}\, A$ *formed by the isomorphism classes of invertible modules is an abelian group. The product of the classes of* P *and* Q *equals the class of* $P\otimes_A Q$*. The neutral element is the class of* A*, and the inverse of the class of* P *is the class of* $P^* = \mathrm{Hom}_A(P,A)$*.*

Whenever $A \to B$ *is a ring-homomorphism, the base change functor* $(-)\otimes_A B$ *takes* $\mathrm{Pic}\, A$ *into* $\mathrm{Pic}\, B$ *making* $\mathrm{Pic}\colon \mathsf{Rings} \to \mathsf{Ab}$ *a functor.*

PROOF: Merely the last statement remains to be commented, and it hinges on the base change functor respecting the tensor product. If $P\otimes_A Q \simeq A$, we find

$$(P\otimes_A B)\otimes_B (Q\otimes_A B) \simeq (P\otimes_A Q)\otimes_A B \simeq A\otimes_A B = B.$$

❑

Fractional and invertible ideals

Over an integral domain A, there is a large class of projective modules of rank one, which we are about to introduce, formed by the so-called *invertible ideals*. Up to a certain equivalence, the invertible ideals constitute a group—*the ideal class group*—which turns out to be isomorphic of the Picard group. Contrary to the invertible modules which

suffer from a certain elusiveness, invertible ideals are concrete, being submodules of the fraction field, and in many cases are a lot easier to lay hands on. The Picard group, on the other hand, generalizes well: many geometric objects like varieties and schemes, are inhabited by creatures called *invertible sheaves* whose isomorphism classes constitute their Picard group.

In Kummer's set-up, with his ideal numbers in centre stage and where the numbers are represented by the principal ideals, fractions will most naturally be represented by principal submodules of the fraction field K of A; in other words, A-submodules of K requiring a single generator. The obvious "idealizations" are the A-submodules of K, which when complying to a minor condition, are called *fractional ideals*. Fractional ideals can in a natural way be added and multiplied, and the invertible ideals are those that possess an inverse.

Fractional ideals (brudne idealer)

(8.21) The precise definition is as follows: An A-submodule $\mathfrak{a} \subseteq K$ is called a *fractional ideal* if there is some non-zero $x \in A$ such that $x\mathfrak{a} \subseteq A$; one may think about x as a common denominator for the elements in $\mathfrak{a}$. Just as for ideals, the fractional ideal $\mathfrak{a}$ is said to be *principal* if it generated by a single element. That is, if it is shaped like $\{ xa \mid x \in A \}$ (which is fractional since the denominator of a serves as a common denominator), and naturally, it will be denote (a).

Principal fractional ideals (brudne hovedidealer)

Two fractional ideals $\mathfrak{a}$ and $\mathfrak{b}$ can be multiplied; exactly as for ideals one defines

$$\mathfrak{a} \cdot \mathfrak{b} = \{ \sum_{i \in I} a_i b_i \mid I \text{ finite}, a_i \in \mathfrak{a} \text{ and } b_i \in \mathfrak{b} \},$$

which obviously is an A-submodule of K, and it is a fractional ideal since if $x \cdot \mathfrak{a} \subseteq A$ and $y \cdot \mathfrak{b} \subseteq B$, it surely holds that $xy \cdot \mathfrak{ab} \subseteq A$.

*Noetherian rings pop up again; we'll come to them! They have the property that all ideals are finitely generated.

Every finitely generated submodule $\mathfrak{a}$ of K is fractional; the product of the denominators of members of a generating set multiplies $\mathfrak{a}$ into A. Over Noetherian domains* the reciprocal holds true since each fractional ideal is isomorphic to a genuine ideal. However, submodules of K requiring infinitely many generators need not be fractional; an example can be subgroup $\mathbb{Z}[p^{-1}]$ of $\mathbb{Q}$ where one finds elements with any power of p as denominator.

*This transporter has elements from the fraction field K; do not confuse it with the transporter introduced in Paragraph 2.9 which has elements merely from A.

(8.22) The *inverse* of a fractional ideal is the fractional ideal $\mathfrak{a}^{-1} = (A : \mathfrak{a})_K = \{ x \in K \mid x\mathfrak{a} \subseteq A \}$*. Obviously $\mathfrak{a}^{-1}\mathfrak{a} \subseteq A$, and when equality occurs, one says that $\mathfrak{a}$ is *invertible*. The invertible ideals form an abelian group $J(A)$ with $\mathfrak{a} \cdot \mathfrak{b}$ as product, $\mathfrak{a}^{-1}$ as inverse and A as neutral element.

Invertible fractional ideals (invertible brudne idealer)

If $\mathfrak{a}$ is principal, say $\mathfrak{a} = (a/b)$, it holds true that $\mathfrak{a}^{-1} = (b/a)$; indeed, in a domain the equality $x \cdot a/b = d$ is equivalent to $x = d \cdot b/a$; so the name inverse is merited. The principal fractional ideals form a subgroup $P(A)$ of the ideal groupe $J(A)$, and the quotient $Cl(A) = J(A)/P(A)$ is the *ideal class group*. Two invertible ideals $\mathfrak{a}$ and

The ideal class group (idealklassegruppen)

$\mathfrak{b}$ belong to the same class in $Cl(A)$ if and only if there is an element $f \in K$ so that $\mathfrak{a} = f \cdot \mathfrak{b}$; or equivalently, there are non-zero elements $a, b \in A$ such that $b \cdot \mathfrak{a} = a \cdot \mathfrak{b}$.

(8.23) Early in the course (Proposition 4.38 on page 100) we saw that in a domain the free ideals are precisely the principal ones, so just as projective modules form a larger class than free modules, the invertible ideals form a larger class than the principal ones. In view of every (finitely generated) projective modules over a local ring being free, the invertible ideals may be described as the ideals that are locally principal.

PROPOSITION 8.24 (CHARACTERIZATION OF INVERTIBLE IDEALS) *Let A be a domain with fraction field K and let $\mathfrak{a}$ be a fractional ideal over $\mathfrak{a}$. Then the following statements are equivalent:*

i) $\mathfrak{a}$ *is invertible;*
ii) $\mathfrak{a}$ *is finitely generated and projective as an A-module;*
iii) $\mathfrak{a}A_{\mathfrak{m}}$ *is principal for every maximal ideal $\mathfrak{m}$ in A.*

PROOF: *i) ⇒ ii)*: There is by hypothesis a relation $1 = \sum_{1 \leqslant i \leqslant r} a_i b_i$ with the a_i's from $\mathfrak{a}$ and the b_i's from $\mathfrak{a}^{-1}$, and multiplying through by any $x \in \mathfrak{a}$ one finds

$$x = \sum_i (b_i x) a_i, \tag{8.1}$$

where we notice that $b_i x \in A$ since $b_i \in \mathfrak{a}^{-1}$. This shows that the a_i's generate a. To prove that $\mathfrak{a}$ is projective, we let E be a free A-module with a basis $e_1, \ldots, e_r$. The assignments $\alpha(e_i) = a_i$ define an A-linear map $\alpha \colon E \to \mathfrak{a}$, and we contend that α is a split surjection. It is surjective by the observation above that the a_i's generate $\mathfrak{a}$, and in view of (8.1), the map $\sigma \colon \mathfrak{a} \to E$ given by $\sigma(x) = \sum_i (b_i x) e_i$ serves as a right section for α (note that $b_i x \in A$ since $b_i \in \mathfrak{a}^{-1}$ and $x \in \mathfrak{a}$).
ii) ⇒ iii): This is just the facts that every finitely generated projective module over a local ring is free (Proposition 8.1 above), and that an ideal in a domain being free means it is principal.
iii) ⇒ i): When *iii)* holds, the inclusion $\mathfrak{a}^{-1}\mathfrak{a} \subseteq A$ becomes an equality when localized at each maximal ideal because principal fractional ideals are invertible, and we conclude by the local nature of being equal (Corollary 7.49 on page 197). ❑

(8.25) Among invertible ideals we now have two group operations; when considered to be invertible modules, the product is given as $\mathfrak{a} \otimes_A \mathfrak{b}$ and the inverse as $\operatorname{Hom}_A(\mathfrak{a}, A)$ whereas when viewed as invertible ideals, the two operations are $\mathfrak{ab}$ and $\mathfrak{a}^{-1}$. Of course, the two ways coincide, which is the principal step towards showing that the Picard group and the ideal class group are isomorphic. Note that for any two ideals $\mathfrak{a}$ and $\mathfrak{b}$ there is an almost tautological map $\iota \colon (\mathfrak{b} : \mathfrak{a})_K \to \operatorname{Hom}_A(\mathfrak{a}, \mathfrak{b})$ that sends an element $x \in (\mathfrak{b} : \mathfrak{a})_K$ in the transporter to the transporting "multiplication-by-x-map".

LEMMA 8.26 *If $\mathfrak{a}$ and $\mathfrak{b}$ are two invertible ideals in the domain A, the multiplication map $\mu\colon \mathfrak{a}\otimes_A\mathfrak{b} \to \mathfrak{a}\mathfrak{b}$ is an isomorphism. Each A-linear map from $\mathfrak{a}$ to $\mathfrak{b}$ is a homothety; that is, the canonical inclusion ι is an isomorphism $(\mathfrak{b}:\mathfrak{a})_K \simeq \mathrm{Hom}_A(\mathfrak{a},\mathfrak{b})$. In particular, it holds true that $\mathfrak{a}^{-1} = (A:\mathfrak{a})_K \simeq \mathrm{Hom}_A(\mathfrak{a},A)$.*

PROOF: The multiplication map μ is surjective by definition of products of ideals. Now, localizing in the multiplicative set $S = A\backslash\{0\}$, we find $S^{-1}\mathfrak{a} = S^{-1}\mathfrak{b} = K$, the fraction field of A, and hence $S^{-1}(\mathfrak{a}\otimes_A\mathfrak{b}) \simeq S^{-1}\mathfrak{a}\otimes_K S^{-1}\mathfrak{b} = K$. Thus $S^{-1}\mu$ is an isomorphism. It follows that $S^{-1}\ker\mu = 0$, and the kernel $\ker\mu$ is a torsion module. Both $\mathfrak{a}$ and $\mathfrak{b}$ are projective, and by *ii)* of Proposition 8.16 on page 208 their tensor product $\mathfrak{a}\otimes_A\mathfrak{b}$ is also projective. It is therefore a direct summand in a free module and consequently torsion free since by assumption A is an integral domain. We conclude that $\ker\mu = 0$.

Each element x from $(\mathfrak{b}:\mathfrak{a})_K$ gives by multiplication a map $\mathfrak{a}\to\mathfrak{b}$, and the assertion in the second claim is that every map $\mathfrak{a}\to\mathfrak{b}$ is shaped like this. This hinges on the simple fact that two elements a and b from $\mathfrak{a}$ commute so that $x = \phi(a)a^{-1}$ will not dependent on the element $a\in\mathfrak{a}$, and hence ϕ will be the homothety by x. Indeed, because ϕ is A-linear, it follows that

$$a\phi(b) = \phi(ab) = \phi(ba) = b\phi(a).$$

Thus $\phi(a)a^{-1} = \phi(b)b^{-1}$. ❑

(8.27) The invertible fractional ideals are not a sparsely populated outskirt in the land of rank one projectives; on the contrary, each isomorphism class of projective rank one modules contains invertible ideals. Any non-zero element in the dual P^* is a map $P\to A$ whose localization $S^{-1}\phi$ in the multiplicative set $S = A\backslash\{0\}$ is an isomorphism. It ensues that $\ker\phi$ is killed by a non-zero element, and this is only possible if $\ker\phi = 0$ (P is contained in a free A-module and A is a domain). Hence P is isomorphic to its image $\phi(A)$, which is an ideal.

The invertible ideals are invertible modules, and if $\mathfrak{a}$ and $\mathfrak{b}$ belong to the same class in $Cl(A)$, they are isomorphic A-modules; indeed $\mathfrak{a} = f\cdot\mathfrak{b}$ for some $f\in K$. Hence there is a natural map $Cl(A)\to\mathrm{Pic}\,A$ that sends the class of an ideal to its isomorphism class. By what we just did it is surjective, and from Lemma 8.26 one deduces painlessly that it is an injective group homomorphism. Hence:

PROPOSITION 8.28 *If A is a domain, the Picard group $\mathrm{Pic}\,A$ and the ideal class group $Cl(A)$ are isomorphic.*

EXAMPLE 8.4 The Picard group of a PID A vanishes since by definition all ideals are principal, and one may show that $\mathrm{Pic}\,A = 0$ also for Noetherian factorial rings A. ☆

EXERCISE 8.8 Let $\mathfrak{a}$ be an invertible ideal which is generated by two elements. Show that $A\oplus A \simeq \mathfrak{a}\oplus\mathfrak{a}^{-1}$. HINT: Copy the staging in the proof of Proposition 8.24 that *i)* ⇒

ii). ★

EXERCISE 8.9 A sequence x, y of two elements in a domain A is said to be *regular* if $ax = by$ implies that $a = \alpha y$ and $b = \alpha x$ for some $\alpha \in A$. For instance, two elements in a UDF without common factors form a regular sequence. Let $\mathfrak{a} \subseteq A$ be an ideal in a domain that contains a regular sequence of two elements, show that $\mathrm{Hom}_A(\mathfrak{a}, A) = A$. ★

8.4 Examples

Modules over principal ideal domains

Any submodule of a free module over a PID is free, which is a rather rare property for a ring to have—ideals in a domain, for instance, are free if and only they are principal—it holds unconditionally, but we shall prove it merely for modules of finite rank to avoid diving into the deep waters of transfinite induction. A simple proof for the non-finite case may be found in Kaplansky's book ([**?**]) (which of course later was superseded Bass's general result), and for those who would appreciate a transfinite swim, we have included an exercise with hints. It follows that a principal ideal domains enjoy the property that all projective modules are free; among the finitily generated modules even the torsion free ones will be free.

The class of finitely generated modules over a PID is one of the very rare classes of modules which are completely classified up to isomorphism. This includes the classical "Main theorem for finitely generated abelian groups", which states that such a group M, up to isomorphism, decomposes as a direct sum of cyclic groups; that is, one has

$$M \simeq \mathbb{Z}^\nu \oplus \bigoplus_i \mathbb{Z}/p_i^{\nu_i}\mathbb{Z},$$

where ν is non-negative integer (the rank of M), the $p_i^{\nu_i}$'s are prime powers; and of course, the sum is finite. Abelian groups that are not finitely generated can be extremely complicate and are a largely unexplored part of the mathematical world—even the apparently simplest cases; *i. e.* subgroups of $\mathbb{Q} \oplus \mathbb{Q}$, seem to form an impenetrable jungle.

(8.29) We are mainly concerned with rings which are principal ideal domains. However, the case of their big brothers, the *Bézout rings*, are of considerable interest—if for nothing else, functions holomorphic in an open domain form a Bézout ring—and as working with Bézout rings adds no complications, we shall do that. A Bézout ring is a ring all whose finitely generated ideals are principal.

Bézout rings (Bézout ringer)

THEOREM 8.30 *Let A be a Bézout rings; that is, a ring where each finitely generated ideal is principal, and let M be a finitely generated A-module.*

i) If M is torsion free, then M is free;
ii) If M is projective it is free, in particular it holds that $\operatorname{Pic} A = 0$.

Since the property of being torsion free obviously is passed to submodules, one has the corollary that every finitely generated submodule of a free module over a Bézout ring is free; note however, that when A is a PID, the requirement that the submodule be finitely generated is automatically fulfilled.

We we shall need the following general lemma.

LEMMA 8.31 *Assume that M is a finitely generated non-zero torsion free module over a domain A. Then there are non-zero A-linear maps $\phi\colon M \to A$.*

PROOF: As usual K denotes the fraction field of A. The module M being torsion free will be a submodule of the non-zero K-vector space $M\otimes_A K$ which spans $M\otimes_A K$ as a K-vector space, and thus we may find a non-zero K-linear map $\psi\colon M\otimes_A K \to K$. As M spans $M\otimes_A K$, the map ψ does not vanish on M, but of course, it does not necessarily assume values in A. To achieve this let $m_1, \dots, m_r$ be generators for M and let a be a common denominator for the images $\psi(m_i)$. Then $\phi = a \cdot \psi$ does the job. ❑

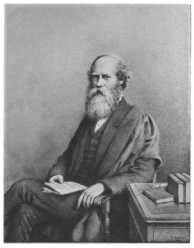
The lemma is not generally valid for modules requiring infinitely many generators; for instance, it holds true that $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Z}) = 0$; indeed, the image of an element from $\mathbb{Q}$ will be an integer divisible by any other integer, and zero is the only such integer.

PROOF OF THEOREM 8.30: We proceed by induction on the rank of M. Any non-zero A-linear map $\phi\colon M \to A$ has an image which is a principal ideal since A is Bézout and M finitely generated, and therefore the image is isomorphic to A. We have thus the split exact sequence:

$$0 \longrightarrow N \longrightarrow M \xrightarrow{\phi} A \longrightarrow 0.$$

The kernel N is torision free and obviously of rank one less than the rank of M. Induction applies, and N is free. But M being isomorphic to $N \oplus A$ is therefore free as well. ❑

(8.32) The classification result of finitely generated modules over a Bézout ring A is established by showing the apparently stronger result that every map between free modules of finite rank over A, or equivalently any matrix with elements from A, can be diagonalized. The arche-typical case of the all-important matrices with integral coefficients, were treated back in 1861 by the Irish mathematician Henry John Stephen Smith. The case of matrices of holomorphic functions was solved by Joseph Wedderburn in 1915. His proof relies only on the base ring being Bézout and is the one we shall present.



Henry John Stephen Smith (1826–1983)
Irish mathematician

Joseph Wedderburn (1882–1948)
Scottish mathematician

THEOREM 8.33 *Let E and F be free modules of finite rank over a Bézout ring A. Any map $\phi\colon F \to E$ can be represented by a diagonal matrix. In other words, if D is a matrix with coefficients in A, there are invertible matrices C and C' with entries in A so that CDC' is diagonal.*

PROOF: We continue with the stage set as in the previous proof. For each surjective map $\pi\colon E \to A$ we may form the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F' & \longrightarrow & F & \xrightarrow{\rho} & A & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \phi|_{F'}} & & \downarrow{\scriptstyle \phi} & & \downarrow & & \\ 0 & \longrightarrow & E' & \longrightarrow & E & \xrightarrow[\pi]{} & A & \longrightarrow & 0 \end{array}$$

where E' and F' denote the kernels of π and $\phi \circ \pi$ respectively, and where the rightmost square is constructed as follows: the image of $\pi \circ \phi$ is a principal ideal; chose a generator g for it and put $\rho = g^{-1} \cdot (\pi \circ \phi)$. The rank of E' is one less than that of E so by induction we may find bases for E' and F' in which $\phi|_{F'}$ is represented as a diagonal matrix.

Chose a basis $\{e_i\}$ for E, and let f be an element of F which is part of a basis and which does not map to zero in E (the zero map is trivial to treat). The expansion of $\phi(f)$ in the basis takes the form $\phi(f) = \sum c_i e_i = d \sum_i b_i e_i$ with d being the greatest common divisor of the c_i's. Then $\sum_i a_i b_i = 1$ for appropriate ring elements a_i. Introduce the projection $\pi\colon E \to A$ by the formula $\pi(\sum x_i e_i) = \sum_i x_i a_i$. If $d = 1$, it holds that $\pi(\phi(f)) = 1$, and $\phi(f)$ (respectively f) forms a basis for E (respectively for F) together with any basis for E' (respectively for F'); thus in that case ϕ is represented by a diagonal matrix.

In case $d \neq 1$, we extend f to a basis $f, f_2, \ldots, f_s$ for F with $f_2, \ldots, f_s$ being one for F'. The trick is to factor ϕ as a product $\phi = \psi \circ \tau$ with both ψ and τ having diagonal matrices. To that end, let $\psi\colon F \to E$ be defined by $\psi(f) = \sum_i b_i e_i$ and $\psi(f_i) = \phi(f_i)$ for $i \geqslant 2$, and $\tau\colon F \to F$ by $\tau(f) = df$ and $\tau(f_i) = f_i$ for $i \geqslant 2$. Obviously τ has a diagonal matrix in any basis for F, and ψ has one by the first part of the proof. ❑

(8.34) As a corollary of the diagonalization theorem one deduces the classification of finitely presented modules over principal ideal domains.

THEOREM 8.35 (MAIN THEOREM FOR MODULES OVER PID'S) *Every finitely generated module M over a principal ideal domain A is isomorphic to a direct sum of cyclic modules. More precisely, it holds true that*

$$M \simeq \nu A \oplus \bigoplus_i A/p_i^{\nu_i} A \tag{8.2}$$

where the p_i's are irreducible elements in A. The integer ν and the integers ν_i's are unambiguously determined by the isomorphism class of M, and the irreducible elements p_i's are unique up to association.

Since A is a PID, the hypothesis that M is finitely presented in Theorem 8.33, may be weakened to M being finitely generated; indeed, over a PID every submodule of a free module of finite rank is of finite rank. This is ensues from the general theory of modules over a Noetherian rings (which we soon shall develop), but an *ad hoc* proof is offered in Exercise 8.10 below.

The first part of the theorem, that M is a direct sum of cyclic modules persists being true for modules of finite presentation over Bézout rings, but the summands are not of the form described in the second statement. The ring Ω of entire functions in the complex plain $\mathbb{C}$ is a Bézout ring, but there are entire functions divisible by infinitely many irreducibles—our old friend $\sin \pi z$ is one example— and if f is one, $\Omega/(f)\Omega$ is cyclic, but not of the prescribed kind.

PROOF: It should be clear that if the map ϕ in the sequence

$$F \xrightarrow{\phi} E \xrightarrow{\pi} M \longrightarrow 0, \tag{8.3}$$

where E and F are finitely generated free modules, has a diagonal matrix in some bases, then its cokernel M is a direct sum of cyclic modules (we leave details to the students).

This shows that M is a direct sum of cyclic modules of the form $A/(f)A$. If $f = pq$ with p and q elements from A without common factors, one may write $1 = ap + bq$, and one easily verifies that ap and bq act as orthogonal idempotents in $A/(f)A$. Consequently $A/(f)A$ decomposes as $A/(f)A \simeq A/(p)A \oplus A/(q)A$. Induction* on the number of irreducible factors finishes the proof.

*This is the only place where we use that A is a PID and not merely a Bézout rings

Finally we attack the uniqueness issue. The number ν equals the rank of M and is of course unambiguously determined. Localizing at a maximal ideal $(p)A$ throws away factors not involving p, so we may assume that A is local with maximal ideal (p) and that the matrix ϕ is diagonal with all entries lying in (p). We contend that two resolutions as in 8.3 are isomorphic in the sense that they enter in a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F_1 & \xrightarrow{\phi_1} & E_1 & \xrightarrow{\pi_1} & M & \longrightarrow & 0 \\ & & \wr\downarrow\beta & & \wr\downarrow\alpha & & \|\downarrow \mathrm{id}_M & & \\ 0 & \longrightarrow & F_2 & \xrightarrow[\phi_2]{} & E_2 & \xrightarrow[\pi_2]{} & M & \longrightarrow & 0. \end{array} \tag{8.4}$$

One then easily finds pairs of bases diagonalizing the two matrices so that the diagonal elements of the two matrices coincide. Once α is in place, β will be the restriction of α to the kernels of the π_i's. To obtain the isomorphism α, note that $\pi_i \otimes \mathrm{id}_k$ are isomorphisms as $\phi_i \otimes \mathrm{id}_k = 0$. By a now standard Nakayama-argument, any map $\alpha\colon E_1 \to E_2$ lifting id_M (that is, rendering the right hand square in (8.4) commutative) will be an isomorphism since $\alpha \otimes \mathrm{id}_K$ equals the isomorphism $(\pi_2 \otimes \mathrm{id}_k)^{-1} \circ (\pi_1 \otimes \mathrm{id}_k)$. ❑

Finitely generated modules over $k[t]$

(8.36) The polynomial ring over a field merits to be mentioned specially:

THEOREM 8.37 *If k is field any finitely generated module M over the polynomial ring $k[t]$ is of the form*

$$M \simeq \nu k[t] \oplus \bigoplus_{i\in I} k[t]/(p_i^{\nu_i})$$

where I is a finite set. Moreover, the non-negative integers ν and ν_i are unambiguously defined by M as is the the sequence $(p_i)_{i\in I}$ of irreducible monic polynomials. In particular, if k is algebraically closed it holds true that

$$M \simeq \nu k[t] \oplus \bigoplus_{i\in I} k[t]/(t-a_i)^{\nu_i}$$

where the a_i's are elements in k.

Exercises

(8.10) Let A be a PID. Show that any submodule of a free module over A of finite rank is finitely generated. HINT: Induction on the rank.

⁂ (8.11) Assume that B is an integral domain which is a finite algebra over a PID A. Show that B is a free A-module of finite rank. If n is the rank, show that every ideal in B can be generated by n elements.

(8.12) *Jordan–Chevalley decompositon.* Let k be an algebraically closed field and V a vector space of finite dimension over k. Show that any $\phi\colon V \to V$ is a sum $\phi = \phi_s + \phi_n$ with ϕ_s and ϕ_n commuting and where ϕ_s is diagonalizable and ϕ_n nilpotent. Show that ϕ_s and ϕ_n are uniquely defined by ϕ. HINT: Consider V to be an $k[t]$-module with t acting via ϕ. On a summand of the type $k[t]/(t-a)^\nu$ put $\phi_s = a\,\mathrm{id}$ and $\phi_n = t - a$.

(8.13) Let $\mathfrak{m}$ be a maximal ideal in the ring A and let ν be a natural number. Show that any projective module over $A/\mathfrak{m}^\nu$ is free. HINT: Nilpotent Nakayama.

⁂ (8.14) *Dedekind-Weber normal form.* The following result is due to Dedekind and Weber: Let k be a field and let $D \in \mathrm{Gl}(n, k[x, x^{-1}])$ be an invertible matrix. Then there are matrices $C \in \mathrm{Gl}(n, k[x])$ and $C' \in \mathrm{Gl}(n, k[x^{-1}])$ so that

$$CDC' = \begin{pmatrix} x^{a_1} & & 0 \\ & \ddots & \\ 0 & & x^{a_r} \end{pmatrix}$$

The restrictions on C and C' are important parts of the theorem. Show the Dedekind–Weber theorem. HINT: The group of units of $k[x, x^{-1}]$ is $\{\, ax^\alpha \mid a \in k^* \text{ and } \alpha \in \mathbb{Z} \,\}$.

(8.15) Show that any finitely generated projective graded module over the polynomial ring $A = k[x_1, \ldots, x_n]$ is free; that is, it is isomorphic to a finite direct sum $\bigoplus_i A(d_i)$. It is

true that every finitely generated projective A-module free, but this is a big theorem. It was conjectured by Jean Pierre Serre and proved independently by Suslin and Quillen. HINT: Graded Nakayama.

(8.16) Assume that A is a domain such that all non-zero ideals are projective. Show that each finitely generated projective module P is of the form $P \simeq \bigoplus_i \mathfrak{a}_i$ where the sum is finite and the $\mathfrak{a}_i$'s are ideals.

(8.17) Assume that A is a PID. An A-module M is a torsion module if any element m is killed by a non-zero ring element a. If $p \in A$ is an irreducible, we let M_p be the set of elements in M killed by some power of p.

a) Show that M_p is a submodule of M, and that $M_p \cap M_q = 0$ if p and q are irreducibles that are not associate.

b) Show that whether M is finitely generated or not, M decomposes at $M = \bigoplus_p M_p$ where the summation extends over a set of representatives for the irreducible elements up to association.

c) Show that the abelian group $\mathbb{Q}/\mathbb{Z}$ is a torsion group whose p-torsion part equals the group $\mathbb{Z}_{p^\infty} = \mathbb{Z}[p^{-1}]/\mathbb{Z}$. Conclude that there is a decomposition $\mathbb{Q}/\mathbb{Z} = \bigoplus_p \mathbb{Z}_{p^\infty}$ where the sum extends over all primes.

(8.18) *For transfinite swimmers.* Submodules of free modules over PID's are free regardless of the free module being finitely generated or not. In this exercise you are guided to give a proof of this, but you are warned that it requires you be initiated in the witchcraft of transfinite induction.

So let E be a free module over the PID A and let $\{e_i\}_{i\in\tau}$ be a basis indexed by a well-ordered set of ordinal type τ. For each ordinal $\sigma < \tau$ we let $E_\sigma = \sum_{i\in\sigma} Ae_\iota$ and put $F_\sigma = F \cap E_\sigma$.

i) If $\sigma < \tau$ show that the quotient $F_{\sigma+1}/F_\sigma$ will be contained in $E_{\sigma+1}/E_\sigma$ and hence it is either zero or isomorphic to A. Conclude that F_σ lies split in $F_{\sigma+1}$. HINT: Prove that $F_\sigma = F_{\sigma+1} \cap E_\sigma$.

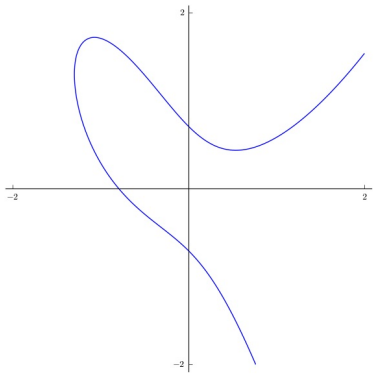
ii) If τ has an immediate predecessor, show that F free.

iii) If τ is a limit ordinal, prove that $F = \bigcup_{\sigma<\tau} F_\sigma$ with the union extending over σ's that are not limit ordinals. Conclude that F is free. HINT: Each F_σ has a basis and lies split in $F_{\sigma+1}$

★

Elliptic curves III

We have already met elliptic curves at several occasions, or to be precise, one should rather say affine elliptic curves on Weierstrass form (there are other standard forms like for instance Tate's normal form; one example: the curve with equation $y^2 + 2xy - x^3 - 1/2 = 0$, whose real points are depicted below). The coordinate ring A of such a curve equals $A = k[x, y]$ with constituting relation $y^2 = p(x)$ where p is a monic polynomial



of degree three with distinct roots, and we must assume that k is an algebraic closed field whose characteristic is not equal to two. This example is about computing the Picard group $\operatorname{Pic} A$, which we shall describe in an *ad hoc* manner. There are general theories and tools in algebraic geometry that make such an exercise easier and place it as a small part in a wider general picture, however, we find an abecedarian approach instructive. It gives use the opportunity to do some amusing and concrete algebra, and one may view it as a motivation for the more advanced xyzetarian technologies.

(8.38) According to the Nullstellensatz in dimension two (Theorem 3.32 on page 78) all maximal ideals in A are of the form $I(P) = (x-a, y-b)$ where $P = (a, b)$ is a point in k^2 lying on the curve C; that is, $b^2 = p(a)$. This allows us to introduce an auxiliary quadratic polynomial $q(x)$ by the relation

$$y^2 - b^2 = p(x) - b^2 = (x-a)q(x), \tag{8.5}$$

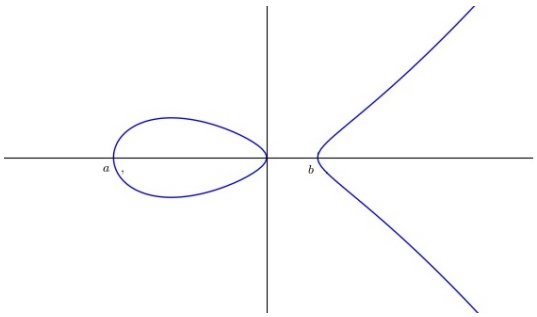
of course it depends on the point, but for simplicity this is not reflected in the notation.

(8.39) Our first *ad hoc* observation is that all ideals in A and consequently all ideals in the local rings $A_{\mathfrak{p}}$, are finitely generated*; actually they are generated by at most two elements. Since A is a free module of rank two over the PID $k[x]$, this follows directly from Exercise 8.11.

*That is, A is Noetherian as we will say later.

LEMMA 8.40 *For each maximal ideal $I(P)$ the local ring $A_{I(P)}$ is a* PID. *In particular, every non-zero ideal in A is invertible. At the point $P = (a, b)$ the maximal ideal $I(P)A_{I(P)}$ is generated by $x - a$ if $b \neq 0$ and by y if $b = 0$.*

There is a simple heuristics behind this lemma. If you take a look at the real curve depicted below, you will se it has vertical tangents near the intersection points with the x-axis, and the projection to the y-axis is locally bijective there. This indicates that one may use the y-coordinate as a parameter in a vicinity of such points. All other points have neighbourhoods where the projection onto the x-axis is one-to-one and where one



thus may use the x-coordinate as a parameter, or $x - a$ if you insist on the parameter being zero at the point under consideration.

PROOF: The proof is divided in two part according to b being zero or not. We first treat the case $b \neq 0$. Then $y + b \notin I(P)$ and is invertible in $A_{I(P)}$, and since $y - b = (y + b)^{-1}(x - a)q(x)$, we see that $I(P)A_{I(P)} = (x - a)$. Assume then that $b = 0$. Differentiating (8.5) shows that $q(a) = p'(a) \neq 0$ (the polynomial $p(x)$ is assumed not to have multiple roots). Therefore $q(x) \notin I(p)$ and is thus invertible in $A_{I(P)}$. Hence $x - a = y^2 q(x)^{-1}$ and $I(P)A_{I(P)} = (y)$. The maximal ideals $I(P)A_{I(P)}$ are therefore all principal, and the claim follows from the next lemma (which is a precursor for Krull's intersection theorem).

Finally, after having established that each local ring $A_{I(P)}$ is a PID, we know that each non-zero ideal in A is locally free of rank one; hence it is projective of rank one and thus invertible by Proposition 8.24 on page 213. ❑

LEMMA 8.41 *Assume that B is a local ring where all ideals are finitely generated and whose maximal ideal $\mathfrak{m}$ is principal. Then B is a* PID.

PROOF: Let t be a generator for $\mathfrak{m}$. We contend that $\mathfrak{a} = \bigcap_i \mathfrak{m}^i = 0$; indeed, assume $\mathfrak{a} \neq 0$ and let $c_1, \ldots, c_r$ be a generator set for it with r minimal. It holds true that $c_1 = tx$ for some $x \in \mathfrak{a}$. One may write $x = a_1 c_1 + \ldots + a_r c_r$ with $a_i \in B$, and thence $c_1 = a_1 t c_1 + \ldots + t a_r c_r$. Now $1 - a_1 t \notin \mathfrak{m}$ and is therefore invertible in B. It ensues that $c_1 = t(1 - a_1 t)^{-1} a_2 c_2 + \ldots + t(1 - a_1)^{-1} a_r c_r$, which is in flagrant contradiction with the c_i's forming a minimal generator set.

This done, let $\mathfrak{a}$ be a non-zero ideal in B. Since $\bigcap_i \mathfrak{m}^i = 0$, there is a largest natural number v so that $\mathfrak{a} \subseteq \mathfrak{m}^v$, and we contend that $\mathfrak{a} = \mathfrak{m}^v = (t^v)$. Indeed, let $x \in \mathfrak{a}$ but $x \notin \mathfrak{m}^{v+1}$. Then $x = at^v$, but $a \notin \mathfrak{m}$. The ring B being local, a is invertible and consequently $\mathfrak{a} = (t^v)$. ❑

PROPOSITION 8.42 *Each non-zero and proper ideal $\mathfrak{a}$ in A is a product of finitely many maximal ideals. In other words, the Picard group* Pic A *is generated by the maximal deals.*

PROOF: Let $f \in \mathfrak{a}$ be non-zero. Then the norm $N(f)$ is a non-zero polynomial which belongs to $\mathfrak{a}$. If the maximal ideal $I(P)$ contains $\mathfrak{a}$, it contains $N(f)$ so that the x-coordinate of P is among the finitely many roots of $N(f)$. We conclude that only finitely many maximal ideals contain $\mathfrak{a}$.

For each point $P \in C$, there is a maximal ν_P so that $\mathfrak{a} \subseteq I(P)^{\nu_P}$ (for most P it will be zero). Localizing in $I(P)$, we find $\mathfrak{a}A_{I(P)} = I(P)^{\nu_P}$ by Lemma 8.41 above, and we may conclude since equality is a local property. ❑

(8.43) Actually, much more is true. Not only is the group $\operatorname{Pic} A$ generated by the maximal ideals, but as we shall see, the association $P \mapsto I(P)$ is a bijection between C and $\operatorname{Pic} A \backslash \{0\}$. This a very specific property of elliptic curves; most other curves only share the property that the Picard group is generated by maximal ideals. The natural question then arises: How is the group law in $\operatorname{Pic} A$ expressed in terms of points on C? Or phrased differently: To which maximal ideal is the product of two maximal ideals isomorphic?

Before answering that question we introduce a natural involution σ on A. It arises from the equation $y^2 = p(x)$ being invariant under the change of the sign of y. We shall use an exponential notation, and denote the action of σ on element $f(x,y)$ from A as $f^\sigma(x,y) = f(x,-y)$, and for each ideal $\mathfrak{a}$ in A the image $\{ f^\sigma \mid f \in \mathfrak{a} \}$ will be denoted by $\mathfrak{a}^\sigma$. The geometric incarnation of σ is just the reflection about the x-axis. Its action on points $P \in C$ will be denoted by $\sigma(P)$, and $\sigma(a,b) = (a,-b)$. It has the three points where C meets the x-axis as fixed points; that is, the three roots of $p(x)$.

We shall answer the rhetoric question above by proving:

THEOREM 8.44 (THE GROUP LAW ON AN ELLIPTIC CURVE) *Associating each point P on the curve C with the maximals ideal in $I(P)$ in A yields a bijection between C and $\operatorname{Pic} A \backslash \{0\}$. The group structure on $C \cup \{0\}$ induced from that on $\operatorname{Pic} A$ has the two properties:*

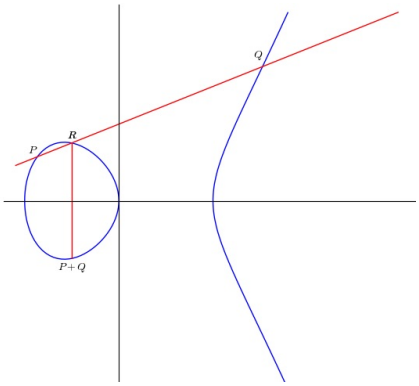
i) $-P = \sigma(P)$;

ii) When $P \neq -P$, it holds true that $P + Q + R = 0$ if and only if P, Q and R are collinear.

Even though the Picard group has a group law induced by the tensor product and so has a multiplicative touch, it is customary to use an additive notation for the induced group structure on C. Notice also that the neutral element 0 does not correspond to a point on C; this reflects the fact that C is an *affine* curve. Adding 0 to C as "the point at infinity" (or closing C up in the projective plane, if you want) yields a so-called complete* curve which is in bijection with the entire $\operatorname{Pic} A$.

The group law is very geometric. To add two points P and Q on C, draw the line through them (which means the tangent to C at P if $Q = P$) and determine the third

*When the ground field is $\mathbb{C}$, one may give C the topology inherited from the standard topology on $\mathbb{C}^2$, and then $C \cup \{0\}$ will be a compactification of C. It is even a Lie group which turns out to be isomorphic to the product $\mathbb{S}^1 \times \mathbb{S}^1$ of two circles



intersection point it has with C; then $P + Q$ will be the reflection of that point through the x-axis.

(8.45) The following lemma reflects the substance of the theorem. We shall explain the lemma, but leave it to the students to carefully carry out the deduction of the theorem.

LEMMA 8.46 *Let P, Q and R be three points on the elliptic curve C. The following three statements hold true:*

i) Each ideal $I(P)$ is projective of rank one and has $I(P)^\sigma$ as inverse;
ii) The ideal $I(P)$ is never principal; i. e. it is never a free module;
iii) If no two of the points are conjugate, the three points are collinear if and only if the ideal $I(P) \cdot I(Q) \cdot I(R)$ is principal;
iv) $I(P) \simeq I(Q)$ if and only if $P = Q$.

PROOF: Proof of *i)*: From Lemma 8.26 above we know that $I(P) \otimes_A I(Q) = I(P) \cdot I(Q)$ for any two points P and Q on C. So our task is to prove that $I(P) \cdot I(P)^\sigma$ is principal. Let the point be $P = (a, b)$. We naively compute the product $I(P) \cdot I(P)^\sigma$:

$$\begin{aligned} I(P) \cdot I(P)^\sigma &= (x-a, y-b)(x-a, y+b) = \\ &= \big((x-a)^2, (x-a)(y-b), (x-a)(y+b), (y^2-b^2)\big) = \\ &= \big((x-a)^2, 2b(x-a), 2y(x-a), (x-a)q(x)\big). \end{aligned}$$

This shows that $I(P) \cdot I(P)^\sigma \subseteq (x-a)$ and to arrive at an equality we have to get rid of either of the factors $2b$, $2y$ or $q(x)$. When $b \neq 0$, we quickly discard the factor $2b$ and are happy (this is where we use that k is of characteristic different from 2).

In case $b = 0$, note that $q(a) \neq 0$, so $q(x)$ and $x - a$ do not have common factors, and we may write $1 = f(x)q(x) + g(x)(x-a)$ with f and g from $k[x]$. This gives the identity $(x-a) = f(x)(x-a)q(x) + (x-a)^2 g(x)$, and we conclude that $(x-a) \in I(P) \cdot I(P)^\sigma$.

Proof of *ii)*: Assume that $I(P)$ is a principal ideal, say generated by f, from which ensues that $(x-a) = I(P)\cdot I^\sigma(P) = (f\cdot f^\sigma)$. Hence $x-a = g\cdot f\cdot f^\sigma$, which is impossible unless f is a unit because $x-a$ is irreducible (Example 3.6 on page 81).

Proof of the "if implication" of *iii)*: The geometric reason behind this is simply, when the three points are aligned, say they are lying on the line $y = \alpha x+\beta$, that the line through them intersects the curve C precisely in the three points (when possible multiplicities are take into account): the x-coordinates of these points are determined from the cubic equation

$$p(x) - (\alpha x+\beta)^2 = 0. \tag{8.6}$$

Denote the three roots by a_1, a_2 and a_2, but be aware that two or all three may coincide. The corresponding y-coordinates will be $\alpha a_i+\beta$. Rebabtising the points as P_1, P_2 and P_3, we contend that

$$(y-(\alpha x+\beta))A = I(P_1)\cdot I(P_2)\cdot I(P_3),$$

which is the proper algebraic way of asserting that the line intersects C exactly in the three points. By Proposition 8.42 above the ideal $(y-(\alpha x+\beta))$ is a product $\mathfrak{m}_1^{\nu_1}\cdot\ldots\cdot\mathfrak{m}_r^{\nu_r}$ of maximal ideals and the exponents ν_i can be found by localization. We examine $(y-(\alpha x+\beta))A_{I(P)}$ for a general point P in C. There are three cases to handle.

i) P is not among the three P_i's. Then $y-(\alpha x+\beta)$ does not vanish at P; it does not belong to $I(P)$, and $(y-(\alpha x+\beta))A_{I(P)} = A_{I(P)}$. The ideal $I(P)$ does not occur as factor in $(y-(\alpha x+\beta))$.

ii) $P = P_i = (a_i, b)$ and $b\neq 0$. Then $y+\alpha x+\beta$ does not vanish in P_i since no two of the P_j's are conjugate. It follows that $(y-(\alpha x+\beta))A_{I(P_i)} = (y^2-(\alpha x+\beta)^2)A_{P_i} = (x-a_i)^{\nu_i}$ where ν_i is the multiplicity of the root a_i.

iii) If $P_i = (a,0)$. Then $\alpha a+\beta = 0$ and $y = y-(\alpha x+\beta)+\alpha(x-a)$. It follows that $((y-(\alpha x+\beta)))A_{I(P)} = (y)$, and the the corresponding multiplicity is one. Differentiating 8.6 one sees that a is simple root of 8.6, since it is a simple zero of $p(x)$.

Proof of *iv)*: Assume that $I(P_1)\simeq I(P_2)$. There are three cases: If the two points are different and non-conjugate points, and consider the line L through P_1 and $\sigma(P_2)$. If $P_2 = \sigma(P_1)$, we note that $b\neq 0$ since the two points are different, and we let L be the tangent to C at P_1. It has the explicit equation $y-y'(a)(x-a)-b$ where $y'(a) = q(a)/2b$. In both cases L intersects C in a third point P_3.

In view of *i)* and the if part of *iii)* , we find

$$A\simeq I(P_3)\otimes_A I(P_1)\otimes_A I(P_2)^\sigma\simeq I(P_3),$$

which is a flagrant contradiction of assertion *ii)*.

Proof of the "only if implication" of *iii)*: Assume finally that $I(P_1) \cdot I(P_2) \cdot I(P_3)$ is principal, and let R be the third intersection point the line through P_1 and P_2 has with C. According to what we just did, $I(P_3)$ and $I(R)$ are then both isomorphic to $I(P_1)^{-1} \otimes_A I(P_2)^{-1}$, and by *iv)* we may conclude that $I(P_3) = I(R)$; that is, $R = P_3$. ❑

The algebraic Möbius band

In Exercise 3.18 on page 83 we examined the ring $A = \mathbb{R}[x, y]$ with constituting relation $x^2 + y^2 = 1$; which is the ring of real polynomial functions on the unite circle, or if one wants, one may interpret A as the ring of trigonometric polynomials by letting $x = \sin\theta$ and $y = \cos\theta$. Exercise 3.18 was about showing that A is not a UFD, and we are about to prove that $\operatorname{Pic} A \simeq \mathbb{Z}/2$. There is in other words up to isomorphism just one rank one projective module which is not free and its square is trivial. In the analogy between projective modules and vector bundles the module P is the algebraic incarnation of the only nontrivial real line bundle on the circle, the Möbius band.

For each point $P = (a, b)$ on the unit circle S, the ideal $I(P) = (x - a, y - b)$ is a maximal ideal in A being the kernel of the evaluation map at P. However, there are other maximal ideals as described in the next lemma, they are however all principal, generated by equations of lines not meeting S.

LEMMA 8.47 *The maximal ideals $\mathfrak{m}$ in A are of the following two types:*

i) Either $\mathfrak{m} = I(P)$ for a point $P = (a, b)$ on the unit circle S,

ii) or $\mathfrak{m}$ is principal and generated by a linear form $\alpha x + \beta y + \rho$ where α, β and ρ are real constants such that $\alpha^2 + \beta^2 = 1$ and $\rho > 1$.

PROOF: Let $\mathfrak{m}$ be a maximal ideal in A, and consider the extension $\mathbb{C}[x, y]$ of A; if $\mathfrak{n}$ is a maximal ideal in $\mathbb{C}[x, y]$ that contains $\mathfrak{m} \cdot \mathbb{C}[x, y]$, it holds that $\mathfrak{n} \cap A = \mathfrak{m}$; likewise, the conjugate ideal $\bar{\mathfrak{n}}$ contains $\mathfrak{m} \cdot \mathbb{C}[x, y]$, and $\bar{\mathfrak{n}} \cap A = \mathfrak{m}$. Now, according to the Nullstellensatz in dimension two (Theorem 3.32 on page 78) it holds true that $\mathfrak{n} = (x - a, y - b)$ with a and b being complex numbers such that $a^2 + b^2 = 1$. If a and b are both real, we are in case *i)*. If not, we evoke Exercise 3.19 on page 83 and conclude that $\mathfrak{n} = (u - c)$ with $u = x + iy$ and $c = a + ib$. Obviously $(u - c) \cdot (\bar{u} - \bar{c}) = 1 + c\bar{c} - u\bar{c} - \bar{u}c$, and writing $u = e^{it}$ and $c = re^{\theta t}$ we the find

$$2^{-1} r^{-1} (1 + c\bar{c} + u\bar{c} - \bar{u}c) = \cos(t - \theta) + (r + r^{-1})/2,$$

which is of the form required in the lemma since $r + r^{-1} > 2$ for all r. But $\mathfrak{n} \cap \bar{\mathfrak{n}} \cap A = \mathfrak{m}$; hence $\mathfrak{m}$ is generated by the real linear form $(u - c) \cdot (\bar{u} - \bar{c})$. ❑

LEMMA 8.48 *Let P and Q be two real points on the unit circle S.*

i) The maximal ideal $I(P)A_{I(P)}$ in the localized ring $A_{I(P)}$ is a principal ideal, but $I(P)$ itself is not principal;

ii) The product $I(P) \cdot I(Q)$ is principal; in particular, $I(P)^2 \simeq A$;
iii) It holds that $I(P) \simeq I(Q)$.

PROOF: Choosing appropriate coordinates, we may assume that $P = (1,0)$, and since $x+1$ does not vanish at P, it is invertible in the local ring $A_{I(P)}$. Hence $x-1 = y^2(x+1)^{-1}$ in $A_{I(P)}$ and $(x-1,y)A_{I(P)} = (y)A_{I(P)}$. However, $I(P)$ is not principal since both y and $x-1$ are irreducible (Exercise 3.18) and neither is a factor of the other.

For the second statement, we choose coordinates so that two points are $(x-a,y-b)$ and $(x+a,y-b)$ (just see to the x-axis being parallel to the line joining P and Q or to the tangent to S at P if $P = Q$). We find by an abecedarian manipulation, stupidly multiplying out and using the identity $x^2+y^2 = a^2+b^2$, that

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

and this ideal is equal to $(y-b)$; indeed, a and b are never simultaneously zero. ❑

We conclude

THEOREM 8.49 *One has* $\operatorname{Pic} A \simeq \mathbb{Z}/2\mathbb{Z}$ *generated by the class of the maximal ideal $I(P)$ for any point P on the unit circle S.*

PROOF: The proof of Proposition 8.42 on page 222 goes word by word through in the present case so that $\operatorname{Pic} A$ is generated by the classes of the maximal ideals. By Lemma 8.48 above it holds that $2[I(P)] = 0$, and thus $[I(Q)] = -[I(P)] = [I(P)]$. ❑

(8.50) There are clear and simple heuristic geometric explanations of these results. Any two points on the circle are connected by a real line intersecting the circle precisely in the two points, and the product of the corresponding maximal ideals is generated by the linear form defining the line. In a similar fashion, the tangent to the circle at a point does not intersect the circle elsewhere, hence the maximal ideals are two-torsion. Lines that do not intersect S intersect the complex curve $x^2+y^2 = 1$ in $\mathbb{C}^2$ in two conjugate points, whose maximal ideal therefore have a product generated by the corresponding linear form, and these are the "other" maximal ideals from case *ii)* in Lemma 8.47.

Exercises

(8.19) Let C be an elliptic curve as in Subsection 8.36 above . Let A be the coordinate ring and K its fraction field. Let furthermore let P and Q be points on C.

a) Show that $I(P) \oplus I(-P) \simeq A \oplus A$;
b) More general show there is an isomorphism $I(P) \oplus I(Q) \simeq A \oplus I(P+Q)$. HINT: There is a natural map $I(P) \oplus I(Q) \to I(P) + I(Q)$; examine its kernel.

(8.20) Let A be a ring. A finite free resolution of an A-module M is an exact sequence

$$0 \longrightarrow F_r \longrightarrow F_{r-1} \longrightarrow \ldots \longrightarrow F_1 \longrightarrow F_0 \longrightarrow M \longrightarrow 0$$

where each F_i is a free module. An A-module M is said to be *stably free* if $M \oplus \nu A$ is free for some non-negative number ν. Show that a stably free projective A-module having a finite free resolution is free.

(8.21) With the notation from the subsection above about elliptic curves (Paragraph 8.45 etc), consider the two matrices

$$\phi = \begin{pmatrix} y-b & -q(x) \\ -(x-a) & y+b \end{pmatrix} \text{ and } \psi = \begin{pmatrix} y+b & q(x) \\ x-a & y-b \end{pmatrix}$$

with coefficients from A. Show that $\phi\psi = \psi\phi = 0$ and that the complex

$$\ldots \xrightarrow{\phi} A \oplus A \xrightarrow{\psi} A \oplus A \xrightarrow{\phi} A \oplus A \xrightarrow{\psi} \ldots$$

which extends infinitely in both directions, is exact. Show that $\operatorname{coker}\phi \simeq I(P)$ and use (part of) the complex above to exhibit an infinite free resolution of $I(P)$. Show that $I(P)$ is not stably free.

★

Lecture 9

Chain conditions

One of the great moments of mathematics was the appearance of Emmy Noether's revolutionary paper *Idealtheorie in Ringbereichen* in 1921 where she introduced the ascending chain condition on ideals and proved the general version of the Primary Decomposition Theorem. The chain conditions have turned out to be extremely useful, and today they permeate both commutative and non-commutative algebra.

Emmy Noether (1882–1935)

German mathematician

9.1 Noetherian modules

We introduced the concept of chains in partially ordered sets already when discussing Zorn's lemma (in Theorem 2.46 on page 49). Recall that a chain $\mathcal{C}$ in a partially ordered set Σ is just a linearly ordered subset; that is, a set such that any two members of the subset $\mathcal{C}$ are comparable.

(9.1) In the present setting, when studying modules over a ring A, we give the term *chain* a more restrictive meaning. The chains we shall consider will all be countable and well ordered. Two sorts of chains will be distinguished, ascending and descending ones. An *ascending chain* in M will be a sequence of submodules $\{M_i\}_{i\in\mathbb{N}_0}$ such that every term M_i is contained in in the successor M_{i+1}; or written out in a display, it is a chain of inclusions like

$$M_0 \subseteq M_1 \subseteq \ldots \subseteq M_i \subseteq M_{i+1} \subseteq \ldots.$$

Ascending chains (oppstigende kjeder)

Descending chains (nedstigende kjeder)

Eventually constant chains (terminerende kjeder)

Noetherian modules (noetherske moduler)

Similarly, a *descending chain* is a sequence $\{M_i\}_{i\in\mathbb{N}_0}$ of submodules fitting into a chain of inclusions shaped like

$$\ldots \subseteq M_{i+1} \subseteq M_i \subseteq \ldots \subseteq M_1 \subseteq M_0.$$

Such chains are said to be *eventually constant* or *eventually terminating* if the submodules become equal from a certain point on; that is, for some index i_0 it holds that $M_i = M_j$ whenever $i, j \geqslant i_0$. Common usage is also to say the chain *stabilizes* at i_0.

(9.2) An A-module M is said to be *Noetherian* if every ascending chain in M is eventually

The Ascending Chain Condition (den oppstigende kjedebetingelsen)
Artinian modules (artinske moduler)

constant. This condition is frequently referred to as the *Ascending Chain Condition* abbreviated to ACC . The module is *Artinian* if every descending chain terminates, a condition also called the *Descending Chain Condition* with the acronym DCC.

A ring A is called *Noetherian* if it is Noetherian as module over itself, and of course, it is *Artinian* if it is Artinian as module over itself. The submodules of A are precisely the ideals, so A being Noetherian amounts to ideals of A satisfying the ACC, and similarly, A is Artinian precisely when the ideals comply with the DCC.

Noetherian and Artinian rings (noetherske og artinske ringer)

(9.3) The two conditions, being Noetherian and Artinian, might look similar, but there is a huge difference between the two. Noetherian and Artinian modules belong in some sense to opposite corners of the category Mod_A. In what follows we shall treat Noetherian modules and Noetherian rings and establish their basic properties, but will lack time to discuss the Artinian modules in any depth, although Artinian rings will be discussed (in Section 9.7 below). In fact, according to a result of Yasuo Akizuki they turn out to be Noetherian as well, they form the class of so-called finite length, and are important both in geometry and number theory.

Emil Artin (1898–1962)
Austrian mathematician

(9.4) The constituting properties of Noetherian modules is asserted in the following theorem. It is due to Emmy Noether and appears as one of the main theorems in her famous paper from 1921.

PROPOSITION 9.5 (MAIN THEOREM FOR NOETHERIAN MODULES) *Let A be a ring and let M be a module over A. The following three conditions are equivalent:*

i) M is Noetherian; that is, it satisfies the ascending chain condition;
ii) Every non-empty family of submodules has a maximal element;
iii) Every submodule of M is finitely generated.

PROOF: Assume first that M is Noetherian and let Σ be a non-empty set of submodules. We must prove that Σ has a maximal element[1]. Assuming the contrary—that there is no maximal elements in Σ—one proves by an easy induction on the length that every finite chain in Σ can be strictly extended upwards. The resulting chain does not terminate, and the ACC is violated.

Next, suppose that every non-empty set of submodules in N possesses maximal elements. Our mission is to prove that every submodule N is finitely generated. To that end, let Σ denote the set of finitely generated submodules. It is clearly non-empty (the zero modules is finitely generated) and consequently has a maximal element N_0. Let $x \in N$ be any element. The module $Ax + N_0$ is finitely generated and contains N_0,

[1] Even though resembling Zorn's lemma this is quite different. The ACC assumption is stronger than what Zorn's lemma asks in that chains are required to be eventually constant not only bounded above; on the other hand the ACC places restrictions only on countable and well ordered chains. Anyhow, it is of interest that the proposition is independent of the Axiom of Choice

so from the maximality of N_0 it ensues that $x \in N_0$. Hence $N = N_0$, and N is finitely generated.

For the third and last implication, assume that all submodules of N are finitely generated, and let an ascending chain

$$M_0 \subseteq M_1 \subseteq \ldots . \subseteq M_i \subseteq M_{i+1} \subseteq \ldots$$

be given. The union $N = \bigcup_i M_i$ is by assumption finitely generated and have say $x_1, \ldots, x_r$ as generators. Each x_j lies in some M_{ν_j}, and the chain being ascending, they all lie in M_ν with $\nu = \max_j \nu_j$. Therefore $N = M_\nu$, and the chain stabilizes at ν. ❑

There are statements for Artinian modules that correspond to the two first claims in the theorem, which are of a kind of order-theoretical nature (you are asked to give a proof in Exercise 9.5 below). However, there is no substitute for the third, about submodules being finitely generated, which draws module theory into the business.

(9.6) The Noetherian modules, as do the Artinian modules, form a subcategory of Mod_A which enjoys a strong closure property. They are what in category theory are called *thick subcategories*. Submodules and quotients of Noetherian modules are Noetherian as is an extension of two, and for of Artinian modules the same holds true.

Thick subcategories (tykke underkategorier)

PROPOSITION 9.7 *Let M', M and M'' be three A-modules fitting in a short exact sequence*

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0.$$

Then the middle module M is Noetherian (respectively Artinian) if and only if the two extremal modules M' and M'' are.

In particular—as follows by a straightforward induction—finite direct sums of Noetherian (or Artinian) modules will be Noetherian (respectively Artinian), and *vice versa*: If a direct sum is Noetherian (or Artinian) it is finite and all the summands are Noetherian (or Artinian).

PROOF: We may without loss of generality identify M' with its image in M. Every chain in M' is then a chain in M, so if M is Noetherian (or Artinian), the same is true for M'. In the same vein, if $\beta\colon M \to M''$ denotes the quotient map, a chain $\{N_i\}$ in M'' lifts to the chain $\{\beta^{-1}(N_i)\}$ in M. Since β is surjective, it holds that $\beta\big(\beta^{-1}(N_i)\big) = N_i$, so that $\{N_i\}$ stabilizes whenever $\{\beta^{-1}(N_i)\}$ does. Hence, if M is Noetherian (or Artinian), so will M''.

To prove the remaining half of the proposition assume that the two extreme modules M' and M'' are Noetherian (or Artinian) and let $\{N_i\}$ be a chain in M. The chain $\{N_i \cap M'\}$ stabilizes at some ν, hence $N_i \cap M' = N_j \cap M'$ for $i, j \geqslant \nu$.

Mapping the N_i's into M'' one obtains the chain $\{\beta(N_i)\}$ in M'', and since M'' by assumption is Noetherian (or Artinian), it stabilizes at some μ. Hence $\beta(N_i) = \beta(N_j)$

for $i, j \geqslant \mu$. For $i, j \geqslant \max(\mu, \nu)$ this gives

$$N_i/N_i \cap M' = \beta(N_i) = \beta(N_j) = N_j/N_j \cap M' = N_j/N_i \cap M',$$

and hence $N_i = N_j$. ❏

(9.8) The properties of being Noetherian or Artinian are retained when a module is localized.

PROPOSITION 9.9 *Let S be a multiplicative set in the ring A and let M be an A-module. If M is Noetherian (respectively Artinian the localized module M_S is Noetherian (respectively Artinian).*

PROOF: The proof is based on the simple observation that for any submodule N of $S^{-1}M$ one has $S^{-1}(\iota^{-1}N) = N$, where $\iota\colon M \to S^{-1}M$ denotes the localization map: indeed, if an element $\iota(y)s^{-1}$ in $S^{-1}M$ belongs to N, so does $\iota(y)$.

Now, any chain $\{N_i\}$ in $S^{-1}M$, whether ascending or descending, induces a chain $\{\iota^{-1}(N_i)\}$ in M, and if this chain stabilizes, say $\iota^{-1}(N_i) = \iota^{-1}(N_j)$ for $i, j \geqslant i_0$, it holds true that $N_i = S^{-1}(\iota^{-1}N_i) = S^{-1}(\iota^{-1}N_j) = N_j$, and the original chain stabilizes at i_0 as well. ❏

Examples

(9.1) *Vector spaces:* A vector space V over a field k is Noetherian (or Artinian) if and only if it is of finite dimension. Indeed, if V is of finite dimension it is the direct sum of finitely many copies of k, hence Noetherian (and Artinian).

If V is not of finite dimension there will be infinite sets $v_1, \ldots, v_i, \ldots$ of linearly independent vectors, and for such the subspaces $V_i =< v_1, \ldots, v_i >$ will form a strictly ascending chain of subspaces; hence V is not Noetherian. A similar argument shows that neither is V Artinian: the spaces $W_i =< v_i, v_{i+1} \cdots >$ form a strictly decreasing chain of subspaces.

(9.2) *Finite product of fields:* The conclusions of the preceding example extend to rings that are finite products of fields; say $A = \prod_{1 \leqslant i \leqslant r} k_i$. Modules over such rings are direct sums $V = \bigoplus_{1 \leqslant i \leqslant r} V_i$ where each V_i is a vector space over k_i with the A-module structure induced by the projection $A \to k_i$. From Proposition 9.7, or rather the succeeding comment, ensues that V is Noetherian (or Artinian) if and only if each V_i is of finite dimension over k_i.

☆

Exercises

⁂ (9.1) Prove the assertion just after Proposition 9.7 that if a direct sum of Noetherian modules is Noetherian, the sum is finite and all the summands are Noetherian.

⁂ (9.2) Show that $\mathbb{Z}$ is a Noetherian $\mathbb{Z}$-module, but that $\mathbb{Z}_{p^\infty} = \mathbb{Z}[p^{-1}]/\mathbb{Z}$ is not. Show

that $\mathbb{Z}_{p^\infty}$ is an Artinian $\mathbb{Z}$-module, but that $\mathbb{Z}$ is not. HINT: The only submodules of $\mathbb{Z}_{p^\infty}$ are the cyclic ones generated by the images of p^{-i} for different i's.

⁂ (9.3) Let $\phi\colon A \to B$ be a map of rings and let M be a B-module. Prove that if M is Noetherian as an A-module, it is Noetherian as an B-module as well. Show by exhibiting examples that the converse is not true in general, but holds true when ϕ is surjective.

⁂ (9.4) Show that a direct sum of finitely many simple modules is both Noetherian and Artinian.

⁂ (9.5) Let M be an A-module. Show that the following two claims are equivalent:

i) M is an Artinian module;

ii) Every non-empty family Σ of submodules of M has a minimal element.

★

9.2 *Noetherian rings*

Recall that a ring A is called *Noetherian* if it is Noetherian as a module over itself. The Noetherian rings form a large natural class of rings with a very rich theory. The lion's share of the rings appearing in classical algebraic geometry are of so-called *essential finite type* over a field k (or over a Noetherian ground ring); that is, they are localizations of finitely generated k-algebras (or algebras over the ground ring). All these rings are Noetherian. Hilbert's basis theorem ensures that algebras finitely generated over a Noetherian base are Noetherian, and by Proposition 9.9 above localizing a ring preserves the property of being Noetherian.

Essential finite type (essensielt av endelig type)

Be aware that although having lots of nice properties, Noetherian rings can be treacherous and show an unexpectedly bad behaviour. Even among local Noetherian rings, which usually are rather tame and well-behaved animals, one finds example with strange properties.

(9.10) The ring A being Noetherian means that any ascending chain of ideals eventually terminates. Applying Proposition 9.5 on page 230 to the ring A itself while remembering that the submodules of A are precisely the ideals, we arrive at the following:

PROPOSITION 9.11 (THE MAIN THEOREM FOR NOETHERIAN RINGS) *For a ring A the following three conditions are equivalent:*

i) *A is Noetherian; that is, the ideals in A comply with the ascending chain condition;*

ii) *Every non-empty family of ideals in A has a maximal element;*

iii) *Every ideal in A is finitely generated.*

It is trivial that fields are Noetherian, and shortly we shall see that polynomial rings over fields are Noetherian too; this is a special case of the celebrated Hilbert's Basis

Theorem. Other examples of Noetherian rings are the principal ideal domains, where ideals are not only finitely generated, but generated by a single element.

(9.12) Quotients of Noetherian rings are Noetherian (Proposition 9.11 on the previous page), but not necessarily subrings. Non-Noetherian domain are obvious examples: they are contained in their fraction fields, which are Noetherian. A subtler example will be given below (Example 9.4).

PROPOSITION 9.13 *Let A be a Noetherian ring and M an A-module. Then M is Noetherian if and only if M is finitely generated.*

PROOF: A finitely generated A-module M can be realized as the quotient of a finite direct sum nA of n copies of A. When A is Noetherian, it follows from Proposition 9.7 on page 231 that nA is Noetherian; indeed, one obtains nA by successive extensions of A by itself. By Proposition 9.7 again, all quotients of nA, in particular M, will be Noetherian. Finally, Noetherian modules are finitely generated since all their submodules are. ❏

(9.14) A converse to Proposition 9.13 does not hold in the sense that rings may have non-zero Noetherian modules without being Noetherian themselves; in fact, this applies to all non-Noetherian rings: simple modules are Noetherian (all submodules are finitely generated!), and every ring possesses non-trivial simple modules by The Fundamental Existence Theorem for Ideals (Theorem 2.49 on page 49). These examples are in some way illustrative; any Noetherian module over a non-Noetherian ring must have a non-trivial annihilator ideal; or phrased in another way, if A has a Noetherian module with global support— what is also called a *faithful* module—it is a Noetherian ring.

Faithful modules (trofaste moduler)

PROPOSITION 9.15 *Assume that M is a module over A. If M is Noetherian, then $A/\operatorname{Ann} M$ is Noetherian as well.*

PROOF: Let $x_1, \dots, x_r$ be generators for M, and consider the map $\phi\colon A \to rM$ that sends x to the tuple $(x \cdot x_1, \dots, x \cdot x_r)$. If x kills all the x_i's, it kills the entire module M, since the x_i's form a generating set, and we may infer that the kernel of ϕ equals the annihilator $\operatorname{Ann} M$. This means that $A/\operatorname{Ann} M$ is isomorphic to a submodule of rM, hence it is Noetherian by Proposition 9.7 above. ❏

Minimal primes in Noetherian rings

(9.16) The minimal prime ideals of an ideal $\mathfrak{a}$ in a ring A (that is, the prime ideals minimal among those containing $\mathfrak{a}$) are in the front line when one starts examining $\mathfrak{a}$. Geometrically, their corresponding closed subsets are the irreducible components of of Spec A. And an important feature—in fact a basic finiteness property—of Noetherian rings is that the set of minimal primes of any ideal is finite. For this reason it is appropriate and natural to include a proof at this stage, which also has the bonus of furnishing a nice and simple illustration of an ever recurring technique called *Noetherian*

Noetherian induction (noethersk induksjon)

induction: If a statement about ideals is not true for all ideals, the set ideals for which it fails—the gang of bad guys, so to say—is then non-empty and consequently will have a maximal member, and working with this maximal scoundrel, one tries to establish a contradiction.

Remembering that the radical of an ideal is the intersection of its minimal primes, one obtains—as a prelude to the general theory of primary decompositions—the corollary that radical ideals in Noetherian rings equal the intersections of their finitely many minimal prime ideals; that is, one has

$$\sqrt{\mathfrak{a}} = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_r \tag{9.1}$$

with the $\mathfrak{p}_i$'s being the minimal primes of $\mathfrak{a}$. And, of course, the minimal primes are unambiguously determined by the ideal $\mathfrak{a}$, and moreover, no inclusion relation among them persists.

PROPOSITION 9.17 *Each ideal $\mathfrak{a}$ in a Noetherian ring has only finitely many minimal prime ideals.*

PROOF: Let Σ be the set of proper ideals in A having infinitely many minimal prime ideals. If Σ is non-empty, it has maximal member, say $\mathfrak{a}$. Obviously $\mathfrak{a}$ is not a prime ideal, so there are elements x and y neither lying in $\mathfrak{a}$, but whose product xy belongs to $\mathfrak{a}$. Then $\mathfrak{a} + (x)$ and $\mathfrak{a} + (y)$ are proper ideals (their product is contained in $\mathfrak{a}$) strictly larger $\mathfrak{a}$, and consequently each has merely finitely many minimal primes. Any prime ideal containing $\mathfrak{a}$ contains either x or y, hence each minimal prime ideal of $\mathfrak{a}$ is either among the finitely many minimal primes of $\mathfrak{a} + (x)$ or the finitely many of $\mathfrak{a} + (y)$. ❑

COROLLARY 9.18 *A radical ideal in a Noetherian ring is an irredundant intersection of finitely many prime ideals. In particular, the nil-radical is the intersection of finitely many prime ideals. The involved prime ideals are unique.*

Examples

Examples of Noetherian rings will soon abound, so here we merely give a few examples of non-Noetherian rings, noting the words of wisdom of Sun Tzu: "If you know the enemy and know yourself".

(9.3) The obvious example of a non-Noetherian ring is the ring $A[x_1, x_2, \ldots]$ of polynomials in infinitely many variables over any ring A. Clearly, the chain of ideals

$$(x_1) \subset (x_1, x_2) \subset \ldots \subset (x_1, x_2, \ldots, x_i) \subset \ldots$$

does not stabilize.

(9.4) One might be mislead by the previous example to believe that non-Noetherian rings are monstrously big. There are, however, non-Noetherian rings contained in the

polynomial ring $\mathbb{Q}[x]$. The simplest example is even a subring of the ring $\mathbb{Z}[p^{-1}][x]$ where p is a natural number greater than one. It is formed by those polynomials in $\mathbb{Z}[p^{-1}][x]$ that assume an integral value at zero; that is, the polynomials $P(x)$ such that $P(0) \in \mathbb{Z}$. In this ring A one finds the following ascending chain of principal ideals

$$(p^{-1}x) \subset (p^{-2}x) \subset \ldots \subset (p^{-i}x) \subset \ldots,$$

which does not stabilize. Indeed, if $p^{-(i+1)}x \in (p^{-i}x)$, one would have $p^{-(i+1)}x = P(x)p^{-i}x$ for some polynomial $P(x) \in A$. Cancelling $p^{-i}x$ would give $p^{-1} = P(x)$, which contradicts that $P(0) \in \mathbb{Z}$.

(9.5) A large class of important non-Noetherian rings are formed by the rings $H(\Omega)$ of holomorphic functions in an open domain Ω in the complex plane. Chains that do not terminate arise from sequences of distinct points in Ω that do not accumulate in Ω. If $\{z_i\}$ is such a sequence, let $\mathfrak{a}_n$ be the ideal of functions in $H(\Omega)$ vanishing in the set $Z_n = \{z_{n+1}, z_{n+2}, \ldots\}$. These ideals clearly form an ascending chain, and from Weierstrass' Existence Theorem ensues that there are functions f_n holomorphic in Ω whose zeros are exactly the points in Z_n. Then $f_n \in \mathfrak{a}_n$, but $f_n \notin \mathfrak{a}_{n-1}$, and the chain can not stabilize at any stage.

☆

Exercises

✲ (9.6) Let $A \subseteq \mathbb{Q}$ be any proper subring. Show that the polynomials in $\mathbb{Q}[t]$ assuming values in A at the origin, is not Noetherian.

(9.7) Let $\{A_i\}_{i \in I}$ be a family of Noetherian rings all different from the null ring.

a) Show that the product $\prod_i A_i$ is Noetherian when I is finite.

b) Show that the product $\prod_i A_i$ is not Noetherian when I is infinite.

(9.8) *The ring of numerical polynomials.* A polynomial $p(x)$ in $\mathbb{Q}[x]$ is called a *numerical* or *integral* polynomial if it assumes integral values on the integers. Every such polynomial has a unique expansion

$$p(x) = \sum c_\nu \binom{x}{\nu}$$

where $\binom{x}{\nu} = x(x-1)\ldots(x-\nu)/\nu!$ and where the c_ν's are integers.

Show that the ring $\mathrm{Int}(\mathbb{Z})$ of numerical polynomials in $\mathbb{Q}[x]$ is not Noetherian. HINT: Show for instance that the ideal $\mathfrak{m} = \{ f \in \mathrm{Int}(\mathbb{Z}) \mid f(0) \text{ is even } \}$ is not finitely generated.

(9.9) *The ring of algebraic integers.* Let $\mathbb{A}$ be the subring of the complex numbers $\mathbb{C}$ whose elements are algebraic integers; that is, they are solutions of equations of the type

$$z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0 = 0$$

where the coefficients a_i are integers. Show that $\mathbb{A}$ is not Noetherian. HINT: For instance, the principal ideals $(\sqrt[2n]{2})$ form an ascending sequence that does not terminate. ★

9.3 *A structure theorem for modules*

As an illustration of the strength and elegance the Noetherian method can show, we offer a structure theorem for finitely generated modules over Noetherian rings—it does not reveal the finer features of a module, but rather describes the overall structure. Every such module is obtained by a series of successive extensions of cyclic modules shaped like $A/\mathfrak{p}$ with the $\mathfrak{p}$'s being prime ideals. It has the important consequence that all additive invariants[2] of Mod_A are determined by their values on the quotients $A/\mathfrak{p}$ with $\mathfrak{p}$ prime.

(9.19) The proof of the structure theorem is built on the following result which is of independent (and fundamental) importance and will be use later.

PROPOSITION 9.20 (MAXIMAL ANNIHILATORS) *Assume that A is a ring and M an A-module. Let $\operatorname{Ann} x$ be maximal among the annihilators of non-zero elements in M. Then $\operatorname{Ann} x$ is a prime ideal.*

Associated prime ideals (assosierte primidealer)

The prime ideals that are annihilators of an element of M are said to be *associated* to M, and the set they form is denoted $\operatorname{Ass} M$.

PROOF: To begin with, observe that $\operatorname{Ann} x$ is a proper ideal as x is non-zero. Let then a and b be ring elements such that $ab \in \operatorname{Ann} x$ and assume that $a \notin \operatorname{Ann} x$. Then $ax \neq 0$. It is generally true that $\operatorname{Ann} x \subseteq \operatorname{Ann} ax$, but since $ax \neq 0$ it holds that $\operatorname{Ann} x = \operatorname{Ann} ax$ because $\operatorname{Ann} x$ is maximal among annihilators of non-zero elements. Now, $bax = 0$, so $b \in \operatorname{Ann} ax = \operatorname{Ann} x$. ❑

COROLLARY 9.21 *Any non-zero module over a Noetherian ring contains a module isomorphic to $A/\mathfrak{p}$ for some prime ideal $\mathfrak{p}$.*

PROOF: Observe that the set of annihilators of non-zero elements is non-empty and has a maximal element since A is Noetherian. Then cite Proposition 9.20 above. ❑

EXERCISE 9.10 Show the slight extension of Proposition 9.20 that if $\operatorname{Ann} x$ is maximal among annihilators of non-zero elements from M that are contained in a fixed prime ideal $\mathfrak{p}$ of A, then $\operatorname{Ann} x$ is prime. ★

(9.22) The ground is now well prepared for the promised structure theorem; here it comes:

[2] An additive invariant is a map $\chi\colon \mathsf{Mod}_A \to G$ where G is a commutative monoid, such that $\chi(M) = \chi(M') + \chi(M'')$ each time

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

is an exact sequence.

THEOREM 9.23 (STRUCTURE OF MODULES) *Let A a Noetherian ring and let M be a non-zero A module. Then M is finitely generated if and only if it possesses a finite ascending chain of submodules $\{M_i\}_{0\leqslant i\leqslant r}$ with $M_0 = 0$ and $M_r = M$ whose subquotients are shaped like cyclic modules $A/\mathfrak{p}_i$ with the $\mathfrak{p}_i$'s being prime; that is, there are short exact sequences*

$$0 \longrightarrow M_{i-1} \longrightarrow M_i \longrightarrow A/\mathfrak{p}_i \longrightarrow 0$$

for $1 \leqslant i \leqslant r$.

PROOF: Let M be finitely generated A module. The set of submodules of M for which the theorem is true is non-empty by Corollary 9.21 and has thus a maximal element, say N. If N were a proper submodule, the quotient M/N would be non-zero and hence contain a submodule isomorphic to $A/\mathfrak{p}$ for some prime $\mathfrak{p}$. The inverse image N' of $A/\mathfrak{p}$ in M would be a submodule containing N and satisfying $N'/N \simeq A/\mathfrak{p}$, so the theorem would also hold for N' violating the maximality of N. ❏

Associated prime ideals and the support

In Section 7.5 of Chapter 7 we introduced the notion of support Supp M of an A-module M as the subset of Spec A formed by the prime ideals $\mathfrak{p}$ such that $M_\mathfrak{p}$ is non-zero. Over Noetherian rings there is an intimate relation between the support of a module and the set Ass M of associated primes. The latter is always a subset of the former, and they have the same minimal elements, so if Supp M is closed, for instance, if M is finitely generated, it equals the closure of the set Ass M.

(9.24) The result follows here; note that we do not require M to be finitely generated, but A needs to be Noetherian.

PROPOSITION 9.25 *Let A be a Noetherian ring and M an A-module.*

i) Then Ass M *is non-empty and* Ass $M \subseteq$ Supp M.

ii) The sets Ass M *and* Supp M *have the same minimal elements.*

PROOF: That Ass M is non-empty is just a restatement of Corollary 9.21 (the combination of maximal annihilators being prime and that A being Noetherian ensures that they exists). If $\mathfrak{p} = \operatorname{Ann} x$ is a prime annihilator, the element x survives in $M_\mathfrak{p}$ according to Lemma 7.46 on page 196, and so $\mathfrak{p}$ belongs to Supp M.

We proceed proving the second statement and let $\mathfrak{p}$ be minimal in the support Supp M. Consider the $A_\mathfrak{p}$-module $M_\mathfrak{p}$. It is a non-zero module whose support is reduced to the maximal ideal $\mathfrak{p}A_\mathfrak{p}$ since $\mathfrak{p}$ is minimal in Supp M, so in view of statement *i)* the maximal ideal $\mathfrak{p}A_\mathfrak{p}$ must be associated to $M_\mathfrak{p}$. The maximal ideal is therefore an annihilator, say $\operatorname{Ann}_{A_\mathfrak{p}} x$, and moreover, the element x may be chosen to lie in M. It holds that $\mathfrak{p} = \mathfrak{p}A_\mathfrak{p} \cap A = \operatorname{Ann}_{A_\mathfrak{p}} x \cap A$. We contend that there is an $s \notin \mathfrak{p}$ so that $\mathfrak{p} = \operatorname{Ann}_A sx$. Indeed, let $a_1, \ldots, a_r$ be generators for $\mathfrak{p}$; for each there is an s_i with

$s_i a_i x = 0$. If $s = s_1 \cdot \ldots \cdot s_r$, it holds that $\mathfrak{p} \subseteq \operatorname{Ann} sx$, and consequently $\mathfrak{p} = \operatorname{Ann} sx$ because evidently $\operatorname{Ann} sx \subseteq \mathfrak{p}$. We conclude $\mathfrak{p}$ is a prime annihilator, hence it belongs to $\operatorname{Ass} M$.

For the reverse inclusion, pick a minimal element $\mathfrak{p}$ from $\operatorname{Ass} M$ and let $\mathfrak{q} \subset \mathfrak{p}$ be a prime ideal. If $M_{\mathfrak{q}} \neq 0$, there must be an element $y \in M$ so that $\operatorname{Ann} y \subseteq \mathfrak{q}$, but as A is Noetherian, there is a maximal such annihilator ideal, which according to the Principle of Maximal Annihilators (9.20 on page 237), or rather the extended version in Exercise 9.10, is prime; and this contradicts that $\mathfrak{p}$ is minimal in $\operatorname{Ass} M$. ❑

(9.26) The union and the intersection of the associated prime ideals have a special significance for the module M, quite parellel to what is the case for ideals. One has:

LEMMA 9.27 *Let A be a Noetherian ring and M a finitely generated A-module. Then*

i) $\bigcap_{\mathfrak{p} \in ass M} \mathfrak{p} = \sqrt{\operatorname{Ann} M}$;

ii) $\bigcup_{\mathfrak{p} \in \operatorname{Ass} M} \mathfrak{p}$ *is the set of zero-divisors in* M.

PROOF: The first follows since $\operatorname{Supp} M$ and $\operatorname{Ass} M$ have the same minimal prime ideals, and since $\bigcup_{\mathfrak{p} \in \operatorname{Supp} M} \mathfrak{p} = \sqrt{\operatorname{Ann} M}$ when M is finitely generated (Proposition 7.55 on page 199). The second follows by the observation that a zero-divisor x lies in an annihilator, hence in a maximal annihilator, which is prime and belongs to $\operatorname{Ass} M$. ❑

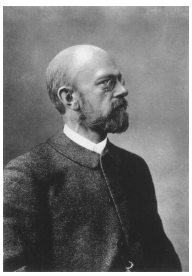
Exercises

(9.11) Let $R = k[t, x_1, x_2, \ldots]$ with constituting relations $x_i = t x_{i+1}$ for $i \geqslant 1$ and let $A = R_{\mathfrak{m}}$ where $\mathfrak{m}$ is the maximal ideal $\mathfrak{m} = (t)$. Consider the module $M = A/(x_1)A$. Show that $\operatorname{Ass} M = \varnothing$, but that M is of global support; *i. e.* $\operatorname{Supp} M = \operatorname{Spec} A$. HINT: A has two non-zero prime ideals $\mathfrak{m}A = (t)A$ and $\mathfrak{p} = \bigcap_{i \geqslant 1} \mathfrak{m}^i$.

★

9.4 *Hilbert's Basis Theorem*

There is almost an infinity of strong results about Noetherian rings, unfortunately we have time to treat too few of them. As a beginning, in this section we shall discuss two. In addition to Hilbert's basis theorem, we treat a criterion for rings being Noetherian due to I.S. Cohen. and in the next section we follow up and give one of Wolfgang Krull's many important results, his intersection theorem.



David Hilbert (1862–1943)

German mathematician

Hilbert's Basis Theorem

As one might think the name indicates, Hilbert's Basis Theorem lies at the basis for the theory of commutative rings, and thereby is paramount for the development of algebraic geometry. It guarantees that most rings appearing in those parts of mathematics are Noetherian. However the name originate from the content of the theorem, that any

ideal in a polynomial ring over a field has a finite basis—the modern version is that polynomial rings over Noetherian rings are Noetherian.

Hilbert proved this theorem as early as in 1890. The proof was published in the paper *Über die Theorie der algebraischen Formen*. Naturally the formulation was slightly different from the modern one (the term Noetherian was of course not in use; Emmy Noether was only eight years old at the time), and the context was confined to polynomial rings over fields or specific rings like the integers, but the spirit was entirely the same. The abstract and non-constructive proof was revolting at a time when that part of mathematics was ruled by long and soporific computations, making it extremely difficult to obtain general results, and it opened up the path to modern algebra. Part of the mythology surrounding the theorem is the exclamation by the "König der Invariant Theorie" Paul Gordan: "Das ist nicht Mathematik, das ist Theologie!". The truth is that Hilbert had proved in a few pages what Gordan and his school had not proved in twenty years. (9.28) There are several different proofs in circulation, and we shall give one of the shortest. These days many constructive proofs are known and good algorithms exist for exhibiting explicit generators for ideals in polynomial rings; however, we shall present a non-constructive proof in the spirit of Hilbert's.

Paul Albert Gordan (1837–1912)

German mathematician

THEOREM 9.29 (HILBERT'S BASIS THEOREM) *If A is a Noetherian ring, then so i the polynomial ring $A[x]$.*

Before giving the proof of Hilbert's basis theorem we state three important corollaries. A straightforward induction on the number variables immediately yields the following:

COROLLARY 9.30 *The polynomial ring $A[x_1, \ldots, x_n]$ over a Noetherian ring A is Noetherian.*

An important special case is when the ground ring A is a field. Since fields are Noetherian, the Basis Theorem tells us that polynomial rings over fields are Noetherian. Moreover, quotients of Noetherian rings are Noetherian, and we obtain directly the next corollary. In particular it says that algebras of finite type over a field, a class of rings that include the coordinate rings of affine varieties, are Noetherian.

COROLLARY 9.31 *Any algebra finitely generated over a Noetherian ring is Noetherian.*

Finally, the last corollary we offer before giving the proof of Hilbert's Basis Theorem, combines that theorem with Proposition 9.9 on page 232 which states that localization preserves Noetherianess. Recall that and A-algebra is said to be *essentially of finite type* if it is the localization of a finitely generated A-algebra.

COROLLARY 9.32 *Any ring essential of finite type over a Noetherian ring is Noetherian.*

PROOF OF HILBERT'S BASIS THEOREM: Let $\mathfrak{a}$ be an ideal in $A[x]$ and for each n let $\mathfrak{a}_n$ be the set of leading coefficients of elements from $\mathfrak{a}$ of degree at most n. Each $\mathfrak{a}_n$ is an ideal

in A, and they form an ascending chain, which since A is assumed to be Noetherian, eventually stabilizes, say for $n = N$. Each $\mathfrak{a}_n$ is finitely generated, so for each $n \leqslant N$ we may chose a finite set of polynomials of degree at most n whose leading coefficients generate $\mathfrak{a}_n$. Let $f_1, \ldots, f_r$ be these polynomials in some order and let $a_1, \ldots, a_r$ be their leading coefficients.

We contend that the f_i's generate $\mathfrak{a}$. So assume not, and let f be of minimal degree ν among the bad guys; that is, among those polynomials in $\mathfrak{a}$ that do not belong to the ideal generated by the f_i's. If the leading coefficient of f is a, it holds that $a \in \mathfrak{a}_\nu$ and we may write $a = \sum_j b_j a_{i_j}$ with the polynomials f_{i_j} whose leading coefficient is a_{i_j}, all are of degree at most the degree of f. The numbers $d_j = (\deg f - \deg f_{i_j})$ are all non-negative, and we may thus form the polynomial $f - \sum_j b_j x^{d_j} f_{i_j}$. It is of degree less than $\deg f$: the term of degree $\deg f$ vanishes by the very choice of the b_j. It does not lie in the ideal generated by the f_i's since f does not, and that contradicts the minimality of $\deg f$. ❑

Cohen's criterion

One may wonder if there are conditions only involving prime ideals that ensure a ring being Noetherian. An ACC-condition on prime ideals is far from sufficient; there are non-Noetherian rings with merely one prime ideal. For instance, let $\mathfrak{m}$ be the ideal generated by all the x_i's in the ring $k[x_1, x_2, \ldots]$ of polynomials in countably many variables. The quotient $k[x_1, x_2, \ldots]/\mathfrak{m}^2$ has only one prime ideal, namely the one generated by the images of all the variables, but is not Noetherian since that ideal is not finitely generated. However, a result of Irvin Cohen's tells us that for a ring to be Noetherian it suffices that the prime ideals are finitely generated.

(9.33) We begin with stating a lemma about maximal ideals that are not finitely generated; it joins the line of results of the type asserting that ideals maximal subjected to some specific condition are prime:

LEMMA 9.34 *Let $\mathfrak{a}$ be maximal among the ideals in A that are not finitely generated. Then $\mathfrak{a}$ is a prime ideal.*

Cohen's criterion ensues easily from this lemma:

PROPOSITION 9.35 (COHEN'S CRITERION) *Assume that all prime ideals in the ring A are finitely generated. Then A is Noetherian.*

PROOF: Assume that A is not Noetherian. The set of ideals that are not finitely generated is then non-empty and according to Zorn's lemma has a maximal element, say $\mathfrak{a}$; indeed, if the union $\bigcup_i \mathfrak{a}_i$ of an ascending chain of ideals were finitely generated, the chain would stabilize (argue as in the last part of the proof of Proposition 9.11 on page 233)

and a member of the chain would be finitely generated. From the lemma we infer that $\mathfrak{a}$ is prime, which is a flagrant contradiction. ❑

PROOF OF LEMMA 9.34: The ring $A/\mathfrak{a}$ is Noetherian since all its ideals are finitely generated. Let a and a' be two members of A and assume that the product aa' lies in $\mathfrak{a}$, but that neither a nor a' lies there. Let $\mathfrak{c} = \mathfrak{a} + (a)$ and $\mathfrak{c}' = \mathfrak{a} + (a')$. These ideals both contain $\mathfrak{a}$ properly and are therefore finitely generated by the maximality of $\mathfrak{a}$, hence the product $\mathfrak{c}\mathfrak{c}'$ is finitely generated. Moreover, it holds that $\mathfrak{c}\mathfrak{c}' \subseteq \mathfrak{a}$, because $aa' \in \mathfrak{c}$. The quotient $\mathfrak{c}/\mathfrak{c}\mathfrak{c}'$ is a finitely generated module over the Noetherian ring $A/\mathfrak{a}$ and contains $\mathfrak{a}/\mathfrak{c}\mathfrak{c}'$. Hence $\mathfrak{a}/\mathfrak{c}\mathfrak{c}'$ is finitely generated, and by consequence $\mathfrak{a}$ is finitely generated as well since $\mathfrak{c}\mathfrak{c}'$ is finitely generated. ❑

Exercises

(9.12) *Hilbert Basis Theorem for Power Series.* Let A be a ring. The purpose of this exercise is to prove that if A is Noetherian, so is the power series ring $A[\![x]\!]$.

a) Let $\{g_i\}_{i\in\mathbb{N}_0}$ be a sequence of power series in $A[\![x]\!]$. Show that $\sum_i x^i g_i$ is a well defined power series in $A[\![x]\!]$.

Let $\phi\colon A[\![x]\!] \to A$ be the map that sends x to zero; *i. e.* a power series $f(x)$ is sent to the constant term $f(0)$. Let $\mathfrak{p}$ be a prime ideal in $A[\![x]\!]$ and assume that $\phi(\mathfrak{p}) = (a_1, \ldots, a_r)$. Chose elements f_i from $\mathfrak{p}$ so that $f_i(0) = a_i$.

b) If $x \in \mathfrak{p}$, show that $\mathfrak{p} = (x, f_1, \ldots, f_r)$.

c) If $x \notin \mathfrak{p}$, show that $\mathfrak{p} = (f_1, \ldots, f_r)$. HINT: Given $f \in \mathfrak{p}$; for each i recursively construct a power series h_i so that $f = h_1 f_1 + \ldots + h_r f_r$.

d) Conclude by Cohen's criterion that $A[\![x]\!]$ is Noetherian whenever A is.

★

Wolfgang Krull (1899–1975)

German mathematician

**The general version concerns an A-module.*

9.5 *Krull's intersection theorem*

The German mathematician Wolfgang Krull was one of the greatest contributor to the development of algebra in the years between the two World Wars, and in this section we shall discuss one of his more famous results, the so-called "Krull's intersection theorem". In its simplest form, the theorem asserts that all the powers $\mathfrak{a}^\nu$ of a proper ideal $\mathfrak{a}$ in a local Noetherian ring do not have common elements apart from 0; that is, it holds true that $\bigcap_\nu \mathfrak{a}^\nu = 0$.

There are several proofs of Krull's intersection theorem, and the one we give is among the shortest possible with the means at hand at the present stage of the course. There is another really simple and elementary proof for the case of the ring* itself due to Hervé Perdry which we present as Exercise 9.19.

EXAMPLE 9.6 To motivate and illustrate the reasons behind Krull's theorem, let us consider the ring of complex polynomials $\mathbb{C}[x_1, \ldots, x_r]$ and a point $a = (a_1, \ldots, a_r)$ in

$\mathbb{C}^r$. The ideal $\mathfrak{m} = (x_1 - a_1, \ldots, x_r - a_r)$ consists precisely of the polynomials that vanish at a, and the members of the powers $\mathfrak{m}^\nu$ are those that vanish to the ν-th order. In this simple situation Krull's theorem expresses the well-known and obvious fact that no non-zero polynomial vanishes to all orders. Another obvious example is found in the ring $\mathbb{Z}$ of integers where of course it holds that $\bigcap_\nu (p^\nu) = 0$ for any integer p, for the simple reason that no integer has infinitely many factors. Of course, Krull's result is a vast generalization of these prosaic examples; ideals in local Noetherian rings are infinitely more intricate than maximal ideals in a ring of complex polynomials or than principal ideals in $\mathbb{Z}$. ☆

(9.36) The show begins with a technical lemma, and again submodules maximal subjected to a specific condition enter the scene, but this time chiefly as catalysts. In general if $N \subseteq M$ is a pair of a module and a submodule and $\mathfrak{a}$ an ideal in A, the intersection $\mathfrak{a}M \cap N$ is not always contained in $\mathfrak{a}N$: elements in N might be divisible in M by elements from $\mathfrak{a}$, but not in N. However under appropriate finiteness conditions, if an element in N is "sufficiently divisible" in M, it will be divisible in N as well; one has the important:

LEMMA 9.37 *Let $\mathfrak{a} \subseteq A$ be a finitely generated ideal and M be a Noetherian A-module with a submodule N. If K is a submodule of M maximal subjected to the condition that $K \cap N = \mathfrak{a}N$, then $\mathfrak{a}^\nu M \subseteq K$ for a sufficiently large $\nu \in \mathbb{N}$. In particular it holds true that $\mathfrak{a}^\nu M \cap N \subseteq \mathfrak{a}N$.*

PROOF: Since $\mathfrak{a}$ is finitely generated, it suffices to show that $x^\nu M \subseteq K$ for every $x \in \mathfrak{a}$ and ν sufficiently big. By the maximality of K, it suffices to prove that $(x^\nu M + K) \cap N = \mathfrak{a}N$. The crucial inclusion is $(x^\nu M + K) \cap N \subseteq \mathfrak{a}N$, the other being clear as $\mathfrak{a}N = K \cap N$.

Now, the transporter submodules $(K : x^i)$ form an ascending chain, which since M is Noetherian, stabilizes at say ν; so that $(K : x^\nu) = (K : x^{\nu+1})$. If $y = x^\nu m + k$ with $m \in M$ and $k \in K$, is a member of $(x^\nu M + K) \cap N$ it holds that $xy \in xN \subseteq K \cap N$ from which ensues that $m \in (K : x^{\nu+1})$ since $xy = x^{\nu+1}m + xk$. Hence $m \in (K : x^\nu)$, and $y \in K \cap N = \mathfrak{a}N$. ❑

PROPOSITION 9.38 *Suppose that A is a ring, that $\mathfrak{a}$ is a finitely generated ideal in A and that M is a Noetherian module over A. Putting $N = \bigcap_i \mathfrak{a}^i M$, one has $\mathfrak{a}N = N$.*

PROOF: By the lemma there is a ν so that $\mathfrak{a}^\nu M \cap N \subseteq \mathfrak{a}N$. But by construction $N \subseteq \mathfrak{a}^\nu M$ and we conclude that $\mathfrak{a}N = N$. ❑

(9.39) Combining Proposition 9.38 above with Nakayama classic, we obtain the classical version of Krull's intersection theorem:

THEOREM 9.40 (KRULL'S INTERSECTION THEOREM) *Let A be ring and $\mathfrak{a}$ an ideal contained in the Jacobson radical of A. Assume that $\mathfrak{a}$ is finitely generated. If M is a Noetherian A-module, it holds true that $\bigcap_i \mathfrak{a}^i M = 0$.*

PROOF: Let $N = \bigcap_i \mathfrak{a}^i M$. Then $\mathfrak{a}N = N$ after Proposition 9.38, and we may finish by applying Nakayama's lemma (Proposition 4.52 on page 106) since N is a submodule of the Noetherian module M and therefore is finitely generated. ❑

COROLLARY 9.41 *Let A be a Noetherian ring and $\mathfrak{a}$ an ideal contained in the Jacobson radical of A. Then $\bigcap_i \mathfrak{a}^i = 0$. In particular, if A is a Noetherian local ring whose maximal ideal is $\mathfrak{m}$, one has $\bigcap_i \mathfrak{m}^i = 0$.*

(9.42) In general it is not true that the intersection of successive powers of a proper ideal vanishes even when the ring is Noetherian. Principal ideals generated by non-trivial idempotents furnish simple counterexamples: if $\mathfrak{a} = (e)$ with e idempotent, one has $\mathfrak{a}^2 = \mathfrak{a}$, and a straightforward induction shows that $\mathfrak{a}^i = \mathfrak{a}$ for all i. Hence $\bigcap_i \mathfrak{a}^i = \mathfrak{a}$. However, the powers of proper ideals in Noetherian integral domains have vanishing intersections:

COROLLARY 9.43 (KRULL'S INTERSECTION THEOREM II) *Assume that $\mathfrak{a}$ is a proper ideal in the Noetherian integral domain A, then $\bigcap_i \mathfrak{a}^i = 0$.*

PROOF: We combine Proposition 9.38 above by Nakayama Extended (Proposition 7.53 on page 198) and exhibit an element $a \in \mathfrak{a}$ so that $(1+a)N = 0$ where $N = \bigcap_i \mathfrak{a}^i$. But $\mathfrak{a}$ being proper, $1 + a$ is non-zero, and consequently $N = 0$ since A is an integral domain. ❑

Exercises

(9.13) Let $N \subseteq M$ be a pair of an A-module and a submodule. Let $x \in A$ be an element. Prove that $xM \cap N = xN$ if and only if x acts as non-zero divisor on M/N.

(9.14) Let ν be a natural number and p a prime. Describe the submodules N of $\mathbb{Z}$ so that $p^{\nu+1}\mathbb{Z} \cap N \subseteq pN$ but $p^{\nu}\mathbb{Z} \cap N \nsubseteq pN$.

(9.15) Let A be a domain which is contained in the Noetherian domain B. Let $\mathfrak{a}$ be an ideal in A. Show that it either holds true that $\mathfrak{a}B = B$ or that $\bigcap_i \mathfrak{a}^i = 0$.

(9.16) Let R be the ring of real functions that are defined and C^∞ on an interval $< -\epsilon, \epsilon >$ round 0. Let $\mathfrak{m}$ be the ideal of those functions in R that vanish at the origin. Show that $\mathfrak{m}$ is a maximal ideal, and that $\bigcap_i \mathfrak{m}^i$ consists of the functions in R all whose derivatives vanish at the origin. Give examples of such functions.

⁂ (9.17) Let A be a local ring with maximal ideal $\mathfrak{m}$ and assume that $\mathfrak{m}$ is a principal ideal generated by a non-nilpotent element.

a) Prove that $\mathfrak{p} = \bigcap_i \mathfrak{m}^i$ is a prime ideal and that all prime ideals in A other than $\mathfrak{m}$ are contained in $\mathfrak{p}$.
b) Prove that if $\bigcap_i \mathfrak{m}^i = (0)$, then the powers $\mathfrak{m}^i$ are the only non-zero ideals in A;
c) Prove that A is Noetherian if and only if $\bigcap_i \mathfrak{m}^i = (0)$.

(9.18) The aim of this exercise is to exhibit a domain A with a principal maximal ideal $\mathfrak{m}$ such the intersection $\bigcap_i \mathfrak{m}^i$ is non-zero. It is in some sense a minimal example of this behaviour, and illustrates how Krull's intersection theorem may fail in non-Noetherian rings.

Let k be a field and let A be the ring $k[t, x_1, x_2 \ldots]$ with constituting relations $x_i = tx_{i+1}$ for $i \in \mathbb{N}$.

a) Show that $\mathfrak{m} = (t)A$ is a maximal ideal and that the ideal $\mathfrak{p} = (x_1, x_2, \ldots)$ generated by all the x_i's is a prime ideal contained in $\mathfrak{m}$.
b) Prove that $\bigcap_i \mathfrak{m}^i = \mathfrak{p}$.
c) Let B be a domain containing k in which there is a principal ideal $\mathfrak{a} = (f)$ such that $\bigcap_i \mathfrak{a}^i \neq (0)$. Show that there is a map of k-algebras $A \to B$.

(9.19) *Perdry's proof of Krull's intersection theorem.* Let $\mathfrak{a}$ be an ideal in a Noetherian ring A and assume that $x \in \bigcap_i \mathfrak{a}^i$. The aim of the exercise is to prove that $x \in x \cdot \mathfrak{a}$. Assume that $a_1, \ldots, a_r$ are generators for $\mathfrak{a}$.

a) Let $\nu \in \mathbb{N}$ be a natural number. Use that $x \in \mathfrak{a}^\nu$ to prove there is a homogenous polynomial $P_\nu(x_1, \ldots, x_r)$ of degree ν in $A[x_1, \ldots, x_r]$ so that $x = P(a_1, \ldots, a_r)$.
b) Let $\mathfrak{c}_n$ be the ideal in $A[x_1, \ldots, x_r]$ generated by $P_1, \ldots, P_n$. Show there is an N so that $\mathfrak{c}_{N+1} = \mathfrak{c}_N$.
c) Show that one has a relation $P_{N+1} = \sum_{1 \leqslant i \leqslant N} Q_i \cdot P_i$ where the Q_i's are homogenous polynomials of positive degree.
d) Conclude that $x \in x\mathfrak{a}$.
e) Deduce hat $\bigcap_i \mathfrak{a}^i = 0$ if $\mathfrak{a}$ is contained in the Jacobson radical of A.

★

9.6 *Modules of finite length*

Finite dimensional vector space are civilized creatures having several features that make them pleasant to work with, one being that they have a dimension. Over any ring there is a class of modules with a numerical invariant attached resembling the dimension of a vector space. This invariant is called *the length*, and the modules in question are said to be of *finite length*, and it will turn out that this is equivalent to the modules being both Noetherian and Artinian. Modules do not possess bases in general, so it is a lot more involved to define the length than the dimension. The trick is to use certain *maximal chains* of submodules—maximal in the sense that there is no room for inserting new modules in the chain—the so-called *composition series*.

Maximal chains (maksimale kjeder)

(9.44) A finite ascending chain $\{M_i\}$ in an A-module M, which begins at the zero module 0 and ends at M, is called a *composition series* if all its subquotients M_{i+1}/M_i are simple modules. By convention simple modules are non-zero, so in particular all

Composition series (komposisjonsserier)

the inclusions $M_i \subseteq M_{i+1}$ are strict. The series when displayed appears like

$$0 = M_0 \subset M_1 \subset \ldots \subset M_{n-1} \subset M_n = M,$$

The length of a composition series (lengden av en komposisjonskjede)
The length of a chain (lengden av en kjede)

where each subquotient M_{i+1}/M_i is shaped like $A/\mathfrak{m}_i$ for some maximal ideal $\mathfrak{m}_i$ in A. The number n is called the *the length* of the series; it is the number of non-zero constituencies. More generally, *the length* of any finite chain will be the number of inclusions; that is, one less than the number of modules.

For a finite chain to be a composition series is equivalent to being a maximal chain; that is, no module can be inserted to make it longer. Such a chain must start at zero and end at M (if not, 0 or M could be adjoined), and there can be no submodules lying strictly between two consecutive terms. Would-be insertions are submodules in between M_i and M_{i+1}, which are in a one-to-one correspondence with submodules of M_{i+1}/M_i—a submodule N corresponds to the quotient N/M_i—so each M_{i+1}/M_i is simple precisely when there are no intermediate submodules. The term *saturated* is also common usage for suchlike chains, but these are neither required to start at the zero module nor to end at M.

Saturated chains (mettede kjeder)

(9.45) At several occasions in the subsequent paragraphs we shall push composition series forwards along A-linear maps, and it is a crucial fact that they stay composition series, though with a slight modification. To be precise, let $\beta\colon M \to N$ be an A-linear map and $\mathcal{M} = \{M_i\}$ a composition series in M. The set $\{\beta(M_i)\}$ of images is obviously a chain in M', but inclusions do not necessarily persist being strict, so there may be repetitions in $\{\beta(M_i)\}$. A part from that, $\{\beta(M_i)\}$ will be a composition series:

LEMMA 9.46 *Let $\beta\colon M \to N$ and let $\{M_i\}$ be a composition series in M. Disregarding possible repetitions, the chain $\{\beta(M_i)\}$ will be a composition series.*

PROOF: A simple diagram-hunt (or a snake argument) yields that γ in the diagram below is surjective. Hence, as M_i/M_{i-1} is simple, the module $\beta(M_i)/\beta(M_{i-1})$ is either zero or isomorphic to M_i/M_{i-1}. Consequently, when repeating terms are discarded, the subquotient of $\{\beta(M_i)\}$ are all simple.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \beta(M_{i-1}) & \longrightarrow & \beta(M_i) & \longrightarrow & \beta(M_i)/\beta(M_{i-1}) & \longrightarrow & 0 \\ & & \uparrow & & \uparrow & & \uparrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & M_{i-1} & \longrightarrow & M_i & \longrightarrow & M_i/M_{i-1} & \longrightarrow & 0. \end{array}$$

❑

⁂ EXERCISE 9.20 Assume that $\alpha\colon N \to M$ is an injective A-linear map. If $\mathcal{M} = \{M_i\}$ is a composition series in M, show that the chain $\{\alpha^{-1}(M_i)\}$ will be a composition series in N after possible repeating terms have been discarded. ★

(9.47) The main result of this section is that once a module has a composition series,

all composition series will have the same length, and the length of any chain will be bounded by that common number. This is a result of Jordan-Hölder type, but one on the weak side—the true Jordan-Hölder theorem states that the subquotients of any two composition series are isomorphic up to a permutation. The original Jodan-Hölder theorem is about (finite) groups, but one finds analogues in many categories, so also in the subcategory of Mod_A of finite lengths modules.

THEOREM 9.48 (WEAK JORDAN-HÖLDER) *Assume that M has a composition series. Then all composition series in M have the same length and any chain may be completed to a composition series.*

The length of a module (lengden til en modul)

The common length of the composition series is called the *length* of the module and denoted $\ell_A(M)$. For modules not of finite length, that is those having no composition series, one naturally writes $\ell_A(M) = \infty$. As a matter of pedantry, the zero module* is considered to be of finite length and its length is zero (what else?). Note that the zero module is the only module of length zero, and that the simple ones are the only ones of length one, indeed, having length one, means that $(0) \subset M$ is the one and only composition series.

*This is a pure formality: by convention the zero module is not included among the simple modules and does therefore not have a composition series!

PROOF: The proof goes by induction on the length of the shortest composition series in a module; this is well defined and finite for all modules concerning us. A module having a composition series of length one is simple, and for those the theorem is obviously true. The induction can begin and the fun can start.

Let $\mathcal{M} = \{M_i\}$ be a composition series of minimal length n in M, which displayed is shaped like

$$0 = M_0 \subset M_1 \subset \ldots \subset M_n = M.$$

The image of $\mathcal{M}$ in the quotient M/M_1 is a composition series of length one shorter than $\mathcal{M}$, hence all composition series in M/M_1 are of length $n-1$ by induction. Denote by $\beta\colon M \to M/M_1$ the quotient map.

Given another composition series $\mathcal{N} = \{N_j\}$ in M. Its length r is at least n, and by induction its image in M/M_1 is a composition series of length $n-1$. Consequently at least one of the inclusions in $\mathcal{N}$ becomes an equality in M/M_1; that is, for some index ν it holds that $\beta(N_\mu) = \beta(N_{\mu+1})$, and we may pick ν to be the least such index. Then $\beta(\mathcal{N})$ displays as

$$0 = \beta(N_0) \subset \beta(N_1) \subset \ldots \subset \beta(N_\nu) = \beta(N_{\nu+1}) \subseteq \ldots \subseteq \beta(N_r) = M. \tag{9.2}$$

We contend that ν is the only index for which equality occurs— this is the fulcrum of the proof from which it clearly ensues that $r = n$: indeed, on the one hand, the length of $\beta(\mathcal{N})$ will then be one less than that of $\mathcal{N}$, and on the other hand, it equals $n-1$ by induction.

From the equality $\beta(N_\nu) = \beta(N_{\nu+1})$ ensues the equality $N_\nu + M_1 \cap N_{\nu+1} = N_{\nu+1}$. Now, M_1 is simple, so either $M_1 \cap N_{\nu+1}$ vanishes or equals M_1. However, it cannot vanish because $N_\nu \neq N_{\nu+1}$, and we infer that $M_1 \subseteq N_{\nu+1}$. It follows that there are strict inclusions $\beta(N_j) \subset \beta(N_{j+1})$ for $j \geqslant \nu + 1$, and hence all inclusions in $\beta(\mathcal{N})$ are strict except the one at stage ν.

The last statement in the theorem, that every chain $\mathcal{M}$ can be completed to a composition series, is easily proved by induction on the length. We may assume that $\mathcal{M}$ has more than two terms. Let M_ν any non-zero term of $\mathcal{M}$ which is different from M, and consider the quotient M/M_ν. Both M_ν and M/M_ν are of length less than M, and the induction hypothesis applies to both. The chain $\mathcal{M}' = \{M_i\}_{i \leqslant \nu}$ in M_ν can therefore be completed to a composition series, and similarly, if $\beta\colon M \to M/M_\nu$ is the quotient map, the chain $\{\beta(M_i)\}_{i>\nu}$ can be completed to one in M/M_ν. Pulling the latter back to M and splicing it with the former yields a composition series in M extending $\mathcal{M}$. ❑ (9.49) A closer look at the proof above reveals that it in fact gives the full Jordan-Hölder theorem:

THEOREM 9.50 (TRUE JORDAN-HÖLDER) *Any two composition series of a module of finite length have up to order the same subquotients.*

PROOF: We resume the proof of the previous theorem, keep the notation and carry on with induction on the length: By induction the two series $\beta(\mathcal{M})$ and $\beta(\mathcal{N})$ have the same subquotients up to order. Now, the subquotients of $\mathcal{M}$ and $\beta(\mathcal{M})$ differ only at the bottom stage M_1, so $\mathcal{M}$ has the subquotient M_1 in addition to those shared with $\beta(\mathcal{M})$. On the other hand, the subquotients of $\mathcal{N}$ and $\beta(\mathcal{N})$ coincide except at a certain stage ν, but in the proof above we showed that $N_\nu + M_1 = N_{\nu+1}$, and since $(N_\nu + M_1)/N_\nu \simeq M_1$, the additional subquotient of $\mathcal{N}$ is isomorphic to M_1 as well. ❑ (9.51) Just like the dimension of vector spaces the length is additive along short exact sequence, which is an indispensable property that makes it possible to compute the length in many cases. Observe also that a submodule (or a quotient) of M having the same length as M must be equal to M.

PROPOSITION 9.52 (ADDITIVITY OF LENGTH) *Assume given a short exact sequence of A-modules.*

$$0 \longrightarrow M' \xrightarrow{\alpha} M \xrightarrow{\beta} M'' \longrightarrow 0.$$

Then M is of finite length if and only if the two others are, and it holds true that $\ell_A(M) = \ell_A(M') + \ell_A(M'')$.

PROOF: Assume first that M is of finite length. Pushing a finite composition series forward along β gives one in M'' and pulling it back along α gives one in M', so M' and M'' are both of finite length.

Assume next that the two modules M' and M'' are of finite length. It suffices to exhibit one composition series of M with the additive property. To this end, we begin with a composition series in M'', say $\{M''_i\}$, and pull it back to M along β. The smallest module in the pulled back chain is $\beta^{-1}(M''_0) = \beta^{-1}(0)$, which equals M', so we may splice $\{\beta^{-1}(M''_i)\}$ with any composition series of M' to obtain one in M, and obviously, the length of the spliced series equals the sum of lengths of the two being spliced. ❑

(9.53) An immediate corollary of Proposition 9.52 is that modules of finite length are both Noetherian and Artinian. Obviously this is true for simple modules (no non-trivial submodules, no non-trivial chains) and hence follows in general by a straightforward induction on the length using Proposition 9.7 on page 231. The converse holds as well:

PROPOSITION 9.54 *An A-module M is of finite length if and only if it is both Noetherian and Artinian.*

PROOF: Assume that M is both Noetherian and Artinian. Since M is Artinian every non-empty set of submodules has a minimal element, so if M is not of finite length, there is a submodule, N say, minimal subjected to being non-zero and not of finite length. It is finitely generated because M is Noetherian and hence Nakayama's lemma applies: There is surjection $\phi\colon M \to k$ onto a simple module k. The kernel of ϕ is of finite length by the minimality of N, and hence N itself is of finite length by Proposition 9.52 above. ❑

(9.55) Be aware that the base ring A is a serious part of the game and can have a decisive effect on the length of a module. If $A \to B$ is a map of rings and M a B-module which is of finite length over both A and B, there is in general no reason that $\ell_A(M)$ and $\ell_B(M)$ should agree. Already when $k \subseteq K$ is a finite non-trivial extension of fields the two lengths differ in that $\dim_K V = [K;k] \dim_k V$ for a vector space V over K. You will find a simple but slightly more subtle example in Example 9.10 below. And of course there are stupid examples like $\mathbb{Q} \subseteq \mathbb{R}$ with $\mathbb{R}$ being of length one over itself, but as a module over $\mathbb{Q}$ its length is infinite (the dimension is even uncountable!)

However, when the map $A \to B$ is surjective, the two lengths agree since then the B-submodules and the A-submodules of M coincide.

(9.56) Unlike what is true for vector spaces, module of the same length need not be isomorphic. Simple examples are the $\mathbb{Z}$-modules $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for different primes p. They are all of length one but no two are isomorphic.

Examples

(9.7) *Vector spaces:* Over fields modules are just vector spaces, and as we verified in Example 9.1 on page 232, for vector spaces it holds that being Noetherian or Artinian is equivalent to being of finite dimension; hence also being of finite length is equivalent to being of finite dimension. And the length of a vector space coincides with its vector

space dimension: indeed, given a basis $v_1, \dots, v_r$, the subspaces $V_i =< v_1, \dots, v_i >$ form a composition series, and any composition series is readily seen to be of this form as the field itself is the only simple module.

In a similar fashion, if the A-module M is killed by a maximal ideal $\mathfrak{m}$ in A, and therefore is a vector space over the field $A/\mathfrak{m}$, one has $\ell_A(M) = \dim_{A/\mathfrak{m}} M$. And M is of finite length over A if and only if it is finitely generated

(9.8) *Finite abelian groups:* The only abelian groups that are of finite length are the finite ones. They are all direct sums of cyclic groups of shape $\mathbb{Z}/p^\nu\mathbb{Z}$ where p is a prime and ν a natural number; that is, such a group M enjoys a finite direct sum decomposition

$$M = \bigoplus_i \mathbb{Z}/p_i^{\nu_i}\mathbb{Z}.$$

*The map $[p]$ is close to being a "multiplication-by-p-map": it sends a class $[x]$ mod $p^{\nu-1}$ to the class $[px]$ mod p^ν.

We contend that $\ell_{\mathbb{Z}}(M) = \sum_i \nu_i$. In other words, the length $\ell_{\mathbb{Z}}(M)$ equals the sum of the multiplicities of the different primes in the prime factorization of the order $|M|$ of M.

By additivity of length it suffices to show that for a each prime p the length of the cyclic group $\mathbb{Z}/p^\nu\mathbb{Z}$ is given as $\ell_{\mathbb{Z}}(\mathbb{Z}/p^\nu\mathbb{Z}) = \nu$, and this one does by an inductive argument based on the standard* short exact sequences

$$0 \longrightarrow \mathbb{Z}/p^{\nu-1}\mathbb{Z} \xrightarrow{[p]} \mathbb{Z}/p^{\nu}\mathbb{Z} \longrightarrow \mathbb{Z}/p\mathbb{Z} \longrightarrow 0.$$

Since $\ell_{\mathbb{Z}}(\mathbb{Z}/p\mathbb{Z}) = 1$, the length $\ell_A(\mathbb{Z}/p^\nu\mathbb{Z})$ increases by one when ν is increased by one, and we are done.

(9.9) Let $A = k[x, y]$ and $\mathfrak{m} = (x, y)$. For each non-negative integer n let $M_n = k[x, y]/\mathfrak{m}^n$. Then there are short exact sequences

$$0 \longrightarrow \mathfrak{m}^{n-1}/\mathfrak{m}^n \longrightarrow M_n \longrightarrow M_{n-1} \longrightarrow 0,$$

so that $\ell_A(M_n) = \ell_A(M_{n-1}) + \ell_A(\mathfrak{m}^{n-1}/\mathfrak{m}^n)$. The module $\mathfrak{m}^{n-1}/\mathfrak{m}^n$ is a vector space over the field $A/\mathfrak{m} = k$ having the classes of the monomials $x^i y^{n-1-i}$ for $0 \leqslant i \leqslant n-1$ as a basis, and hence $\ell_A(\mathfrak{m}^{n-1}/\mathfrak{m}^n) = \dim_k \mathfrak{m}^{n-1}/\mathfrak{m}^n = n$. We conclude that $\ell_A(M_n) = \ell_A(M_{n-1}) + n$ and induction on n yields that

$$\ell_A(M_n) = \sum_{i=1}^{n} i = \binom{n+1}{2}.$$

(9.10) We let $A = \mathbb{Z}$ and $B = \mathbb{Z}[i] = \mathbb{Z}[x]/(x^2+1)$ and let $M = \mathbb{Z}[i]/(105)\mathbb{Z}[i]$. The prime factorization of 105 is $105 = 3 \cdot 5 \cdot 7$, and one checks easily that $x^2 + 1$ is irreducible mod 3 and 7 but decomposes over $\mathbb{F}_5$; the primes 3 and 7 persist being primes in $\mathbb{Z}[i]$,

but 5 splits up in the product $5 = (2+i)(2-i)$. The Chinese Remainder Theorem gives a decomposition as B-modules

$$M = \mathbb{F}_3(i) \oplus \mathbb{F}_5 \oplus \mathbb{F}_5 \oplus \mathbb{F}_7(i),$$

where $\mathbb{F}_3(i)$ and $\mathbb{F}_7(i)$ are fields and the two $\mathbb{F}_5$'s are $\mathbb{Z}[i]$ modules with i acting as multiplication by 2 on one and by -2 on the other. We conclude that $\ell_{\mathbb{Z}}(M) = 6$ but $\ell_{\mathbb{Z}[i]}(M) = 4$.

☆

Exercises

(9.21) Compute the length of $\mathbb{Z}[i]/525$ both as a $\mathbb{Z}$-module and as a $\mathbb{Z}[i]$-module.

(9.22) Show that the length $\ell_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z})$ equals the total number of prime factors of n (counted with multiplicity).

(9.23) Let $n = p_1 \cdot \ldots \cdot p_r$ be the prime factorization of a square free natural number and consider a quadratic extension $\mathbb{Z} \subset A$. Show that $\ell_{\mathbb{Z}}(A/(n)A) = \ell_A(A/(n)A)$ if and only if each prime factor p_i persists being a prime in A.

(9.24) Let M be a module of finite length over A and let $f \in A$. Show that the quotient M/fM and the annihilator $(0 : f)_M = \{\, x \in M \mid fx = 0 \,\}$ are of the same finite length; that is, $\ell_A(M/fM) = \ell_A((0 : f)_M)$

(9.25) *Modules of finite length over* PID*'s.* Let A be PID and let $f \in A$ be an element. Show that $\ell_A(A/(f)A)$ is the number of prime factors in f (counted with multiplicity).

(9.26) Assume that M is an A-module of finite length and that $\mathfrak{a}$ is an ideal contained in the Jacobson radical of A. Show that for some integer n it holds true that $\mathfrak{a}^n M = 0$. HINT: Consider the descending chain $\mathfrak{a}^i M$ and remember Nakayama's lemma.

(9.27) A frequently met situation in algebraic geometry is that a ring A is a k-algebra; that is, it contains a ground field k (for instance, algebras like $k[x_1, \ldots, x_r]/\mathfrak{a}$ are of this type). Then any A-module is a vector space over k. Assume that A in addition to being a k-algebra is local ring. Denote the maximal ideal by $\mathfrak{m}$ and let $k(\mathfrak{m}) = A/\mathfrak{m}$ be the residue class field.

a) Assume that k maps isomorphically onto $k(\mathfrak{m})$. Prove that a module M is of finite length over A if and only if is of finite dimension over k and in case it holds true that $\dim_k M = \ell_A(M)$.

b) Assume merely that $k(\mathfrak{m})$ is finite extension of the image of k. Prove that $\dim_k M = [k(\mathfrak{m}) : k] \cdot \ell_A(M)$.

(9.28) *Modules of finite length over finite product of fields.* Let $A = \prod_{1 \leqslant i \leqslant r} k_i$ be a finite product of fields. Show that an A-module $V = \bigoplus_{1 \leqslant i \leqslant r} V_i$, where V_i is a vector space over k_i, is of finite length if and only if each V_i is of finite dimension over k_i, and in that

case it holds true that

$$\ell_A(\bigoplus_i V_i) = \sum_i \dim_{k_i} V_i.$$

★

Finite length and support

(9.57) We finish of the story about modules of finite length with a criterion for a module to be of finite length in terms of the support, and a structure theorem, which essentially says that modules of finite length are direct sums of "local contributions"—but of course, it says nothing about how the local contributions are shaped.

PROPOSITION 9.58 *A finitely generated module M over a Noetherian ring is of finite length if and only if its support* Supp M *is a finite union of closed points.*

PROOF: Assume to begin with that M is of finite length and let $\{M_i\}$ be a composition series. Citing Proposition 7.59 on page 200 we infer that the support of M satisfies $\operatorname{Supp} M = \bigcup_i \operatorname{Supp} M_i/M_{i+1} = \bigcup_i \{\mathfrak{m}_i\}$ where $M_i/M_{i+1} \simeq A/\mathfrak{m}_i$ are the subquotients of the composition series $\{M_i\}$. So the support is a finite union of closed points*.

*Recall that the closed points in Spec A are precisely the maximal ideals. It may well happen that a subset of Spec A is closed and finite without all points being closed; for instance, Spec $\mathbb{Z}_{(p)}$ is finite.

For the other implication we resort to the Structure Theorem (Theorem 9.23 on page 238) which assures that there is a descending chain $\{M_i\}$ of submodules in M whose subquotients are shaped like $A/\mathfrak{p}_i$ with $\mathfrak{p}_i$ being prime. Again by Proposition 7.59 it holds that $\operatorname{Supp} M = \bigcup_i V(\mathfrak{p}_i)$. Now, if Supp M consists of finitely many closed points, all the prime ideals $\mathfrak{p}_i$'s will be maximal and consequently all the subquotients M_i/M_{i-1} will be fields. Hence M is of finite length (in fact, the chain $\{M_i\}$ will be a composition series). ❑

One part of the proposition, that the support of a module of finite length is finite and discrete, holds true without hypotheses on A, however the other implication depends on A being Noetherian. An example can be $A = B/\mathfrak{m}^2$ where $B = k[x_i | i \in \mathbb{N}]$ is the polynomial ring in countably many variables and $\mathfrak{m} = (x_i | i \in \mathbb{N})$. Then A is neither Artinian nor Noetherian, but has only one prime ideal $\mathfrak{m}/\mathfrak{m}^2$.

There are many rings that are not Artinian, but whose spectrum is finite. For instance the localization $\mathbb{Z}_p$ of the integers at the prime (p) is not Artinian, but Spec $\mathbb{Z}_{(p)}$ has only two points namely (0) and $p\mathbb{Z}_{(p)}$. All members of the support being maximal, is therefore a crucial part of the hypothesis.

The structure of modules of finite length

(9.59) Here comes the promised result about the structure of modules of finite length over a ring A, they compose as the direct sum of their localizations. Given an A-module M. For each maximal ideal $\mathfrak{m} \subseteq A$ there is a localization maps $\iota_{\mathfrak{m}} : M \to M_{\mathfrak{m}}$, and we may combine them into a map $\phi : M \to \bigoplus_{\mathfrak{m} \in \operatorname{Supp} M} M_{\mathfrak{m}}$.

PROPOSITION 9.60 (STRUCTURE OF FINITE LENGTH MODULES) *Assume that M is a module of finite length over the ring A. Then the a canonical ϕ is an isomorphism*

$$M \simeq \bigoplus_{\mathfrak{m} \in \operatorname{Supp} M} M_{\mathfrak{m}}.$$

PROOF: The proof will be an application of the local to global principle, more precisely the one asserting that being an isomorphism is a local property of A-linear maps (Proposition 7.49 on page 197). Our task is therefore to establish that the localizations $\phi_{\mathfrak{m}}$ of ϕ are isomorphisms for all $\mathfrak{m}$. To cope with the double localizations that appear, we notice the following lemma:

LEMMA 9.61 *Let M be a module of finite length over a ring A and let the maximal ideal $\mathfrak{m}$ belong to* Supp *M. Then $M_{\mathfrak{m}}$ is of finite length and has support $\{\mathfrak{m}\}$.*

PROOF OF THE LEMMA: Let $\{M_i\}$ be a composition series in M with $M_i/M_{i-1} \simeq A/\mathfrak{m}_i$. Now we contend that

$$(A/\mathfrak{m}_i)_{\mathfrak{m}} = \begin{cases} 0 & \text{when } \mathfrak{m} \neq \mathfrak{m}_i; \\ A/\mathfrak{m} & \text{when } \mathfrak{m} = \mathfrak{m}_i. \end{cases}$$

Indeed, $\mathfrak{m} \neq \mathfrak{m}_i$ is equivalent to $\mathfrak{m}_i \nsubseteq \mathfrak{m}$ (both are maximal ideals), and hence to there being elements in $\mathfrak{m}_i$ not in $\mathfrak{m}$. So some element killing $A/\mathfrak{m}_i$ gets inverted in the localization, and it holds that $(A/\mathfrak{m}_i)_{\mathfrak{m}} = 0$ when $\mathfrak{m} \neq \mathfrak{m}_i$. Furthermore, it obviously holds that $(A/\mathfrak{m})_{\mathfrak{m}} = A/\mathfrak{m}$ (elements not in $\mathfrak{m}$ act as inversions on the field $A/\mathfrak{m}$). We infer that after possible repetitions are discarded, the chain $\{(M_i)_{\mathfrak{m}}\}$ is a composition series in $M_{\mathfrak{m}}$ with all subquotients equal to $A/\mathfrak{m}$. ❑

With this lemma up our sleeve, it follows painlessly that $\phi_{\mathfrak{m}}$ is an isomorphism for all $\mathfrak{m}$. To fix the ideas let $\operatorname{Supp} M = \{\mathfrak{m}_1, \ldots, \mathfrak{m}_r\}$. If $\mathfrak{m} \notin \operatorname{Supp} M$, we have $M_{\mathfrak{m}} = 0$ and $(\bigoplus M_{\mathfrak{m}_i})_{\mathfrak{m}} = 0$, so $\phi_{\mathfrak{m}}$ is the zero map (which is an isomorphism in this case). If $\mathfrak{m}$ is one of the $\mathfrak{m}_i$'s, say $\mathfrak{m} = \mathfrak{m}_j$, the lemma gives

$$\Big(\bigoplus_i M_{\mathfrak{m}_i}\Big)_{\mathfrak{m}} = \Big(\bigoplus_i M_{\mathfrak{m}_i}\Big)_{\mathfrak{m}_j} \simeq \bigoplus_i \big(M_{\mathfrak{m}_i}\big)_{\mathfrak{m}_j} = M_{\mathfrak{m}_j},$$

and $\phi_{\mathfrak{m}} : M_{\mathfrak{m}} \to M_{\mathfrak{m}}$ is the identity map. ❑

9.7 *Artinian rings*

We now turn to the rings whose ideals satisfy the descending chain condition; that is, rings that are Artinian modules over themselves. Even though the definitions may appear symmetric, the class of Artinian rings is astonishingly different from the class of

Noetherian rings. The latter is a large class encompassing almost all rings one meets in algebraic geometry, whereas the Artinian ones merely serve special (but important) purposes. They are the tiny, little brothers among the Noetherian rings—but even though being "small", they are far from being insignificant, and they can indeed carry a great lot of subtleties.

(9.62) It turns out, as we shorty shall see, that Artinian rings are Noetherian. This is specific for Artinian *rings*, but far from being true for Artinian modules. The Artinian rings are characterized among the Noetherian ones by the property that all their prime ideals are maximal and that the maximal ideals are finite in number. In geometric terms, their spectra are finite sets and the Zariski topology is discrete. For a finite spectrum, the Zariski topology being discrete is equivalent to all points being closed, but this is no more true for infinite spectra (see Exercise 9.33 for an example).

Yasuo Akizuki (1902–1984) Japanese mathematician

The theorem we are about to prove is due to the Japanese mathematician Yasuo Akizuki. There is an analogue version valid for non-commutative rings, proved about at the same time as Akizuki proved his theorem, which usually is contributed to Charles Hopkins and Jacob Levitzki, but the commutative version is Akizuki's:

THEOREM 9.63 (AKIZUKI'S THEOREM) *An Artinian ring is Noetherian. Hence A has only finitely many prime ideals which all are maximal.*

(9.64) Recall that a module which is both Noetherian and Artinian is of finite length. This shows that Artinian rings are of finite length (regarded as modules over themselves) and hence come with a natural numerical invariant, the length $l_A(A)$; that is, the number of (simple) subquotients in a composition series. The second statement in the Theorem is a consequence of A being of finite length as proven in Proposition 9.54, but we have stated it like that since the proof follows the reverse path: one first proves that Spec A is discrete and finite and subsequently that A is Noetherian.

The proof of Akizuki's theorem $\Longrightarrow$

The proof of Akizuki's theorem is organized as a sequence of three lemmas. The first is about Artinian domains:

LEMMA 9.65 *An Artinian domain A is a field.*

PROOF: Let $f \in A$ be a non-zero member of A. The principal ideals (f^i) form a descending chain which must ultimately be constant; that is, $(f^{\nu+1}) = (f^\nu)$ for some ν. Then $f^\nu = af^{\nu+1}$ for some $a \in A$, and cancelling f^ν, which is permissible as A is a domain, we find $1 = af$; *i. e.* f is invertible. ❑

We proceed with proving the second statement in the theorem, the one about prime ideals being maximal and finite in number. This is the easy piece, that A is Noetherian, is deeper.

LEMMA 9.66 *An Artinian ring A has only finitely many prime ideals, and they are all maximal. Hence, if J denotes the radical of A, the quotient A/J is a finite product of fields.*

PROOF: We have already established the first assertion, if $\mathfrak{p}$ is a prime in A, the quotient $A/\mathfrak{p}$ is an Artinian domain, hence a field by Lemma 9.65 above.

As to the second statement, assume that $\{\mathfrak{m}_i\}_{i\in\mathbb{N}}$ is a countable set of different maximal ideals in A. For each natural number r consider the ideal $N_r = \bigcap_{i\leqslant r} \mathfrak{m}_i$. They form a descending chain, and A being Artinian it holds true that $N_\nu = N_{\nu+1}$ for some ν. This means that $\bigcap_{i\leqslant\nu} \mathfrak{m}_i \subseteq \mathfrak{m}_{\nu+1}$, and by Proposition 2.28 on page 40, one of the $\mathfrak{m}_i$'s must lie in $\mathfrak{m}_{\nu+1}$, contradicting the assumption that the $\mathfrak{m}_i$'s are different.

The last assertion ensues from the The Chinese Remainder Theorem. If $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ are the prime ideals in A, the radical J equals $J = \bigcap_i \mathfrak{m}_i$, and since all the $\mathfrak{m}_i$'s are maximal, they are pair-wise comaximal. Hence the Chinese Remainder Theorem gives an isomorphism $A/J \simeq \prod_i A/\mathfrak{m}_i$. ❑

The preceding lemma implies that the radical J of A coincides with intersection of the maximal ideals in A; that is, J is also the Jacobson radical of A. The elements are nilpotent, but A is not *a priori* Noetherian, so J is not *a priori* a nilpotent ideal. However, our next lemma says it is.

LEMMA 9.67 *The radical J of an Artinian ring A is nilpotent; that is, $J^n = 0$ for some n. Moreover, J is Noetherian.*

This lemma concludes the proof of Akizuki's theorem. Since A/J, being the product of finite number of fields is Noetherian, we infer that A is Noetherian citing Proposition 9.7 on page 231.

PROOF: The descending chain of powers $\{J^\nu\}$ becomes stationary at a certain stage; that is, there is an r such that $J^{r+1} = J^r$. Putting $\mathfrak{a} = \operatorname{Ann} J^r$ one finds

$$(\mathfrak{a} : J) = \{\, x \mid xJ \subseteq \operatorname{Ann} J^r \,\} = \operatorname{Ann} J^{r+1} = \mathfrak{a}.$$

If $\mathfrak{a} = A$, then $J^r = 0$ and we are happy, so assume that $\mathfrak{a}$ is a proper ideal, and let $\mathfrak{b}$ be a minimal ideal strictly containing $\mathfrak{a}$; such exist since A is Artinian. Let $x \in \mathfrak{b}$ but $x \notin \mathfrak{a}$, then $\mathfrak{b} = \mathfrak{a} + Ax$. If $\mathfrak{a} + Jx = \mathfrak{b}$, it follows that $Ax/Ax \cap \mathfrak{a} = \mathfrak{b}/Ax \cap \mathfrak{a} = J \cdot Ax/Ax \cap \mathfrak{a}$, and Nakayama's lemma* yields that $Ax/Ax \cap \mathfrak{a} = 0$; that is, $Ax \subseteq \mathfrak{a}$, which is not the case. Hence $\mathfrak{a} + Jx$ is strictly contain in $\mathfrak{b}$, and by minimality $\mathfrak{a} + Jx = \mathfrak{a}$. Hence $x \in (\mathfrak{a} : J) = \mathfrak{a}$, which is a contradiction.

*$Ax/Ax \cap \mathfrak{a}$ is finitely generated over A and J is the Jacobson radical of A.

The final step of the proof of Akizuki's theorem is an induction argument to show that J is Noetherian. For ν sufficiently big, we saw above that $J^\nu = 0$, and for each ν there is a short exact sequence:

$$0 \longrightarrow J^{\nu+1} \longrightarrow J^\nu \longrightarrow J^\nu/J^{\nu+1} \longrightarrow 0.$$

Submodules and quotients of Artinian modules are Artinian (as explained in Proposition 9.7 on page 231), so it follows that J^ν, and therefore also $J^\nu/j^{\nu+1}$, is Artinian. But $J^\nu/J^{\nu+1}$ is a module over A/J which we just proved is a finite product of fields, and over such rings any Artinian module is Noetherian (Example 9.2 on page 232); and we are through by descending induction on ν. ❑

The structure of Artinian rings

Since Artinian rings are of finite length over themselves, we may apply the Structure Theorems (Proposition 9.58 and Corollary 9.60 both on page 252) to obtain the following description:

THEOREM 9.68 *Let A be an Artinian ring. Then* Spec A *is finite and discrete, and the localisation maps $A \to A_\mathfrak{m}$ induce an isomorphism*

$$A \simeq \prod_{\mathfrak{m}\in\operatorname{Spec} A} A_\mathfrak{m}.$$

If A is Noetherian and Spec A *is finite and discrete, then A is Artinian.*

Saying Spec A is finite and discrete is just another way of saying that all prime ideals in A are maximal and finite in number. Anticipating the notion of *Krull dimension*, a ring all whose prime ideals are maximal is said to be of Krull dimension zero. Hence a Noetherian ring A is Artinian if and only if its Krull dimension equals zero.

The theorem says nothing about local Artinian rings, even if they might appear small and innocuous, they can be extremely intricate creatures.

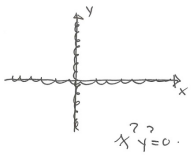
Examples

(9.11) *Local rings at minimal primes, multiplicities of components:* A Noetherian ring with merely one prime ideal is necessarily Artinian. The prime ideal must be maximal, Supp $A =$ Spec A has one single point, and we may hence resort to Proposition 9.58. A particular case of this situation arises when one localizes a ring at a minimal prime: If $\mathfrak{p}$ is a minimal prime in the (Noetherian) ring A, the local ring $A_\mathfrak{p}$ has the sole prime ideal $\mathfrak{p}A_\mathfrak{p}$, since its prime ideals correspond to those in A contained in $\mathfrak{p}$, and so $A_\mathfrak{p}$ is Artinian. The closed subset $V(\mathfrak{p})$ is one of the irreducible components of Spec A, and the length $\ell_A(A)_\mathfrak{p}$ is called the *multiplicity of the component* $V(\mathfrak{p})$.

Multiplicity of a component (multiplisitet til en komponent)

(9.12) *Multiplicities of components:* Consider the ring $A = k[x,y]$ with constituting relation $x^2y^2 = 0$. One checks that both $k[x]$ and $k[y]$ are polynomial rings and that in A it holds true that $0 = (x^2) \cap (y^2)$. So the only minimal primes of A are the principal ideals (x) and (y). We contend that $A_{(x)} = K[x]/(x^2)$ where K is the rational function field $K = k(y)$; it is easily seen that this ring is of length two; indeed, Exercise 7.32 on page 189 yields the equality $k[X,Y]_{(X)} = K[X]_{(X)}$, and we may conclude that

$A_{(x)} = \big(k[X,Y]/(X^2Y^2)\big)_{(X)} = K[X]/X^2$, and the claim follows. The geometric picture of $V(X^2Y^2)$ is the union of the X-axis and the Y-axis in the plane Spec $k[X,Y]$, but with the multiplicities two attached, and one may think about them as being "doubled". The multiplicities reappears as the lengths of the localized rings $A_\mathfrak{p}$ at minimal primes.



(9.13) *Intersection multiplicities:* In the theory of plane curves the so-called Bézout's theorem is central; it predicts the number of intersection points between two plane curves. Important ingredients of the formulation are the "local intersection multiplicities", which manifests themselves algebraically as the lengths of certain Artinian rings. The equations of the curves are polynomials f and g in $A = k[x,y]$, and we assume that f and g are without common factors.

In view of Proposition 3.30 on page 77 any prime ideal containing (f,g) is then maximal and by 9.17 on page 235 there are only finitely many such, say $\mathfrak{m}_1, \dots, \mathfrak{m}_r$ (they are the minimal primes of (f,g)); hence Supp $A/(f,g)A$ is finite and discrete and $A/(f,g)A$ is an Artinian ring. It decomposes as $A/(f,g)A = \prod_i A_{\mathfrak{m}_i}/(f,g)A_{\mathfrak{m}_i}$. The local intersection multiplicity at $\mathfrak{m}_i$ is $(f,g)_{\mathfrak{m}_i} = \ell(A)_{\mathfrak{m}_i}/(f,g)A_{\mathfrak{m}_i}$, and the total number of intersection points is

$$\ell(A/(f,g)A) = \sum_i \ell(A_{\mathfrak{m}_i}/(f,g)A_{\mathfrak{m}_i}).$$

In Bézout's theorem there are also local contributions at infinity occurring, and including those, the theorem states that the total count will be $\deg f \cdot \deg g$. For instance, two lines can be parallel, and their only intersection point then lies at infinity.

(9.14) Consider the intersection of the line $y = \alpha x$ and the curve $y^2 = x^3$; so that we put $f = y - \alpha x$ and $g = y^2 - x^3$. Assume first $\alpha \neq 0$, we then

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

So $A/(f,g) \simeq k[x]/(x^2) \times k$. The support has two points, the origin where the multiplicity is two and the point (α^2, α^3) where the multiplicity is one. However if $\alpha = 0$, we find

$$(y - \alpha x, y^2 - x^3 = (y, x^2 - x^3) = (y, x^3),$$

so that in this case $A/(f,g) = k[x]/(x^3)$ whose support is concentrated in the origin, and its multiplicity there is three.

☆

Exercises

(9.29) Let $A = k[X,Y]/(X^aY^b)$ where a and b are natural numbers, and let x and y stand for the images of X and Y in A. Show that the equality $(0) = (x^a) \cap (y^b)$ holds

in A and that (x) and (y) are the minimal primes of A. Show that $A_{(x)} = K[X]/(X^a)$ where $K = k(Y)$ and that this ring is of length a.

(9.30) Let n and m be two natural numbers and let $\mathfrak{a}$ be the ideal $\mathfrak{a} = (x^m, y^n)$ in $k[x, y]$. Show that $A = k[x, y]/\mathfrak{a}$ is Artinian and compute its length.

(9.31) Let n, m and r be three natural numbers and let $A = k[x, y, z]/(x^n, y^m, z^r)$. Prove that A is Artinian and compute its length.

(9.32) Show that if $A = k[x, y]/\mathfrak{a}$ is of length two, then after a linear change of coordinates, $A = k[x, y]/(x, y^2)$. If A is of length three, show that either $A = k[x, y]/(x, y)^2$ or there is a linear change of coordinates such that $A = k[x, y]/(x, y^3)$.

(9.33) Consider the direct product $R = \prod_i \mathbb{Z}/2\mathbb{Z}$ of countably many copies of $\mathbb{Z}/2\mathbb{Z}$.

a) Prove that R is not Noetherian by showing that the sets $\mathfrak{a}_r = \{ (x_i) \mid x_i = 0 \text{ for } i \geqslant r \}$ (*i. e.* the sets of strings with a zero tail of given length) form an ascending chain of ideals which is not stationary.

b) Show that every element in R is idempotent.

c) Let $\mathfrak{m}$ be a maximal ideal in R. Show that the localization $R_\mathfrak{m}$ satisfies $R_\mathfrak{m} \simeq \mathbb{Z}/2\mathbb{Z}$. Conclude that all prime ideals in R are maximal. HINT: The only idempotents in a local ring are 0 and 1.

d) Show that Spec R is a compact Hausdorff space. Conclude that the Zariski topology is not discrete.

e) Sow that the direct sum $D = \bigoplus_i \mathbb{Z}/2\mathbb{Z}$ is an ideal in R and that Spec R is the disjoint union of Supp D and Supp R/D.

Lecture 10

Primary decomposition

The story about primary decomposition originates in the the Fundamental Theorem of Arithmetic. Recall that this primordial theorem states that any natural number can be expressed as a product of prime numbers whose factors are unambiguously determined up to order. The early 19^{th} century mathematicians when beginning to explore the algebraic number fields discovered before long that the integers in such fields do not share this property unconditionally; there are rings of algebraic integers for which the analogue of the Fundamental Theorem does not hold; the factors are not always unique. We already saw examples of this phenomenon in Chapter 3.

However, for a large class of rings ubiquitous in algebraic number theory—the so-called Dedekind domains—the situation could be saved by using prime ideals instead of prime numbers; any non-zero and proper ideal in a Dedekind domain is a product of powers of prime ideals in an unambiguous way (up to order as usual).

Dedekind domains, even though they are ubiquitous in number theory, are rather special rings, and the question arose quickly what is generally true. Emanuel Lasker* was one of the first to give a partial answer; he established primary decomposition for ideals in rings finitely generated over fields. The final breakthrough came with Emmy Noether's famous 1921-paper [**?**]. Her results were profoundly more general and her proofs enormously easier and more translucent than the previous. In the formulation of the general decomposition theorem—valid for Noetherian rings—products are replaced by intersections and powers of prime ideals by so-called *primary ideals*. Every ideal $\mathfrak{a}$ in a Noetherian ring has a such a primary decomposition: one may express $\mathfrak{a}$ as a finite intersection $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r$ of primary ideals. The uniqueness of the intervening ideals however, is only partially true—there are so-called *embedded components* around that disturb the picture.

We have already seen an important instance of the theorem. With Corollary 9.18 on page 235 we established that radical ideals in Noetherian rings are intersections of finitely many prime ideals, and the involved prime ideals are unique when the

**In addition to be an eminent mathematician, Lasker was World Chess Champion for 27 years.*

Emanuel Lasker (1868–1941)

German mathematician

intersection is irredundant. The case of ideals not being radical is rather more complex, prime ideals will not be sufficient, and the primary ideals enter the story.

Primary decompositions also have a highly significant geometric aspect. In a geometric language a primary decomposition of an ideal $\mathfrak{a}$ corresponds to a decomposition of $V(\mathfrak{a})$ into a union of closed, irreducible* subsets called the *irreducible components* of X. For instance, the subset given by $xyz = 0$ in $\mathbb{C}^3$ has the three coordinate planes as irreducible components. Not that since $V(\mathfrak{a}) = V(\sqrt{\mathfrak{a}})$, the topology does not capture the full primary decomposition, but only the representation of $\sqrt{\mathfrak{a}}$ as an intersection of the minimal primes.

*A topological space is irreducible *if it is not the union of two proper, closed subsets.*

Irreducible spaces (irreduktibele rom)
Irreducible components (irreduktible komponenter)

The road map of this chapter is as follows: we begin with introducing the new important players, the primary ideals, and establish their basic properties. Next follows the announcement of the main existence theorem—the Lasker-Noether theorem— a discussion around it and its proof, and the two uniqueness theorems are proven.

Homogeneous ideals are in widespread use in algebraic geometry, and an important fact is that their primary decompositions may be chosen to stay within the realm of homogeneous ideals, and we have included a proof of that. The same holds for monomial ideals, and as well for ideals invariant under certain other groups, which we treat in an appendix.

10.1 Primary ideals

As alluded to in the introduction to this chapter, one needs a notion of *primary ideals* that generalizes the notion of "prime powers" for integers. The naive try would be just to use powers of prime ideals, but this turns out to be far too simple—the issue is of a much subtler character.

(10.1) The property of an ideal $\mathfrak{q}$ being primary is best introduced as a property of the quotient $A/\mathfrak{q}$. To motivate the definition, let p be a prime number and n any integer, and consider the multiplication-by-n-map $\mathbb{Z}/p^r\mathbb{Z} \to \mathbb{Z}/p^r\mathbb{Z}$. It is either bijective or nilpotent*, and the important point here is that this characterizes prime powers; indeed, if $m = rp^s$ with p relatively prime to r and $r > 1$, it holds that $\mathbb{Z}/m\mathbb{Z} \simeq \mathbb{Z}/p^s\mathbb{Z} \times \mathbb{Z}/r\mathbb{Z}$ and multiplication by p will be injective on one factor and nilpotent on the other, so it is neither injetive nor nilpotent.

When n is prime to p it will be bijective, and when n has p as factor, it will be nilpotent.

Inspired by this exquisite property of prime powers, we call a proper* ideal $\mathfrak{q}$ a *primary ideal* if the following condition is satisfied:

As is the case for prime ideals, we insist on primary ideals being proper.

Primary ideals (primæridealer)

- For every element $x \in A$ the multiplication map $A/\mathfrak{q} \to A/\mathfrak{q}$ that sends an element y to $x \cdot y$ is either injective or nilpotent.

The "multiplication-by-x-map" is frequently called by the more scientific name "the *homothety* by x". Since the radical $\sqrt{\mathfrak{q}}$ consists of ring elements with a power lying in $\mathfrak{q}$; that is, those inducing nilpotent homotheties, the condition may be reformulated as:

Homotheties (homotetier)

- ❑ If $xy \in \mathfrak{q}$, then either $y \in \mathfrak{q}$ or $x \in \sqrt{\mathfrak{q}}$.

EXAMPLE 10.1 If f is an irreducible element in a unique factorization domain, then the principal ideals (f^n) generated by a power of f are primary: if $xy = zf^n$ and f^n does not divide y, it follows that f divides x, and so $x \in \sqrt{(f^n)} = (f)$. ☆

Basic properties of primary ideals

Here we turn to some of the basic properties of primary ideals. Their radicals are prime ideals (as would be expected from offshoots of prime powers), and they behave well with respect to intersections, localizations and the formation of quotient. Even though having a prime radical is a primordial property of primary ideals, the notions are not equivalent—the analogy with prime powers must not be pushed too far.

(10.2) The first property we shall discuss is that primary ideals have radicals which are prime. Once this is established, it follows that the radical $\sqrt{\mathfrak{q}}$ is the smallest among the prime ideals containing $\mathfrak{q}$. Indeed, as for any ideal, the radical of $\mathfrak{q}$ is contained in each prime ideal that contains $\mathfrak{q}$ In other words, the ring $A/\mathfrak{q}$ has just one minimal prime ideal.

PROPOSITION 10.3 *If $\mathfrak{q}$ is a primary ideal in the ring A, the radical $\sqrt{\mathfrak{q}}$ is a prime ideal, and it is the smallest prime ideal containing $\mathfrak{q}$.*

PROOF: Assume that $xy \in \sqrt{\mathfrak{q}}$, but $y \notin \sqrt{\mathfrak{q}}$; then $x^n y^n$ lies in $\mathfrak{q}$ for some n, but $y^n \notin \mathfrak{q}$, so some power of x^n lies there. Hence $x \in \sqrt{\mathfrak{q}}$. ❑

It is customary to say that a primary ideal $\mathfrak{q}$ is $\mathfrak{p}$-*primary* when $\mathfrak{p} = \sqrt{\mathfrak{q}}$, which also is phrased as $\mathfrak{p}$ *belongs to* $\mathfrak{q}$. The converse of Proposition 10.3 does not hold in general; the radical being prime is not sufficient for an ideal to be primary. Example 10.4 below is an easy and concrete instance of this. Even powers of prime ideals need not be primary. Example 10.5 below is the standard example of this phenomenon, and a more elaborate example in a polynomial ring is found in Exercise 10.3 below. However, if the radical of $\mathfrak{q}$ is maximal, $\mathfrak{q}$ will be primary:

$\mathfrak{p}$-primary ideals ($\mathfrak{p}$-primære idealer)

PROPOSITION 10.4 *An ideal $\mathfrak{q}$ whose radical is maximal, is primary.*

PROOF: Assume that the radical $\sqrt{\mathfrak{q}}$ is maximal and write $\mathfrak{m} = \sqrt{\mathfrak{q}}$. Because $\mathfrak{m}$ is both maximal and the smallest prime containing $\mathfrak{q}$, the ring $A/\mathfrak{q}$ is a local ring with maximal ideal $\mathfrak{m}/\mathfrak{q}$ as the only prime ideal. Therefore the elements of $\mathfrak{m}/\mathfrak{q}$ are nilpotent while those not in $\mathfrak{m}/\mathfrak{q}$ are invertible. ❑

COROLLARY 10.5 *The powers $\mathfrak{m}^n$ of a maximal ideal $\mathfrak{m}$ are $\mathfrak{m}$-primary.*

PROOF: The radical of $\mathfrak{m}^n$ equals $\mathfrak{m}$. ❑

LEMMA 10.6 *Assume that $B \subseteq A$ is an extension of rings and that $\mathfrak{q}$ is a $\mathfrak{p}$-primary ideal in A. Then $\mathfrak{q} \cap B$ is a $\mathfrak{p} \cap B$-primary ideal in B.*

PROOF: Since $B/\mathfrak{q} \cap B \subseteq A/\mathfrak{q}$, the multiplication by x in $B/\mathfrak{q} \cap B$ is either injective or nilpotent since this holds for multiplication by x in $A/\mathfrak{q}A$. That $(\sqrt{\mathfrak{q}}) \cap B = \sqrt{(\mathfrak{q} \cap B)}$ is trivial. ❑

Examples

Many examples in this section will be monomial ideals, and later on we shall dedicate a special subsection to them (Subsection 10.5 on page 279). The method described there, may be applied to several of the precent examples, but at this stage of the course we are confined to using *ad hoc* methods.

(10.2) The ideal $(y, x)^2$ is a primary ideal in the polynomials ring $k[x, y, z]$. To see this, make $k[x, y, z]$ a graded ring by assigning the weights $\deg x = \deg y = 1$ and $\deg z = 0$ to the variables. The ideal $(x, y)^2 = (x^2, xy, y^2)$ is formed by the polynomials all whose homogenous components are of degree at least two. Consider now a polynomial f, with expansion $f = f_0 + f_1 + \dots$ in homogeneous components, and assume it does not belong to $\sqrt{(x, y)^2} = (x, y)$. This means that $f_0 \neq 0$. Let g be another polynomial whose expansion in homogeneous components is $g = g_s + g_{s+1} + \dots$, and assume that $fg \in (x, y)^2$. We find that $fg = f_0 g_s +$ heigher terms, and so as $f_0 g_s \neq 0$, we infer that $s \geqslant 2$; that is, $g \in (x, y)^2$.

(10.3) Let k be a field and $r \leqslant n$ two natural numbers. Each of the monomial ideals $(x_1^{\alpha_1}, x_2^{\alpha_2}, \dots, x_r^{\alpha_r})$ in the polynomial ring $k[x_1, \dots, x_n]$ is primary. A detailed proof will be given in Lemma 10.43 below. The powers $(x_1, \dots, x_r)^n$ are primary as well as you are asked to prove in Exercise 10.2 on page 264.

(10.4) The ideal $\mathfrak{a} = (x^2, xy)$ in the polynomial ring $k[x, y]$ has a radical that is prime, but $\mathfrak{a}$ is not primary. The radical of (x^2, xy) equals (x), which is prime, but in the quotient $k[x, y]/(x^2, xy)$ multiplication by y is neither injective nor nilpotent (y kills the class of x, but no power of y lies in (x^2, xy)). One decomposition of (x^2, xy) as an intersection of primary ideals is

$$(x^2, xy) = (x) \cap (x^2, y).$$

Checking the equality is not hard. One inclusion ($\subseteq$) is trivial, and the other holds since a relation $z = ax = bx^2 + cy$ implies that x divides c (the polynomial ring is UFD), and hence $z \in (x^2, xy)$. Notice that both ideals in the intersections are primary: (x) since it is prime and (x^2, y) because the radical equals (x, y) which is maximal. There

are also other decompositions of $\mathfrak{a}$ into intersections of primary ideals; for instance, it holds true that

$$(x^2, xy) = (x) \cap (x, y)^2.$$

Indeed, (x^2, xy) consists of the polynomials with x as factor that vanish at least to the second order at the origin. This exemplifies that primary decompositions are not unique in general.

(10.5) *The quadratic cone and a power of a prime that is not primary:* This is the standard example of a prime ideal whose square (or any of its powers, for that matter) is not primary; we already discussed it in a slightly different context (Example 7.15 on page 186).

It goes as follows: let $A = k[x, y, z]$ with constituting relation $z^2 - xy$. This is a graded ring (the relation is homogeneous) and the elements x, y and z form a basis for the degree one part (the relation is of degree two). The ideal $\mathfrak{p} = (z, x)$ is prime (kill it, and you get the polynomial ring $k[y]$), but $\mathfrak{p}^2$ is not primary; indeed, yx lies there being equal to z^2, but neither does x lie in $\mathfrak{p}^2$ (for degree reasons) nor does y lie in $\mathfrak{p}$ (the degree one part of $\mathfrak{p}$ has basis x and z).

One decomposition of $\mathfrak{p}^2$ into primary ideals is shaped like

$$(z, x)^2 = (z^2, zx, x^2) = (yx, zx, x^2) = (x, y, z)^2 \cap (x).$$

The ideal $(x, y, z)^2$ has the maximal ideal (x, y, z) as radical and is therefore primary. The ideal (x) is more interesting. Killing x, we obtain the ring $A/(x) = k[y, z]$ with constituting relation $z^2 = 0$, whose elements are either non-zero divisors or nilpotent*, and so (x) is a primary ideal. Its radical equals (z, x).

*The elements are of the form $a(y) + b(y)z$ with $a, b \in k[y]$, and one easily sees that this is a non-zero divisor unless $a = 0$, but then the square is zero.

The quadratic cone (den kvadratiske kjeglen)

The origin of the name *the quadratic cone* lies in the fact that the geometric locus C in $\mathbb{C}^3$ where $z^2 - xy = 0$ is a cone, which means that the line connecting the origin to any point on C lies in C: indeed, such a line is parameterized as (ta, tb, tc) where $(a, b, c) \in C$ is the point and t the running parameter, and since obviously $(tc)^2 - (ta)(tb) = t^2(c^2 - ab) = 0$, the line is contained in C. And the reason for *quadratic* in the name is simply because the equation of C is of the degree two.

☆

(10.7) The intersection of finitely many $\mathfrak{p}$-primary ideals persist being $\mathfrak{p}$-primary. In the analogy with the integers this reflects the simplistic fact that the greatest common divisor of some powers of the same prime number is a power of that prime.

PROPOSITION 10.8 *If $\{q_i\}$ is a finite collection of $\mathfrak{p}$-primary ideals, then the intersection $\bigcap_i \mathfrak{q}_i$ is $\mathfrak{p}$-primary.*

PROOF: Recall that the formation of radicals commutes with taking finite intersection (Lemma 2.62 on page 52), and therefore one has $\sqrt{\bigcap_i \mathfrak{q}_i} = \bigcap_i \sqrt{\mathfrak{q}_i} = \mathfrak{p}$. Assume next

that $xy \in \bigcap_i \mathfrak{q}_i$, but $y \notin \bigcap_i \mathfrak{q}_i$; that is, $xy \in \mathfrak{q}_i$ for each i, but $y \notin \mathfrak{q}_\nu$ for some ν. Since $\mathfrak{q}_\nu$ is $\mathfrak{p}$-primary x lies in the radical $\sqrt{\mathfrak{q}_\nu}$ of $\mathfrak{q}_\nu$, which equals $\mathfrak{p}$, but as we just checked, $\mathfrak{p}$ is as well the radical of the intersection $\bigcap_i \mathfrak{q}_i$. ❑

The hypothesis that the intersection be finite cannot be ignored. Powers $\mathfrak{m}^i$ of a maximal ideal are all primary and have the same radical, namley $\mathfrak{m}$, but at least when A is a Noetherian domain, their intersection equals (0) by Krull's Prinicipal Ideal Theorem; the zero ideal might be primary, but certainly not $\mathfrak{m}$-primary (in most cases). There are however instances when infinite intersections of $\mathfrak{p}$-primary ideals are $\mathfrak{p}$-primary (one is described in Exercise 10.14 below).

(10.9) The property of being primary is compatible with localizations, at least when these are performed with respect to multiplicative sets disjoint from the radical.

PROPOSITION 10.10 *Let S a multiplicative set in the ring A and let $\mathfrak{q}$ be a $\mathfrak{p}$-primary ideal. Assume that $S \cap \mathfrak{p} = \varnothing$. Then $S^{-1}\mathfrak{q}$ is $S^{-1}\mathfrak{p}$-primary, and it holds true that $\iota_S^{-1}(S^{-1}\mathfrak{q}) = \mathfrak{q}$.*

PROOF: Localizing commutes with forming radicals (Proposition 7.23 on page 183) so the radical of $S^{-1}\mathfrak{q}$ equals $S^{-1}\mathfrak{p}$. Assume that $x/s \cdot y/s' \in S^{-1}\mathfrak{q}$, but that $y/s' \notin S^{-1}\mathfrak{q}$. Then $txy \in \mathfrak{q}$ for some $t \in S$, and obviously it holds that $y \notin \mathfrak{q}$. Hence tx lies in the radical $\mathfrak{p}$ of $\mathfrak{q}$, and since $t \notin \mathfrak{p}$, we conclude that $x \in \mathfrak{p}$; in other words x/s lies in $S^{-1}\mathfrak{p}$.

To verify that $\iota_S^{-1}(S^{-1}\mathfrak{q}) = \mathfrak{q}$ let $x \in A$ be such that $\iota_S(x) \in S^{-1}\mathfrak{q}$. This means that $sx \in \mathfrak{q}$ for some $s \in S$, but by hypothesis $\mathfrak{p} \cap S = \varnothing$ so that $s \notin \mathfrak{p}$; thence $x \in \mathfrak{q}$ because $\mathfrak{q}$ is primary. ❑

(10.11) The third property we shall discuss permits us, when studying a given primary ideal $\mathfrak{q}$, to replace A by $A/\mathfrak{q}$ and $\mathfrak{q}$ by the zero ideal, which occasionally makes arguments cleaner and notationally simpler.

PROPOSITION 10.12 *Let A be a ring and $B = A/\mathfrak{a}$ where $\mathfrak{a}$ is an ideal in A. Assume that $\mathfrak{q}$ is an ideal in A containing $\mathfrak{a}$. Then the image $\mathfrak{q}B = \mathfrak{q}/\mathfrak{a}$ of $\mathfrak{q}$ in B is primary if and only if $\mathfrak{q}$ is. The radical of the image equals the image of the radical; or in symbols, $\sqrt{(\mathfrak{q}/\mathfrak{a})} = (\sqrt{\mathfrak{q}})/\mathfrak{a}$.*

PROOF: This is pretty obvious because by the Isomorphism Theorem (Theorem 2.21 on page 37) it holds that $A/\mathfrak{q} \simeq B/\mathfrak{q}B$, so the multiplication-by-what-ever-maps are the same. ❑

In particular, we observe that the ideal $\mathfrak{q}$ is primary if and only if the zero ideal (0) is a primary ideal in the quotient $A/\mathfrak{q}$.

Exercises

(10.1) With notation as in Example 10.5 on the previous page, show that $\mathfrak{p}^n$ is not primary for any $n \geqslant 2$. HINT: Show that $xy^{n-1} \in \mathfrak{p}^n$ but $y^{n-1} \notin \mathfrak{p}^n$.

⁂ (10.2) Let $\mathfrak{p}$ be the ideal in the polynomial ring $k[x_1, \ldots, x_n]$ over a field k generated by the r first variables; that is, $\mathfrak{p} = (x_1, \ldots, x_r)$. Show that every power $\mathfrak{p}^m$ is $\mathfrak{p}$-primary.

HINT: Consider $k[x_1, \dots, x_n]$ to be a polynomial ring over $A = k[x_{r+1}, \dots, x_n]$; it is naturally graded with A being the part of degree zero, and $\mathfrak{p}^m$ wil be the ideal of elements whose homogeneous components all are of degree at least equal to m.

(10.3) The example in 10.14 of a prime ideals whose powers are not primary, was an ideal in the coordinate ring of a cone. Such a phenomenon can take place even for ideals in a polynomial ring, and the present exercise (which we have borrowed from Melvin Hochster) illustrates this.

To facilitate the computations we give the polynomial ring $R = k[x, y, z]$ a grading by assigning degrees to the variables: $\deg x = 3$, $\deg y = 4$ and $\deg z = 5$.

Consider the map $\phi\colon k[x, y, z] \to k[t]$ defined by $x \mapsto t^3$, $y \mapsto t^4$ and $z \mapsto t^5$. It preserves the degrees (when $k[t]$ is given the usual grading with the degree of t being one) hence the kernel $\mathfrak{p}$ is a homogeneous prime ideal. The image of ϕ is the ring $k[t^3, t^4, t^5]$.

a) Show that the polynomials $f = xz - y^2$, $g = x^3 - yz$ and $h = yx^2 - z^2$ all lie in $\mathfrak{p}$.
b) Verify that the homogeneous polynomials of degree less than 8 in R are x, x^2 and xy, and prove that no element of $\mathfrak{p}$ is of degree less than 8.
c) Show that $g^2 - fh$ is divisible by x and that $u = (g^2 - fh)x^{-1}$ is homogeneous of degree 15.
d) Conclude that $\mathfrak{p}^2$ is not a primary ideal. HINT: $xu \in \mathfrak{p}^2$ but neither u nor any power of x lies in $\mathfrak{p}^2$.

★

10.2 *The Lasker-Noether theorem*

The road towards the final result about primary decomposition has two stages. The existence of a minimal decomposition is one, and the uniqueness, or partly uniqueness as one rather should say, is another. This bifurcation was already apparent when we factored elements in UFD's; every element in a Noetherian ring is a finite product of irreducible elements, but only factorizations into prime elements are unique (up to order and units). In this section we shall accomplish the first stage and establish the existence of minimal primary decompositions of ideals in Noetherian rings. The result is named after Emanuel Lasker and Emmy Noether.

Minimal primary decompositions

(10.13) Given a collection $\{S_i\}$ of set. It might very well happen that the intersection $\bigcap_i S_i$ does not change if one throws away one or more of the S_i's (for instance, if $S_1 \subseteq S_2$, one stupidly has $S_1 \cap S_2 = S_1$), and in that case one says that intersection is *redundant*. In the opposite case, that all the S_i contribute to the intersection, or in other words, when $\bigcap_i S_i \nsubseteq \bigcap_{i \neq j} S_i$ for all j, the intersection is called *irredundant*.

Redundant intersections (redundant snitt)

Irredundant intersections (irreduntante snitt)

The superfluous sets of an intersection are precisely those S_j such that $\bigcap_{i\neq j} S_i \subseteq S_j$, and one may render the intersection irredundant by just discarding them.

Primary decomposition (primær dekomposisjon)

(10.14) Now, let $\mathfrak{a}$ be an ideal in the ring A. A *primary decomposition* of $\mathfrak{a}$ is an expression of $\mathfrak{a}$ as a finite intersection of primary ideals; that is, an equality like

$$\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r \tag{10.1}$$

where the $\mathfrak{q}_i$'s are primary ideals. We have already seen a few examples of such decompositions (Examples 10.4 and 10.5 above).

**And it can be in non-trivial ways too; we shortly return to those.*

Without further constraints there are several trivial* ways such a decomposition can be ambiguous. First of all, it can be a redundant intersection. Secondly, a $\mathfrak{p}$-primary ideal can be the intersections of other $\mathfrak{p}$-primary ideals, sometimes even in infinitely many different ways (there is an upcoming Example 10.6). The first type of ambiguity is coped with by just discarding superfluous ideals, and Proposition 10.8 above helps us coping with the second. We just group those $\mathfrak{q}_i$'s with the same radical together and replace them by their intersection, which will be primary and will have the same radical.

Minimal or reduced primary decompositions (minimale eller reduserte primærdekomposisjoner)

The primary decomposition (10.1) is called *minimal* or *reduced* if all the radicals $\sqrt{\mathfrak{q}_i}$ are different and the intersection is irredundant. We have proven:

LEMMA 10.15 *Any primary decomposition $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r$ can be rendered a minimal one; that is, an irredundant intersection with the radicals $\sqrt{\mathfrak{q}_i}$ being distinct.*

EXAMPLE 10.6 Consider $\mathfrak{m}^2 = (x^2, xy, y^2)$ in the polynomial ring $k[x, y]$ (where $\mathfrak{m} = (x, y)$). For all scalars α and β with $\alpha \neq 0$ one has the equality

$$\mathfrak{m}^2 = (x^2, xy, y^2) = (x^2, y) \cap (y^2, \alpha x + \beta y).$$

Indeed, one easily checks that $\mathfrak{m}^2 \subseteq (x^2, y) \cap (y^2, \alpha x + \beta y)$ (since $\alpha \neq 0$), and the other inclusion amounts to the two lines generated by the class of y and the class of $\alpha x + \beta y$ in the two dimensional vector space $\mathfrak{m}/\mathfrak{m}^2$ being distinct so that their intersection is reduced to the origin. ☆

EXAMPLE 10.7 If the ring A is a PID, there is nothing much new. The prime ideals are the principal ideals (p) generated by an irreducible p. The (p)-primary ideals are those generated by powers of p; that is, those on the form (p^v). In general, if $f = p_1^{v_1} \ldots p_r^{v_r}$ is a factorisation of f into a product of irreducible elements, the primary decomposition of (f) is unambiguous and it is given as

$$(f) = (p_1)^{v_1} \cap \ldots \cap (p_r)^{v_r}.$$

The same applies to *principal* ideals in any UFD, where irreducible elements are prime. ☆

(10.16) Finally in this paragraph, we notice that as a direct consequence of Proposition 10.10 on page 264 primary decompositions localize well:

PROPOSITION 10.17 *Assume that S is a multiplicatively closed subset of the ring A and that $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r$ is a primary decomposition of an ideal $\mathfrak{a}$ and denote the radical of $\mathfrak{q}_i$ by $\mathfrak{p}_i$. Then it holds true that $S^{-1}\mathfrak{a} = S^{-1}\mathfrak{q}_1 \cap \ldots \cap S^{-1}\mathfrak{q}_r$. Moreover, either $S^{-1}\mathfrak{q}_i$ is primary with radical $S^{-1}\mathfrak{p}_i$ or $S^{-1}\mathfrak{q}_i = S^{-1}A$.*

The resulting decomposition of $S^{-1}\mathfrak{a}$ is not always irredundant even if the one one starts with is. The primes $\mathfrak{p}_i$ meeting S blow up to the entire ring $S^{-1}A$ and will not contribute to the intersection; they can thus be discarded, and one may write

$$S^{-1}\mathfrak{a} = \bigcap_{S\cap\mathfrak{p}_i=\varnothing} S^{-1}\mathfrak{q}_i.$$

A particularly interesting case arises when one takes S to be the complement of one of the $\mathfrak{p}_i$'s, say $\mathfrak{p}_\nu$. Then* $\mathfrak{a}A_{\mathfrak{p}_\nu} = \mathfrak{q}_\nu A_{\mathfrak{p}_\nu}$, and $\mathfrak{a}A_{\mathfrak{p}_\nu}$ is a *primary* ideal in $A_{\mathfrak{p}_\nu}$.

*Recall the notation $A_\mathfrak{p}$ for A localised at the prime ideal $\mathfrak{p}$, that is in the multiplicative set $A\backslash\mathfrak{p}$.

Existence of Primary Decompositions

In rings that are not Noetherian, ideals may or may not have a finite primary decomposition, but in Noetherian rings they always have one. The proof is an application of a technique called Noetherian induction (the principle of assailing a "maximal crook").

PROPOSITION 10.18 *In a Noetherian ring A each ideal $\mathfrak{a}$ is the intersection of finitely many primary ideals.*

PROOF: Since the ring A is assumed to be Noetherian, the set of ideals for which the conclusion fails, if non-empty, has a maximal element $\mathfrak{a}$ (the "maximal crook"). Replacing A by $A/\mathfrak{a}$ we may assume that the zero ideal is the only crook, and aim for a contradiction. So we assume (0) is not the intersection of finitely many primary ideals in A (in particular it is not primary), but that all non-zero ideals are.

Because (0) is not primary, there will be two elements x and y in A with $xy = 0$, but with $x \neq 0$ and y not nilpotent. The different annihilators $\operatorname{Ann} y^i$ form an ascending chain of ideals, and hence $\operatorname{Ann} y^\nu = \operatorname{Ann} y^{\nu+1}$ for some ν. We contend that $(0) = \operatorname{Ann} y \cap (y^\nu)$. Indeed, if $a = by^\nu$ is an element in (y^ν) that lies in $\operatorname{Ann} y$, one has $ay = by^{\nu+1} = 0$, therefore $b \in \operatorname{Ann} y^{\nu+1} = \operatorname{Ann} y^\nu$, and it follows that $a = by^\nu = 0$. Now, $x \in \operatorname{Ann} y$ is a non-zero element, and since y is not nilpotent, both ideals (y^ν) and $\operatorname{Ann} y$ are non-zero, and both are therefore finite intersections of primary ideals. The same then obviously holds true for (0), and thus the zero ideal (0) is not crooked, contradicting the assumption it were. ❑

THEOREM 10.19 (THE LASKER-NOETHER THEOREM) *Every ideal in a Noetherian ring has a minimal primary decomposition.*

PROOF: Start with any decomposition of an ideal $\mathfrak{a}$ into primary ideals (there is at least one according to the proposition above). By Lemma 10.15 on page 266 it can be made minimal by regrouping ideals with the same radical and discarding redundant ones. ❑

10.3 The Uniqueness Theorems

There are two main uniqueness issues concerning primary decompositions. One may ask if the radicals of the components are unique, or one may ask if the primary components themselves are unique. The first question is answered by an unconditionally yes (even if the ring is not Noetherian), but to the second the answer is no in general (also in Noetherian rings), although partially being a yes. We shall exclusively work with Noetherian rings. The proofs basically go through also in the non-Noetherian case, but they are not as clean cut as for Noetherian rings, and we refrain from that exercise (see chapter four in Atiyah and Macdonald's book [**?**] if you are interested). But of course, the statements apply only to ideals that have a finite primary decomposition.

The First Uniqueness Theorem

(10.20) The way to show that the radicals of the primary components are invariants of an ideal, is to characterize them without referring to the decomposition. The idea is to consider the collection of *transporter ideals* $(\mathfrak{a} : x)$ when x varies in A, and it turns out that in the Noetherian case the radicals of the primary components of $\mathfrak{a}$ are precisely the prime ideals among these.

PROPOSITION 10.21 *Let $\mathfrak{a}$ be an ideal in a Noetherian ring A. The radicals that occur in an minimal primary decomposition of $\mathfrak{a}$, are precisely those transporter ideals $(\mathfrak{a} : x)$ with x in A that are prime.*

Passing to the quotient $A/\mathfrak{a}$ and observing that the quotient $(\mathfrak{a} : x)/\mathfrak{a}$ equals the annihilator $(0 : [x])$ of the class $[x]$ in $A/\mathfrak{a}$, one may give the theorem the equivalent formulation (remember Proposition 10.12 on page 264), which is the one we shall prove:

PROPOSITION 10.22 (PRINCIPLE OF ANNIHILATORS) *The radicals arising from a minmal primary decomposition of the zero ideal in a Noetherian ring are precisely those ideals among the annihilators* $\operatorname{Ann} x$ *of elements x from the ring that are prime.*

PROOF: Fix a minimal primary decomposition of the zero ideal (0). There are two implications to prove. We begin with letting $\mathfrak{q}$ be one of the components, and letting $\mathfrak{p} = \sqrt{\mathfrak{q}}$ denote the radical we aim at exhibiting an element x such that $\mathfrak{p} = \operatorname{Ann} x$.

Denote by $\mathfrak{c}$ the intersection of the components in the decomposition other than $\mathfrak{q}$. Then $\mathfrak{c} \cap \mathfrak{q} = 0$, and $\mathfrak{c} \neq 0$ since the decomposition is irredundant.

Let $x \in \mathfrak{c}$ be a non-zero element such $\operatorname{Ann} x$ is maximal among the annihilators of non-zero elements of $\mathfrak{c}$. We contend that $\operatorname{Ann} x = \mathfrak{p}$, and begin with showing the inclusion $\operatorname{Ann} x \subseteq \mathfrak{p}$. Because $x \neq 0$, it holds that $x \notin \mathfrak{q}$, and hence $xy = 0$ implies that $y \in \mathfrak{p}$ as $\mathfrak{q}$ is $\mathfrak{p}$-primary.

In order to show the other inclusion pick a $y \in \mathfrak{p}$ and assume that $xy \neq 0$. Some power of y lies in $\mathfrak{q}$ and therefore kills x. Hence there is a natural number n so that $y^n x = 0$, but $y^{n-1}x \neq 0$. By the maximality of $\operatorname{Ann} x$ it holds true that $\operatorname{Ann} x = \operatorname{Ann} y^{n-1}x$, and consequently $y \in \operatorname{Ann} x$, which contradicts the assumption that $xy \neq 0$.

For the reverse implication, assume that $\operatorname{Ann} x$ is a prime ideal. Let I be the set of indices such that $\mathfrak{q}_i$ does not contain x. Then $\bigcap_{i\in I} \mathfrak{q}_i \subseteq \operatorname{Ann} x$ because

$$x \cdot \bigcap_{i\in I} \mathfrak{q}_i \subseteq \bigcap_{i\notin I} \mathfrak{q}_i \cdot \bigcap_{i\in I} \mathfrak{q}_i \subseteq \bigcap_{i\notin I} \mathfrak{q}_i \cap \bigcap_{i\in I} \mathfrak{q}_i = (0).$$

Consequently it holds true that the product of appropriate powers of the corresponding radicals $\mathfrak{p}_i$ is contained in $\operatorname{Ann} x$. Since $\operatorname{Ann} x$ is supposed to be prime, it follows that $\mathfrak{p}_\nu \subseteq \operatorname{Ann} x$ for at least one $\nu \in I$. On the other hand, it holds true that $(0) = x \cdot \operatorname{Ann} x \subseteq \mathfrak{q}_\nu$ from which ensues that $\operatorname{Ann} x \subseteq \mathfrak{p}_\nu$ because $\mathfrak{q}_\nu$ is $\mathfrak{p}_\nu$-primary and $x \notin \mathfrak{p}_\nu$. ❑

As a corollary we arrive at the first uniqueness theorem:

THEOREM 10.23 (THE FIRST UNIQUENESS THEOREM) *The radicals occurring in an minimal primary decomposition of an ideal in a Noetherian ring are unambiguously determined by the ideal.*

Isolated and embedded components

(10.24) The radicals of the primary components are of course tightly related to the ideal, vaguely analogous to the prime factors of an integer, and they merit a proper name. They are called the *associated prime ideals* of $\mathfrak{a}$, and the set they constitute is denoted by $\operatorname{Ass} A/\mathfrak{a}$. In particular, $\operatorname{Ass} A$ will be the set of prime ideals associated to zero. The term is certainly ambiguous, having a different meaning when used for modules, but it is practical, and the two interpretations are closely related: the associate primes of $\mathfrak{a}$ are the associate primes of the A-module $A/\mathfrak{a}$ (and the notation $\operatorname{Ass} A/\mathfrak{a}$ is consistent).

Associated prime ideals (assosierte primidealer)

(10.25) There are no inclusion relations between the components of an irredundant primary decomposition (irredundance means precisely this), but that does not exclude inclusion relations between the associated primes. In Example 10.4, for instance, we found that $(x^2, xy) = (x) \cap (x, y)^2$ with the associated primes being (x) and (x, y). This leads us to distinguish between *isolated* and *embedded* associated primes. The former are

Embedded associated prime (embeddede assosierte primidealer)

Isolated associated primes (isolerte assosierte primidealer)

those being minimal in Ass A; that is, they do not contain any other associated prime, whereas the latter are those that do. In the example above, (x) is an isolated prime whilst (x, y) is embedded*.

Primary components with an isolated radical are called *isolated components* and those with an embedded radical are called *embedded components*.

Isolated components (isolerte komponenter)

Embedded components (embeddede komponenter)

(10.26) Early in the course, when discussing the radical of an ideal, we proved that the radical $\sqrt{0}$ of A equals the intersection of all minimal primes in A (Paragraph 2.59 on page 2.59); that is, $\sqrt{0} = \bigcap \mathfrak{p}$, the intersection extending over the minimal elements of Spec A.

On the other hand, we just expressed the radical $\sqrt{0}$ as the intersection of the prime ideals minimal in Ass A so that $\sqrt{0} = \bigcap \mathfrak{p}$ where the intersection extends over the minimal elements in Ass A. When the intersections of two finite families of prime ideals are equal and there are no inclusion relations between members of either family, the families coincide (Lemma 2.36 on page 43). Hence the sets Spec A and Ass A have the same minimal primes. We have proved:

PROPOSITION 10.27 *In a Noetherian ring A the sets* Spec A *and* Ass A *have the same minimal elements; in other words, the minimal primes of A are precisely the isolated associated primes. In particular, there are finitely many minimal primes.*

*You might be puzzled by the notion embedded components since they are not contained in, but on the contrary contain other associated primes. The usage comes from geometry since inclusions between varieties are the reverse of those between ideals.

(10.28) We have seen the intersection of the the associated primes of A is the set of nilpotent elements in A, and their union turns out to play a particular role at least in Noetherian rings, it equals the set of zero divisors:

PROPOSITION 10.29 *The set of zero-divisors in a Noetherian ring A equals $\bigcup_{\mathfrak{p}\in \operatorname{Ass} A} \mathfrak{p}$, the union of the associated primes.*

PROOF: Let $\operatorname{Ann} z$ be maximal among the annihilators of non-zero elements in A. Then $\operatorname{Ann} z$ is prime and hence an associated prime of A. Indeed, if $xyz = 0$ and $xz \neq 0$, it it ensues from the maximality of $\operatorname{Ann} z$ that $\operatorname{Ann} z = \operatorname{Ann} xz$ because obviously $\operatorname{Ann} z \subseteq \operatorname{Ann} xz$. Hence $y \in \operatorname{Ann} z$, and as any annihilator ideal is, it is contained in a maximal one, we are through. ❑

EXAMPLE 10.8 We offer one more example and consider the ideal

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

in the polynomial ring $\mathbb{C}[x, y, z]$ and aim at determining a primary decomposition.

To get an idea of where to start we resort to geometry, and take a look at the zero-locus $V(\mathfrak{a})$ inside $\mathbb{C}^3$. It is given by $x^2y = y^2z = z^2x = 0$, and is easily seen to be the union of the three coordinate-axes. This means that there must be some components supported along each axis, and no component can be supported elsewhere. So let us

consider the x-axis. Localizing at x, *i. e.* passing to $A_x = \mathbb{C}[x, x^{-1}, y, z]$ (thus eliminating the two other axes) we see that $\mathfrak{a}A_x = (y, y^2z, z^2) = (y, z^2)$; so we suspect this to be one of the components; by symmetry we find two more suspects, and in fact, it holds true that

$$(x^2y, y^2z, z^2x) \subseteq (y, z^2) \cap (x, y^2) \cap (z, x^2).$$

This is however not the whole story. The element xyz lies in the intersection to the right, but not in $\mathfrak{a}$. Now, clearly $(x, y, z) \subseteq (\mathfrak{a} : xyz)$ and (x, y, z) being maximal, it holds that $(\mathfrak{a} : xyz) = (x, y, z)$, so there must be an (x, y, z)-primary component. After a few tries (and a some failures) one finds the equality

$$(x^2y, y^2z, z^2x) = (y, z^2) \cap (x, y^2) \cap (z, x^2) \cap (x^2, y^2, z^2).$$

The associated primes are (z, y), (x, y), (z, x) and (x, y, z).

Once one has a good guess, it is relatively easy to check if it is correct. All involved ideals are generated by monomials, and monomial ideals have the nice property that a polynomial is a member if and only if all the monomial terms of the polynomial are. Hence it suffices to check that each monomial in the ideal to the right also lies in the one to the left. But monomials in (x^2, y^2, z^2) have either x^2, y^2 or z^2 as a factor, and by symmetry we may assume it is y^2. Lying in (z, x^2) too, our monomial must have either z or x^2 as a factor and thereby also zy^2 or x^2y^2; but both these lie in $\mathfrak{a}$, and we are done. ☆

The second uniqueness theorem

We now come to the uniqueness issue for the primary components. Already our first example (Exampe 10.4 on page 262) showed that they are not unique. We found that

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

and both (x^2, y) and $(x, y)^2$ are minimal primary components. Notice that they have the same radical (x, y) (they must!), and so they are embedded components. The bad news is that (x^2, xy) have infinitely many minimal primary decomposition with different embedded components (Example 10.9 below), but there is good news too: the other component is unique. This is generally true: the Second Uniqueness Theorem states that isolated components are unique. The reason for this is that isolated components may be retrieved from the ideal by localizing at the corresponding associated prime ideals, and according to the First Uniqueness Theorem these primes are independent of the decomposition; if $\mathfrak{q}$ is the component and $\sqrt{\mathfrak{q}} = \mathfrak{p}$, it holds true that $\mathfrak{q} = \iota^{-1}(\mathfrak{a}A_\mathfrak{p})$ where $\iota\colon A \to A_\mathfrak{p}$ is the localization map (or simply that $\mathfrak{q} = A \cap \mathfrak{a}A_\mathfrak{p}$ when ι is injective).

THEOREM 10.30 (THE SECOND UNIQUENESS THEOREM) *The isolated primary components of an ideal $\mathfrak{a}$ in a Noetherian ring A are unambiguously defined by the ideal.*

PROOF: We shall concentrate on one of the isolated associated prime ideals $\mathfrak{p}$ of $\mathfrak{a}$, but the main player will be a $\mathfrak{p}$-primary component $\mathfrak{q}$ from one of the minimal primary decompositions of $\mathfrak{a}$. The salient point is, as already announced, the equality

$$\mathfrak{q} = \iota^{-1}(\mathfrak{a}A_{\mathfrak{p}}), \tag{10.2}$$

from which the theorem ensues as isolated prime ideals are invariants of $\mathfrak{a}$.

To establish (10.2) one writes the decomposition of $\mathfrak{a}$ as $\mathfrak{a} = \mathfrak{q} \cap \bigcap_i \mathfrak{q}_i$ where the intersection extends over the primary components different from $\mathfrak{q}$. Localizing at $\mathfrak{p}$ one finds

$$\mathfrak{a}A_{\mathfrak{p}} = \mathfrak{q}A_{\mathfrak{p}} \cap \bigcap_i \mathfrak{q}_i A_{\mathfrak{p}} = \mathfrak{q}A_{\mathfrak{q}} \tag{10.3}$$

since the $\mathfrak{q}_i$'s blow up when localized; that is, $\mathfrak{q}_i A_{\mathfrak{p}} = A_{\mathfrak{p}}$. Indeed, since $\mathfrak{p}$ is isolated, $\mathfrak{p}_i \nsubseteq \mathfrak{p}$ holds for all i. Taking inverse images of both sides of (10.3) and citing Proposition 10.10 on page 264 we conclude that $\iota^{-1}(\mathfrak{a}A_{\mathfrak{p}}) = \mathfrak{q}$. ❑

Examples

(10.9) For any natural number n the equality

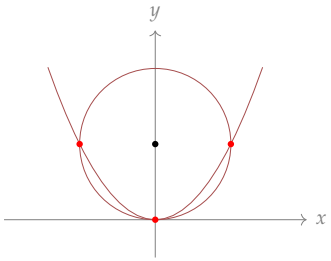
$$(x^2, xy) = (x) \cap (x^2, xy, y^n) \tag{10.4}$$

holds true in the polynomial ring $k[x, y]$, and this is an example of infinitely many different minimal primary decompositions of the same ideal. That the left side of (10.4) is included in the right is obvious; to check the other inclusion, let a belong to the right side. Then

$$a = p \cdot x = q \cdot x^2 + r \cdot y^n + sxy,$$

with p, q r and s polynomials in $k[x, y]$. It follows that x divides r, and hence that $a \in (x^2, xy)$.

(10.10) *Intersection of two conics:* So far all our examples have merely involved monomial ideals, but of course most ideals are not monomial. Primary decompositions are notoriously strenuous to lay hands on, and the monomial ideals are among the easiest to handle, hence their tendency to appear in texts. However, we are obliged to give at least one example of a more mainstream situation. It illustrates also that the decomposition is largely of a geometric nature; that is, at least the isolated associated prime ideals are; the primary components may conceal subtler structures—in this case, they conceal the tangency of two intersecting curves.



We shall analyse the familiar case of the intersection of two quadratic curves; the unit circle centred at $(0,1)$ and the standard parabola given by $y = x^2$. So let $\mathfrak{a} = (x^2 + (y-1)^2 - 1, y - x^2)$ in $k[x,y]$, where k is any field of characteristic different from 2. A standard manipulation shows that the common zeros of the two polynomials are the points $(1,1)$, $(-1,1)$ and $(0,0)$, and the same manipulations give

$$\mathfrak{a} = (x^2 + (y-1)^2 - 1, y - x^2) = (x^2(x^2-1), y - x^2).$$

Any prime ideal $\mathfrak{p}$ containing $\mathfrak{a}$ must contain either x, $x-1$ or $x+1$. It contains y if x lies in it, and because $y - x^2 = y - (x+1)(x-1) - 1$, one has $y - 1 \in \mathfrak{p}$ in the two other cases. We conclude that the (x,y), $(x-1,y-1)$ and $(x+1,y-1)$ are the only prime ideals containing $\mathfrak{a}$; and since they all three are maximal, the associated primes are found among them, and there can be no embedded component.

To determine the primary components of $\mathfrak{a}$, we localize (as in Theorem 10.30 on page 271). In the local ring $A = k[x,y]_{(x+1,y-1)}$, where both x and $x-1$ are invertible, we obtain the equality

$$\mathfrak{a}A = (x^2(x^2-1), y - x^2) = (x+1, y - x^2) = (x+1, y-1).$$

In similar fashion, in $B = \mathbb{C}[x,y]_{(x-1,y-1)}$ both x and $x+1$ are invertible, and one has

$$\mathfrak{a}B = (x^2(x^2-1), y - x^2) = (x-1, y-1).$$

Finally, in $C = k[x,y]_{(x,y)}$ both $x+1$ and $x-1$ have inverses, and we see that

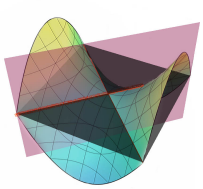
$$\mathfrak{a}C = (x^2, y).$$

Since there are no embedded components, we conclude that

$$\mathfrak{a} = (x-1, y-1) \cap (x+1, y-1) \cap (x^2, y).$$

When the characteristic of k equals two, things evolve in a slightly different manner. In that case, the two ideals $(x-1,y-1)$ and $(x+1,u-1)$ conicide and $x^2+1 = (x+1)^2$. We find

$$\mathfrak{a} = ((x+1)^2, y-1) \cap (x^2, y).$$



(10.11) *A saddle surface and the union of two planes:* Consider the intersection of the "saddle surface" S given in $\mathbb{C}^3$ by $z = xy$ and the union of the xy-plane and the xz-plane, which has the equation $yz = 0$.

The plane $z = 0$ intersects S in the union of the x-axis and the y-axis, and the plane $y = 0$ in the x-axis. The x-axis thus appears twice in the intersection, which algebraically is manifested by the occurrence of a non-prime primary component in the decomposition of the ideal $\mathfrak{a} = (z - xy, zy)$.

Because $zy = (z - xy)y + xy^2$, it holds that $\mathfrak{a} = (xy^2, z - xy)$, and consequently one finds

$$\mathfrak{a} = (xy^2, z - xy) = (x, z - xy) \cap (y^2, z - xy);$$

by observing that any prime containing $\mathfrak{a}$ contains either (x, z) or (y, z) so that (10.11) holds true when localized at any prime ideal, and hence equality persists since equality is a local property. Now, $(x, z - xy) = (x, z)$ is a prime ideal, and $(y^2, z - xy)$ is (y, z)-primary (for instance, since $k[x, y, z]/(y^2, z - xy) \simeq k[x, y]/(y^2)$), so we have found a primary decomposition of $\mathfrak{a}$.

☆

Exercises

✲ (10.4) Show that for any scalar $a \in k$ it holds true that

$$(x^2, xy) = (x) \cap (x^2, y + ax)$$

and that this is a minimal primary decomposition. Show that different scalars a give different decompositions.

✲ (10.5) Determine a minimal primary decomposition of $\mathfrak{a} = (x^3, y^2x^2, y^3x)$ in the polynomial ring $A = k[x, y]$.

✲ (10.6) Let $\mathfrak{a}$ be the ideal in the polynomial ring $A = k[x, y, z]$ given as $\mathfrak{a} = (yz, xz, xy)$. Show that the minimal primary decomposition of $\mathfrak{a}$ is shaped like

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

Show that the maximal ideal (x, y, z) is associated to the square $\mathfrak{a}^2$ and determine a minimal primary decomposition of $\mathfrak{a}^2$. HINT: Consider $(\mathfrak{a} : xyz)$.

(10.7) Determine the primary decomposition of the ideal (13) in the ring $\mathbb{Z}[i]$ of Gaussian integers.

(10.8) With the notation of Section 8.36 on page 220 about elliptic curves, consider the ring $A = k[x, y]$ with constituting relation $y^2 = p(x)$ where $p(x)$ is a monic cubic

polynomial with distinct roots. Determine the primary decomposition of the principal ideals $(x-a)$ and $(y-b)$ where $a, b \in k$.

(10.9) Let k be a field. Let $\mathfrak{p}_s = (x_1, \ldots, x_s)$ in $k[x_1, \ldots, x_n]$ be the ideal generated by the s first variables. Consider

$$\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3 \ldots \mathfrak{p}_r = (x_1)(x_1, x_2(x_1, x_3, x_3) \ldots (x_1, \ldots, x_r).$$

Prove that the following equality holds true

$$\mathfrak{a} = \mathfrak{p}_1 \cap \mathfrak{p}_2^2 \cap \ldots \cap \mathfrak{p}_{r-1}^{r-1} \cap \mathfrak{p}_r^r,$$

and that this is a minimal primary decomposition of $\mathfrak{a}$. HINT: Show that if $\mathfrak{c}$ is any ideal generated by monomials of degree $s-1$ in $x_1, \ldots, x_s$, then $\mathfrak{c}\mathfrak{p}_s = \mathfrak{c} \cap \mathfrak{p}_s^s$; then use induction on r.

(10.10) Concider the map $k[x, y, z] \to k[t, z]$ such that $x \mapsto t^2 - 1$ and $y \mapsto t(t^2-1)$ and $z \mapsto z$. Let $\mathfrak{p} = (z^2 - (x+1), y - zx)$. Determine the minimal primary decomposition of $\mathfrak{p}k[t, z]$.

(10.11) *Symbolic powers.* We have seen that the powers $\mathfrak{p}^n$ of a prime ideal $\mathfrak{p}$ in A not necessarily are $\mathfrak{p}$-primary, unless $\mathfrak{p}$ is maximal. But there are primary ideals canonical associated to the powers $\mathfrak{p}^n$; the so-called *symbolic powers* $\mathfrak{p}^{(n)}$. These arise in the following way: The ideal $\mathfrak{p}A_\mathfrak{p}$ is maximal in the local ring $A_\mathfrak{p}$ and its powers are therefore primary by Proposition 10.4 on page 261. Pulling primary ideals back along the localization map ι results in primary ideals (Proposition 10.10 on page 264). The ideal $\mathfrak{p}^{(n)} = \iota^{-1}(\mathfrak{p}^n A_\mathfrak{p})$ (or when ι is injective, $\mathfrak{p}^{(n)} = A \cap \mathfrak{p}^n A_\mathfrak{p}$) will therefore be primary, and this pullback is the n-th symbolic power of $\mathfrak{p}$.

Symbolic powers (symbolske potenser)

a) Show that if n and m are natural numbers, it holds that $\mathfrak{p}^{(n)} \cdot \mathfrak{p}^{(m)} \subseteq \mathfrak{p}^{n+m}$;
b) Show that $\mathfrak{p}^n = \mathfrak{p}^{(n)}$ if and only if $\mathfrak{p}^n$ has no embedded components;
c) With the notation as in Example 10.5 on page 263 determine the symbolic square $\mathfrak{p}^{(2)}$ of the ideal $\mathfrak{p} = (x, y)$.

★

Example: Reduced rings and a criterion of Serre

As an example of how a primary decomposition can be used, we give a criterion due to Serre for a Noetherian ring to be reduced which goes under name of the R_0–S_1–criterion.

The nil-radical $\sqrt{(0)}$ of a Noetherian ring A consists of the nilpotent elements in A, so one expects to be able to read off from the primary decomposition of the zero ideal whether A is reduced or not. The radical is expressed as the intersection

$$\sqrt{(0)} = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_r$$

of the finitely many minimal primes $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ of A. Now, A is reduced precisely when $\sqrt{(0)} = (0)$, and we may infer that A reduced if and only if the intersection of the

minimal primes equals (0):

$$(0) = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_r; \tag{10.5}$$

The equation (10.5) is a minimal primary decomposition of (0), so we conclude that A is reduced precisely when the primary components and the minimal primes of A are the same. In particular, A has no embedded components.

PROPOSITION 10.31 *A ring A is reduced if and only if it abides to the two requests that follows.*

i) For each minimal prime $A_\mathfrak{p}$ is a field;

ii) For each non-minimal prime ideal $\mathfrak{p}$, the local ring $A_\mathfrak{p}$ has no zero divisor.

The first requirement is usually is called R_0–condition and the second the S_1-condition (notions which will be perfectly logical when you have heard about regular rings, depth and height), hence the name the R_0–S_1–criterion.

PROOF: The first resquest is equivalent to all isolated primary components being prime since if $\mathfrak{q}$ is one of the sort and $\mathfrak{p} = \sqrt{\mathfrak{q}}$, the quotient $\mathfrak{p}/\mathfrak{q}$ equals the nil-radical of $A_\mathfrak{q}$, *i. e.* it is equivalent to $A_\mathfrak{p}$ being a field (see also proposition 7.29 on page 187 about the total ring of fractions of reduced rings). The second statement is equivalent to A being without embedded components; indeed, it may be reformulated as each annihilator $(0 : x)$ being contained in a minimal prime ideal. ❑

EXERCISE 10.12 Let A be a Noetherian ring without embedded components and let $\mathfrak{q}_1, \ldots, \mathfrak{q}_r$ be the isolated components of (0). Show that the total ring of fractions $K(A)$ of A is the product

$$K(A) = A_{\mathfrak{p}_1} \times \ldots \times A_{\mathfrak{p}_r},$$

where $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$ and where each $A_{\mathfrak{q}_i}$ is an Artinian local ring. ★

10.4 *The homogeneous case*

It is of great interest to know that the subsidiary ideals of a homogeneous ideal arising from a minimal primary decomposition are homogeneous. This not so much because most ideals one meets in examples or exercises are homogeneous, but because of the ubiquity of the so-called *projective varieties* in algebraic geometry—these are subvarieties of projective space defined by homogeneous ideals.

In what follows, we shall prove that associated primes and isolated components of homogeneous ideals are homogeneous. When it comes to the more elusive embedded components, all we can hope for is that they may be chosen to be homogeneous. Fortunately this is the case, and the Lasker-Noether Theorem is fully valid in a graded context.

(10.32) We start with a little lemma (in fact, we already met a version of it in Exercise 1.14 as early as on page 24).

LEMMA 10.33 *Let R be a graded ring and assume that x and y are two elements such that $xy = 0$. Let x_e be the homogeneous term of x of lowest degree. Then $x_e^r y = 0$ for some natural number r. In particular, $x_e^r y_i = 0$ for each homogenous component y_i of y.*

PROOF: Let $x = x_e + \ldots + x_n$ and $y = y_f + \ldots + y_m$ respectively be the expansions of x and y in homogeneous components. We shall show by induction that $x_e^{\nu+1} y_{f+\nu} = 0$ for all ν, and any r with $r \geqslant m - f + 1$ will then do. As $x_e y_f$ is the term of xy of lowest degree, it holds that $x_e y_f = 0$, and the induction can start. Expanding the product xy one finds for the homogeneous component of degree $e + f + \nu$ that

$$x_e \cdot y_{f+\nu} + x_{e+1} \cdot y_{f+\nu-1} + \ldots + x_{e+j} \cdot y_{f+\nu-j} + \ldots + x_{e+\nu} \cdot y_f = 0.$$

Multiplying through by x_e^ν gives $x_e^{\nu+1} \cdot y_{f+\nu} = 0$ because if $j \geqslant 1$, it holds true by induction that

$$x_e^\nu \cdot y_{f+\nu-j} = x_e^{j-1} \cdot x_e^{\nu-j+1} \cdot y_{f+\nu-j} = 0.$$

❑

(10.34) The first step in establishing the graded Noether–Lasker Theorem is to see that primes associated to homogeneous ideals are homogeneous.

PROPOSITION 10.35 *Let R be a graded ring and let $\mathfrak{p}$ be a prime ideal associated to the homogeneous ideal $\mathfrak{a}$. Then $\mathfrak{p}$ is homogeneous.*

PROOF: Replacing R by $R/\mathfrak{a}$, we may assume that $\mathfrak{a} = 0$. Let $\mathfrak{p}$ be a prime associated to R. Any associated prime is the annihilator of some element, and there is an $a \in R$ so that $\mathfrak{p} = (0 : a)$. Assuming that $x \in (0 : a)$ we have to show that each homogeneous components of x also lies in $(0 : a)$. If x_e is the homogeneous component of x of lowest degree, the lemma tells us that for some ν the power x_e^ν kills a . Hence $x_e^\nu \in (0 : a)$ and by consequence $x_e \in (0 : a)$ as $(0 : a)$ is a prime ideal. Noticing that $x - x_e$ also lies in $(0 : a)$, we finish the proof by the number of non-vanishing homogeneous terms in x. ❑

(10.36) The next step is to treat the primary components, and as alluded to above, embedded components are not unique and therefore not forced to be homogeneous, but they may be chosen homogeneous. For any ideal $\mathfrak{a}$ we shall denote by $\mathfrak{a}^\sharp$ the largest homogeneous ideal contained in $\mathfrak{a}$; that is, the ideal generated by all homogeneous elements belonging to $\mathfrak{a}$.

LEMMA 10.37 *Let $\mathfrak{q}$ be a primary ideal in the graded ring R whose radical is homogeneous. Then $\mathfrak{q}^\sharp$ is primary and $\sqrt{\mathfrak{q}^\sharp} = \sqrt{\mathfrak{q}}$.*

PROOF: We start out by proving that $\mathfrak{q}^\sharp$ has $\sqrt{\mathfrak{q}}$ as radical. Obviously $\sqrt{\mathfrak{q}^\sharp} \subseteq \sqrt{\mathfrak{q}}$. Attacking the other inclusion, we pick a member $x \in \sqrt{\mathfrak{q}}$. Since $\sqrt{\mathfrak{q}}$ is homogeneous all the homogeneous components of x lie in $\sqrt{\mathfrak{q}}$, and we may safely assume x to be

homogeneous. Now, $x^\nu \in \mathfrak{q}$ for some ν, and as any homogeneous member of $\mathfrak{q}$ belongs to $\mathfrak{q}^\sharp$, it follows for free that $x^\nu \in \mathfrak{q}^\sharp$, and we are through.

Next, suppose that $xy \in \mathfrak{q}^\sharp$ and that $y \notin \mathfrak{q}^\sharp$. The task is to show that $x \in \sqrt{\mathfrak{q}^\sharp}$, or in other words that $x \in \sqrt{\mathfrak{q}}$, since the two radicals coincide. If y_f is the homogeneous term of y of lowest degree, we may assume that $y_f \notin \mathfrak{q}^\sharp$, since if y_f lie in $\mathfrak{q}^\sharp$, we may replace y by $y - y_f$, and after repeating this procedure if needed, we shall end up with an element whose term of lowest degree do not belong to $\mathfrak{q}^\sharp$. With y_f well placed outside of $\mathfrak{q}^\sharp$, it follows that y_f is not a member of $\mathfrak{q}$ (all homogeneous elements of $\mathfrak{q}$ belongs to $\mathfrak{q}^\sharp$ by definition). Now, xy_f lies in $\mathfrak{q}^\sharp$ since all the homogeneous components of xy do, and therefore it lies in $\mathfrak{q}$. Hence $x \in \sqrt{\mathfrak{q}}$. ❏

(10.38) Finally we are prepared for the graded version of the Lasker–Noether theorem; needless to say, the two uniqueness theorem persist, they can of course be applied to any minimal primary decompositions.

PROPOSITION 10.39 (GRADED LASKER–NOETHER THEOREM) *A homogeneous ideal in a Noetherian graded ring has a minimal primary decomposition with all components being homogeneous, and all its associated primes are homogeneous.*

PROOF: Observe first that all the prime ideals associated to a homogeneous ideals are homogeneous (Proposition 10.35 above).

It is fairly clear that $(\mathfrak{a} \cap \mathfrak{b})^\sharp = \mathfrak{a}^\sharp \cap \mathfrak{b}^\sharp$ (the homogeneous elements in $\mathfrak{a} \cap \mathfrak{b}$ are the homogenous elements the lie in both $\mathfrak{a}$ and $\mathfrak{b}$) so starting out with a minimal primary decomposition $\mathfrak{a} = \bigcap_i \mathfrak{q}_i$ of a homogeneous ideal $\mathfrak{a}$ and applying the $\sharp$-construction to it, one arrives at a decomposition

$$\mathfrak{a} = \mathfrak{a}^\sharp = \bigcap_i \mathfrak{q}_i^\sharp, \tag{10.6}$$

and according to Lemma 10.37, this is a primary decomposition. Moreover, the radicals of the ideals $\mathfrak{q}_i^\sharp$ being the same as the radicals of the $\mathfrak{q}_i$'s, we can conclude that (10.6) is a minimal primary decomposition. ❏

A structure theorem for graded modules

This is a natural place to include a graded version of the structure theorem for fintiely generated modules (Theorem 9.23) we proved in Chapter 9, and the proof goes through with minor modifications.

(10.40) The changes are of two types. We want all intervening submodules to ge graded subsmodules and all prime ideals to be homogenoeus, and this ensured by Proposition 10.35 above. Secondly we want the chain and the ensuing maps in the ensuing short exact sequences to be maps in the category of graded modules, that is, we want the to preserve homogeneous elements and degrees. For this the shift operator

is needed. Recall that $M(d)$ has the same homogeneous elements as the graded module $M = \bigoplus_n M_n$, but the degrees are shifted: $(M(d))_n = M_{n+d}$. In short, we insist the claim being stated in the context of the category GrMod_R.

THEOREM 10.41 (STRUCTURE OF GRADED MODULES) *Let R a Noetherian graded ring and let M be a non-zero graded A module. Then M is finitely generated if and only if it possesses a finite ascending chain of graded submodules $\{M_i\}_{0\leqslant i\leqslant r}$ with $M_0 = 0$ and $M_r = M$, and such that all subquotients are shaped like $R/\mathfrak{p}_i(m_i)$ with $\mathfrak{p}_i$ a homogeneous prime ideal.*

In other words, there are short exact sequences in GrMod_R

$$0 \longrightarrow M_{i-1} \longrightarrow M_i \longrightarrow A/\mathfrak{p}_i(m_i) \longrightarrow 0$$

for $1 \leqslant i \leqslant r$.

PROOF: Assume M is finitely generated (the other implication is straightforward and left to the zealous students). Just as in the non-graded case, the proof is by Noetherian induction. The set of graded submodules for which the theorem is true is non-empty (zero submodule is there) and has a maximal element N since M is Noetherian. Assume that N is proper. The quotient M/N is then a non-zero graded module and has an associated homogeneous prime ideal $\mathfrak{p} = (0 : x)$ where x is homogenous. Shifting degrees by $m = -\deg x$, we obtain a map $R/\mathfrak{p}(m) \to M/N$ of graded modules, which induces the short exact sequence

$$0 \longrightarrow N \longrightarrow N' \longrightarrow R/\mathfrak{p}(m) \longrightarrow 0.$$

Here N' is the inverse image in M of $Rx \simeq R/\mathfrak{p}$, and the sequence shows that N' is strictly larger than N and has a chain of the required kind. ❑

10.5 *The case of monomial ideals*

Just as in the case of homogenous ideals, primary decompositions of monomial ideals stay within the family, so to say: associated primes and isolated components will automatically be monomial, and the embedded components can be chosen to be. We shall give a simple and constructive proof of this useful fact which also furnishes an easy algorithm that computes the decompositions. It is a sure but laborious way to a decomposition, but since there are relatively few and easily understood monomial prime ideals, there will often be smarter *ad hoc* approaches.

(10.42) The crux of the proof is Lemma 10.43 below, but before stating it we need a new notion: a monomial $f \in \mathfrak{a}$ is called *primitive* (relative to $\mathfrak{a}$) if no monomial that strictly divides f, belongs to $\mathfrak{a}$; in other words, if one lowers the power of a variable occurring in f, the resulting monomial does not belong to $\mathfrak{a}$. Every monomial ideals has a generating set consisting of primitive monomials.

Primitive monomials (primitive monomer)

Note that a monomial belongs to a monomial ideal if and only if it is a multiple of one of the primitive generators: indeed, monomials with different multi-exponents are linearly independent over k, and a monomial do not get lower multi-exponents when multiplied by a polynomial.

LEMMA 10.43 (THE JCO–ALGORITHM) *Let $\mathfrak{a}$ be a monomial ideal in the polynomial ring $A = k[x_1, \ldots, x_n]$ over the field k.*

i) If $\mathfrak{a}$ is generated by pure powers; that is, if for some subset $I \subseteq \{1, \ldots, n\}$ it holds true that $\mathfrak{a} = (x_i^{\alpha_i} | i \in I)$, then $\mathfrak{a}$ is primary;

ii) Let f be a primitive generator for $\mathfrak{a}$ and assume that $f = f_1 f_2$ where f_1 and f_2 are relatively prime monomials. Then it holds true that:

$$\mathfrak{a} = \big(\mathfrak{a} + (f_1)A\big) \cap \big(\mathfrak{a} + (f_2)A\big).$$

PROOF: Proof of *i)*: assuming that $xy \in \mathfrak{a}$ and $x \notin \mathfrak{a}$ we are to prove that a power of y belongs to $\mathfrak{a}$. From $x \notin \mathfrak{a}$, we infer that at least one of the monomial terms of x does not belong to $\mathfrak{a}$, and because $\mathfrak{a}$ is monomial, all the monomial terms of xy lies in $\mathfrak{a}$ since xy lies there. So when the claim is shown for monomials, we may conclude that a power of every monomials term of y lies in $\mathfrak{a}$, and a standard argument with the multinomial formula then yields that a high power of y lies there.

Assume then that both x and y are monomial. As x is a monomial not in $\mathfrak{a}$, at each one of the variables x_ν appears in x with an exponent less than α_ν, hence as y is a multiple of one of the generators, x_ν divides y for at least one index ν. Consequently y^{α_ν} belongs to $\mathfrak{a}$.

Proof of *ii)*: The ideals $\mathfrak{a} + (f_1)A$ and $\mathfrak{a} + (f_2)A$ are both monomial, so it suffices to prove that each monomial in the intersection belongs to $\mathfrak{a}$ (the other inclusion is obvious).

For a monomial x it holds for each i that $x \in \mathfrak{a} + (f_i)A$ if and only if x either lies in $\mathfrak{a}$ or is divisible by f_i. Consequently x lies in both the ideals $\mathfrak{a} + (f_i)A$ if and only if it either lies in $\mathfrak{a}$ or is divisible by both f_1 and f_2. Since f_1 and f_2 are relatively prime, the latter condition means that r is divisible by $f_1 f_2$, hence also in that case it belongs to $\mathfrak{a}$. ❏

PROPOSITION 10.44 (MONOMIAL LASKER–NOETHER) *Let $A = k[x_1, \ldots, x_n]$ and let $\mathfrak{a}$ be a monomial ideal. Then the associated primes and the isolated primary components of $\mathfrak{a}$ are monomial, and additionally, the embedded components may be chosen to be monomial as well.*

PROOF: We begin by determining a generating sets for $\mathfrak{a}$ consisting of primitive generators. The algorithm lies open in the day: if the generators of $\mathfrak{a}$ are pure powers, $\mathfrak{a}$ is primary. If not, at least one variable has a factorization $f = f_1 f_2$ with the f_i's involving

different variables, and they are therefore relatively prime. The lemma gives

$$\mathfrak{a} = (\mathfrak{a} + (f_1)) \cap (\mathfrak{a} + (f_2)A),$$

and the chosen generator set for $\mathfrak{a}$ induces generator sets for each of the ideals $(\mathfrak{a} + (f_i))$: just adjoint f_i to the old set and discard those generators that cease to be primitive. We iterate and feed each $(\mathfrak{a} + (f_i))$ into the algorithm unless it is primary in which case we leave it as it is. The process must eventually terminate since *e.g.* the sum of the degrees of the generators goes strictly down in each step.

And at the end, we have written $\mathfrak{a}$ as the intersection of monomial ideals that are primary. It will probably not be minimal, so some tiding up might be necessary. ❑

Examples

(10.12) An embedded component of a monomial ideal need not be monomial Exercise 10.4 on page 274 exhibits one: There you proved (or should have) that in the polynomial ring $k[x, y]$ it holds that

$$(x^2, xy) = (x) \cap (x^2, y + ax)$$

for all $a \in k$.

(10.13) According to the lemma $(x^n, xy, y^m) = (x, y^m) \cap (y, x^n)$ in $k[x, y]$, and all three ideals have (x, y) as radical and thus are (x, y)-primary. This illustrates that a risk with the algorithm is that primary ideals might be split up into intersections of ideals that at the end have to be coalesced if one goes for a minimal primary decomposition.

(10.14) Consider $\mathfrak{a} = (x^3y, xy^3, xz^3, zx^3, yz^3, zy^3)$ in the polynomial ring $A = k[x, y, z]$. Inverting z, we see that $\mathfrak{a}A_z = (x, y)A_z$ due to occurrence of the generators z^3y and z^3x. Similarly, $\mathfrak{a}A_x = (y, z)A_x$ and $\mathfrak{a}A_y = (x, z)A_y$. All these ideals are primary, in fact prime, and the equalities exclude the prime ideals (x), (y) and (z) from being associate, and we deduce that

$$\mathfrak{a} = (x, y) \cap (x, z) \cap (y, z) \cap \mathfrak{q} \tag{10.7}$$

where $\mathfrak{q}$ is (x, y, z)-primary. We claim that one may take $\mathfrak{q} = (x^3, y^3, z^3)$. Indeed, the inclusion $\subseteq$ is clear, and the proposition allows us merely to work with monomials when checking the other. So let the monomial $f = x^\alpha y^\beta z^\gamma$ belong to the right side of (10.7). Since $f \in \mathfrak{q}$, at least one of the exponents, say α, must be three or more, and as f lies in (z, y), either β or γ must be one or more, and hence f lies in $\mathfrak{a}$.

To illustrate the algorithmic approach we run the algorithm for our ideal $\mathfrak{a}$. At each stage the branching monomials are underlined in red, and all simplifications are tacitly

performed:

$$\begin{aligned}
\mathfrak{a} &= (\underline{x^3y}, xy^3, xz^3, \underline{zx^3}, yz^3, zy^3) \\
&= (x^3, \underline{xy^3}, xz^3, yz^3, zy^3) \cap (y, xz^3, \underline{zx^3}) \\
&= (x, \underline{yz^3}, zy^3) \cap (x^3, y^3, \underline{xz^3}, yz^3) \cap (y, z) \cap (y, \underline{xz^3}, x^3) \\
&= (x, y) \cap (x, z^3, \underline{zy^3}) \cap (x, y^3, \underline{yz^3}) \cap (x^3, y^3, z^3) \cap (y, z) \cap (y, z^3, x^3) \\
&= (x, y) \cap (x, z) \cap (x, z^3, y^3) \cap (x^3, y^3, z^3) \cap (y, z) \cap (y, z^3, x^3) \\
&= (x, y) \cap (x, z) \cap (y, z) \cap (x^3, y^3, z^3)
\end{aligned}$$

The last equality just tidies things up by making the intersection irredundant.

☆

⁂ EXERCISE 10.13 Exhibit, by way of the algorithm in Lemma 10.43, a minimal primary decomposition of $\mathfrak{a} = (x^\alpha y, z^\beta x, y^\gamma z)$ in the polynomial ring $k[x, y, z]$ where α, β and γ are natural numbers. ★

10.6 *Primary decomposition of modules*

The primary decomposition of ideals is easily generalised to submodules. In a Noetherian module, each submodule has a primary decomposition, which enjoy properties entirely analogous to the uniqueness properties in the ideal case. The proofs follow the same patterns and are built around the same ideas, but obviously there are necessary changes in the terminology. In this section we shall sketch the development, but shall leave most of the details to the students.

Primary modules

Primary ideals play a essential role in the theory of primary decomposition of ideals, and when extending the theory to modules, our first task is to establish a corresponding concept in the category of modules, and it is a carbon copy: a given a submodule N of the module M, is said to be *primary* if it fulfils the following condition:

Primary submodul (primære undermoduler)

- the homotheties in M/N induced by ring elements are either injective or nilpotent.

There is, as in the ideal case, an equivalent formulation in terms of the transporter ideal $(N : M) = \{ x \in A \mid xM \subseteq N \}$:

- if $xz \in N$ then either $z \in N$ or $x^n \in (N : M)$ for some natural number n.

Some of the most basic properties of primary modules are summarized in the following proposition. They are natural generalizations to modules of well-known properties of primary ideals. The proofs are basically the same as the proofs of the parallel properties in the ideal case, and, as we warned, they are left to the students. Note that the *radical*

radical of a submodule (radikalet til en undermodul)

of a submodule is defined as the radical of the relevant transporter ideal; that is, as $\sqrt{(N : M)}$.

PROPOSITION 10.45 *Let A be ring and M an A-module.*

i) The radical of a primary submodule is a prime ideal;

ii) Finite intersections of $\mathfrak{p}$-primary submodules are $\mathfrak{p}$-primary;

Let $N \subseteq M$ and $L \subseteq N$ be submodules and $S \subseteq A$ a multiplicative set.

iii) If N is a $\mathfrak{p}$-primary submodule and $S \cap \mathfrak{p} = \varnothing$, then $S^{-1}N$ is $S^{-1}\mathfrak{p}$-primary submodule of $S^{-1}M$;

iv) If N is $\mathfrak{p}$-primary in M, then N/L is a $\mathfrak{p}$-primary submodule of M/L.

Primary decomposition

Just as for ideals, a primary decomposition of a submodule is a presentation

$$N = N_1 \cap \ldots \cap N_r,$$

where each N_i is a primary submodule of M. It is *minimal* if it is irredundant and the radicals are different, and every primary decomposition may be render a minimal one by discarding redundant components and coalescing components with identical radicals. The components of a minimal primary decomposition are said to *primary components* of N. The Lasker–Noether theorem has a counterpart in the module category:

minimal primary decompositions (minimale primær dekomposisjoner)
Primary components (primærkomponenter)

THEOREM 10.46 *Every Noetherian module has a minimal primary decomposition*

PROOF: We attack by Noetherian induction. Consider the set of submodules of M for which the conclusion does not hold true. If the theorem fails for M, this set is non empty and has a maximal element N. We may replace M by M/N, so that every non-zero submodule has a primary decomposition but the zero-submodule (0) does not have one.

Then (0) is obviously not primary, and there are elements $x \in A$ and $z \in M$ so that $xz = 0$ but the homothety by x is not nilpotent. The kernels of the homotheties $[x^i]$ by the powers x^i form an ascending sequence of submodules and because M is Noetherian, must stabilize at a certain point, say at ν. Then $\ker[x^\nu] = \ker[x^{\nu+1}]$. We contend that $(0) = \ker[x] \cap x^\nu M$. Indeed, if $xz = 0$ and $z = x^\nu w$ it follows that $w \in \ker[x]^{\nu+1}$, hence in $\ker[x]^\nu$. By consequence $z = x^\nu w = 0$. Both $\ker[x]$ and $x^\nu M$ are non-zero submodule, and hence both are expressed as finite intersections of primary submodules, Evidently the same then holds for their intersection, which contradicts that no such expression exists for (0). ❑

Just as with ideals, one says that a primary components is isolated if its radical is minimal among the radicals of the components, and it is called *embedded* in the opposite case.

embedded components (embedded komponent)

Primary decompositions of submodules enjoy uniquiness properties in complete parallel to ideals; we state the two uniqueness theorems, but leave it to the interested student to work out the proofs (highly recommended) which *mutatis mutandis* aer the same as for ideals:

THEOREM 10.47 (THE TWO UNIQUENESS THEOREMS) *Let M a finitely generated module over a Noetherian ring A and let N be a submodule.*

i) The radicals of the primary components of N coincide with its associated prime ideals, and unambiguously defined by the submodule N;

ii) The isolated primary components of N are unique.

10.7 *Appendix: Primary decomposition and group actions*

In many situations there is group G acting on the ring A, and the ideal $\mathfrak{a}$ we study is invariant under G. Many rings arising from linear algebra or representation theory (for example coordinate rings of generic determinantal varieties) come endowed with natural actions of Lie groups. This is also the case when $\mathfrak{a}$ is a homogeneous ideals in the polynomial ring $R = k[x_1, \ldots, x_n]$ over a field k. Indeed, the multiplicative group k^* acts on R through homotheties; *i. e.* the action is given as $t \cdot f(x_1, \ldots, x_n) = f(tx_1, \ldots, tx_n)$, and f is homogenous of degree d it holds true that $f(tx) = t^d f(x)$, hence homogeneous ideals are invariant. Similarly, the group $(k^*)^n$ acts on $k[x_1, \ldots, x_r]$ with the action of $\lambda = (\lambda_1, \ldots, \lambda_n)$ on x_i being $\lambda \cdot x_i = \lambda_i x_i$. Then, in multi-index notation, it holds for a monomial $x\alpha$ that $\lambda \cdot x^\alpha = \lambda^\alpha x^\alpha$, and so monomial ideals will be invariant.

(10.48) A relevant question is if a G-invariant ideal $\mathfrak{a}$ has a primary decomposition that is invariant* under G. If G is finite, this is certainly not true; for instance, in $\mathbb{Z}[i]$ the primary decomposition of (5) is $(5) = (2+i) \cap (2-i)$, but complex conjugation swaps the two prime ideals $(2+i)$ and $(2-i)$. However, if the group G has no non-trivial finite quotient it holds true.

*The action induces an action on the ideals by setting $\mathfrak{a}^g = \{ga \mid g \in G, a \in \mathfrak{a}\}$

PROPOSITION 10.49 *Let G be a group without non-trivial finite quotients. Let A be a Noetherian ring on which the group G acts and let $\mathfrak{a}$ be a G-invariant ideal. Then $\mathfrak{a}$ has a G-invariant primary decomposition.*

PROOF: Let $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r$ be a minimal primary decomposition of $\mathfrak{a}$ with $\operatorname{Ass} A = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$ being the set of associated primes. For each $g \in G$ one has

$$\mathfrak{a} = \mathfrak{a}^g = \mathfrak{p}_1^g \cap \ldots \cap \mathfrak{p}_r^g, \tag{10.8}$$

and since g acts through an isomorphism of A, each ideal $\mathfrak{q}_i^g$ persists being primary, but will have $\mathfrak{p}_i^g$ as a possible new radical. Anyhow, (10.8) gives a minimal primary decomposition of $\mathfrak{a}$. The associated primes being unique, the group G permutes them,

and this action is manifested through a group homomorphism from G to the symmetric group Sym Ass A. By hypothesis, G has no non-trivial finite quotient so the image of G in the finite group Sym Ass A must be trivial, and consequently each $\mathfrak{p}_i$ is invariant under G.

A similar argument shows that isolated components of $\mathfrak{a}$ also are invariant under G.

As to the embedded components, which are not unique, things are slightly more complicated. They will not all be invariant, but we will be content when just finding one that is. So let $\mathfrak{q}$ be one having radical $\mathfrak{p}$. Since $\mathfrak{p}$ is invariant and $\mathfrak{p}^N \subseteq \mathfrak{q}$ implies that $(\mathfrak{p}^g)^N \subseteq \mathfrak{q}$, we may apply Exercise 10.15, and conclude that the intersection $\bigcap_{t\in G} \mathfrak{q}^g$ is $\mathfrak{p}$-primary. We contend it is a component of $\mathfrak{a}$; indeed:

$$\mathfrak{a} = \bigcap_{g\in G} \mathfrak{a}^g = \bigcap_{g\in G} (\mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_r) = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_s \cap \bigcap_{g\in G} q_{s+1}^g \cap \ldots \cap \bigcap_{g\in G} \mathfrak{q}_r^g,$$

where the s first components of $\mathfrak{a}$ are the isolated ones. ❑

Exercises

(10.14) Let $\{\mathfrak{q}_i\}_{i\in I}$ be a collection of primary ideals (of any cardinality) all with the same radical $\mathfrak{p}$. Assume that there is a natural number n so that $\mathfrak{p}^n \subseteq \mathfrak{q}_i$ for all i. Prove that the intersection $\bigcap_{i\in I} \mathfrak{q}_i$ is primary with $\mathfrak{p}$ as radical.

(10.15) Let the group G act on the Noetherian ring A and let $\mathfrak{q}$ be a $\mathfrak{p}$-primary ideal. Assume that $\mathfrak{p}$ is invariant under G. Prove that $\bigcap_{g\in G} \mathfrak{q}^g$ is $\mathfrak{p}$-primary and invariant under G.

(10.16) Show that if k is an algebraically closed field, the group $(k^*)^n$ has no finite quotient.

(10.17) Show that a connected Lie-group has no finite quotient.

★

Lecture 11

Krull dimension

Dimension is generally a complicated and subtle notion, and only in some good cases is there a satisfactory definition. Vector spaces have a dimension as do manifolds (or at least each connected component has). Manifolds are locally isomorphic to open sets in some euclidean space, and the dimension of that euclidean space is constant along each connected component, and is the dimension of the component.

There is another and naive approach to the concept of dimension. For example, in three dimensional geometric gadgets—which are called *threefolds*—one might heuristically imagine increasing chains of subgadgets of length three: points in curves, curves in surfaces and surfaces in the threefold. This may be formalized by using closed and irreducible subsets of a topological space as "subgadgets", and the dimension will be the maximal length of such a chain, or rather the supremum of the lengths as they might not be bounded. This definition works for any topological space, but the ensuing dimension does not carry much information unless the topology is "Zariski-like" (the only irreducible subspaces in a Hausdorff space are the points, so with this definition of dimension all Hausdorff spaces will be of dimension zero). Translated into algebra, where prime ideals correspond to closed irreducible subsets, this leads to the concept of Krull dimension of a ring; the supremum of the lengths of chains of prime ideals.

For varieties, or equivalently for algebras finitely generated over fields, there is another good candidate for the dimension, namely the transcendence degree over the ground field of the function field; that is the fraction field $K(A)$ of the coordinate ring in case the variety is affine . This may be motivated by the fact the Krull dimension of the polynomial ring $k[x_1, \ldots, x_n]$ equals n (which is not obvious, but follows from the Prinicipal Ideal Theorem), and obviously the transcendence degree of $k(x_1, \ldots, x_n)$ is n. That the two coincide, is one of the important consequences of Noether's Normalization Lemma.

11.1 *Definition and basic properties*

(11.1) Let A be a ring. We shall consider strictly ascending and finite chains $\{\mathfrak{p}_i\}$ of prime ideals

$$\mathfrak{p}_0 \subset \mathfrak{p}_1 \subset \ldots \subset \mathfrak{p}_\nu.$$

Krull dimension (Krull-dimensjon)

Saturated chains (mettede kjeder)

Maximal chains (maksimale kjeder)

Recall that the integer ν is called *the length* of the chain; it is one less than the number of prime ideals, or if you want, it equals the number of inclusions. The *Krull dimension* of A will be the supremum of the lengths of all such chains. It is denoted by $\dim A$. A chain is said to be *saturated* if there are no prime ideals in A lying strictly between two of the terms, and it is *maximal* if additionally the chain cannot be lengthened, neither upwards nor downwards; so the smallest term of a maximal chain is a minimal prime and the larges a maximal ideal.

(11.2) Even if it happens that each chain of prime ideals in A is finite, there might be arbitrary long chains, and the Krull dimension will be infinite. It is easy to find examples among non-Noetherian rings whose Krull dimension is infinite—the ring in Example 11.1 below is an obvious example of one with an infinite ascending chain—but even Noetherian rings might have infinite Krull dimension; (the first example, which we shall discuss in Example 14.1 on page 357, was discovered by Masayoshi Nagata). However, these examples live on the fringe of the Noetherian society, and rings met in mainstream algebraic geometry will all have finite dimension, and notably, every local Noetherian ring has finite Krull dimension, as we shall see.

Note that in a ring whose Krull dimension is finite, the supremum is achieved, and A will have saturated chains of maximal length; that is, of length equal to $\dim A$.

Examples

(11.1) The polynomial ring in $A = k[x_1, \ldots, x_r, \ldots]$ in infinite many variables is of infinite Krull dimension. Each of the ideals $\mathfrak{p}_r = (x_1, \ldots, x_r)$ is prime, and they form an infinite ascending chain. There is also an infinite descending chain of prime ideals in A whose members are the ideals $\mathfrak{q}_r = (x_r, x_{r+1}, \ldots)$.

(11.2) *Rings of dimension zero:* A ring A is zero-dimensional when no chain of prime ideals has more than one term; or in other words, when all prime ideals are both minimal and maximal. If A in addition is Noetherian, there are according to the Lasker-Noether Theorem (Teorem 10.19 on page 268) only finitely many minimal prime ideals. Hence A has only finitely many prime ideals and they are all maximal, and as this is a property that characterise Artinian rings among the Noetherian rings (Theorem 9.68 on page 256), a Noetherian ring is of dimension zero if and only if it is Artinian. In particular, fields and finite products of fields are of dimension zero. But also all rings

shaped like quotients $A/\mathfrak{a}$ of Noetherian rings by an ideal $\mathfrak{a}$ that is $\mathfrak{m}$-primary for a maximal ideal $\mathfrak{m}$, are of dimension zero.

(11.3) *Domains of dimension one:* It is worthwhile casting a glance on one-dimensional rings as well. In a one-dimensional domain the zero ideal is a prime ideal, and saturated chains are all of the form $(0) \subset \mathfrak{p}$. All non-zero prime ideals are therefore maximal, and they are also minimal over the zero ideal (they are what we later on will be calling the *height one primes*, one might also call them *subminimal*). Examples can be the principal ideal domains; in particular polynomial rings $k[x]$ in one variable over fields and, of course, the integers $\mathbb{Z}$ will all be one-dimensional.

Height one primes (høyde-en primidealer)

If A is not a domain, there may as well be saturated chains of length zero; in other words, some of the minimal prime ideals might also be maximal (see Example 11.5 below).

(11.4) *Some domains of dimension two:* The polynomial ring $A[x]$ over a PID A is of dimension two. Back in Chapter 3 (Proposition 3.30 on page 77) we showed, at least when A has infinitely many prime ideals, that the maximal ideals of $A[x]$ are of the type $\big(g(x), p\big)$ where p is a prime element in A, and $g(x)$ a polynomial that is irreducible modulo p. Furthermore, we saw that the non-maximal non-zero prime ideals all are principal, they are either equal to (p) for a prime element p in A or to $\big(g(x)\big)$ for an irreducible and primitive polynomial $g(x)$. From this one easily deduces that all maximal chains in $A[x]$ are of length two. Among the rings of this kind we find the polynomial rings $k[x, y]$ and $\mathbb{Z}[x]$.

The case when the PID A merely has finitely many maximal ideals, is more involved. The dimension of $A[x]$ will still be two, but the description of the ideals is different: there may be maximal ideals that are principal (see Example 11.8 below).

☆

⁂ EXERCISE 11.1 Let A be a Noetherian domain and let $x \in A$ a prime element so that the principal ideal (x) is prime. Show that $\dim A_{(x)} = 1$. HINT: Show that the powers $(x^n)A_{(x)}$ are all the ideals in $A_{(x)}$. ★

⁂ EXERCISE 11.2 Let $\mathfrak{p} = (y, z)$ in the polynomial ring $R = k[x, y, z]$. Show that $\dim R_{\mathfrak{p}} = 2$ HINT: : Excerise 7.8 on page 181 may be useful. ★

A useful inequality

(11.3) Any chain $\{\mathfrak{p}_i\}$ of prime ideals in A may be broken in two at any stage, say at $\mathfrak{p} = \mathfrak{p}_\nu$, and thus be presented as the concatenation of a lower chain, formed by the members of the chain contained in $\mathfrak{p}$, and an upper chain, formed by those containing $\mathfrak{p}$. And of course, one may as well splice two chains provided one ends at the prime where the other one begins. A chain thus split, appears as:

$$\mathfrak{p}_0 \subset \mathfrak{p}_1 \subset \ldots \subset \mathfrak{p}_{\nu-1} \subset \mathfrak{p}_\nu = \mathfrak{p} \subset \mathfrak{p}_{\nu+1} \subset \ldots \subset \mathfrak{p}_n.$$

The primes contained in $\mathfrak{p}$ are in a one-to-one correspondence with the prime ideals in the localization $A_\mathfrak{p}$, hence the lower chains correspond to chains in $A_\mathfrak{p}$. Moreover, the prime ideals containing $\mathfrak{p}$ correspond to prime ideals in the quotient $A/\mathfrak{p}$, hence the upper chains correspond to chains in $A/\mathfrak{p}$, and considering the suprema of the lengths of such splices, we arrive at the following formula:

PROPOSITION 11.4 *Let A be a ring and $\mathfrak{p}$ a prime ideal. Then*

$$\dim A_\mathfrak{p} + \dim A/\mathfrak{p} \leqslant \dim A.$$

Note that the proposition is still valid if one or more of the dimensions are infinite, provided the usual convention that $n + \infty$ equals ∞ is in force.

In some rings there are maximal chains—that is, saturated chains which cannot be lengthened either way—of different lengths, and in that case the Krull dimension is the length of the longest. For any prime ideal in a shorter chain, the inequality in (11.4) will be strict. It is easy to give such examples when the ring A is not a domain (but still is finitely generated over a field). Quite simply, coordinate rings of algebraic sets with irreducible components of different dimension will do (see Examples 11.5 and 11.6 below).

(11.5) More generally, if the ring A has several minimal prime ideals, the space Spec A will have several irreducible components, as in Example 11.5 below where the point $(0,1)$ and the x-axis are the components. The dimension of A will be the largest of the dimensions of the components, or if there are infinitely many, the supremum. In algebraic terms this translates into:

PROPOSITION 11.6 *If $\{\mathfrak{p}_i\}$ are the minimal primes of A, then* $\dim A = \sup_i \dim A/\mathfrak{p}_i$.

PROOF: The intersection of the prime ideals in a chain being prime, any maximal chain starts at minimal prime (Exercise 2.21 on page 53); furthermore chains of prime ideals beginning at a prime $\mathfrak{p}$ are in a one-to-one correspondence with chains of prime ideals in $A/\mathfrak{p}$. ❑

Height of ideals (høyden til idealer)

(11.7) It is common usage to call $\dim A_\mathfrak{p}$ the *the height* of $\mathfrak{p}$, or more generally for any ideal $\mathfrak{a}$ in A the heigh is the least height of any prime ideal containing $\mathfrak{a}$; that is

$$\operatorname{ht}\mathfrak{a} = \min_{\mathfrak{a}\subseteq\mathfrak{p}} \operatorname{ht}\mathfrak{p}.$$

Codimension of an ideal (kodimensjonen til et ideal)

One speaks also about the *height of* $\mathfrak{p}$ *over* $\mathfrak{q}$ when $\mathfrak{p} \supset \mathfrak{q}$ are two primes. It equals the supremum of lengths of chains connecting $\mathfrak{q}$ to $\mathfrak{p}$; or equivalently to $\dim(A/\mathfrak{q})_\mathfrak{p}$.

The height $\operatorname{ht}\mathfrak{p}$ is also called the *codimension* of $V(\mathfrak{p})$, since when the inequality in Proposition 11.4 is an equality, it holds that $\operatorname{ht}\mathfrak{p} = \dim A_\mathfrak{p} = \dim A - \dim A/\mathfrak{p}$.

Catenary rings (katenære ringer)

(11.8) Rings with the property that for each pair of nested prime ideals all saturated chains connecting the two have the same length, are called *catenary*. The rings in the

examples just mentioned are both catenary despite having maximal chains of different length, so being catenary is a weaker property than having maximal chains of uniform length. There are many examples of Noetherian domains that are not catenary, but these are rather exotic constructs you wouldn't tumble over when practicing algebraic geometry.

⁂ EXERCISE 11.3 Show that for a ring to be catenary it suffices that the defining property holds for the nested pairs consisting of a minimal and a maximal ideal. Show that if $A/\mathfrak{p}$ is catenary for each minimal prime $\mathfrak{p}$, then A is catenary. ★

EXERCISE 11.4 Show that $k[x,y]$ and $\mathbb{Z}[x]$ are catenary. HINT: Proposition 3.30 on page 77. ★

In several large and important classes of rings all members are catenary. For instance, as we shall see later, all algebras which are finitely generated over a field are catenary as are their localizations. For domains from this class of rings, the stronger property that all maximal chains are of equal length also holds true, we shall say they are of *uniform altitude*. However this does not necessarily persist being true for localizations of such as Examples 11.7 and 11.8 below shows.

Uniform altitude
(uniform høyde)

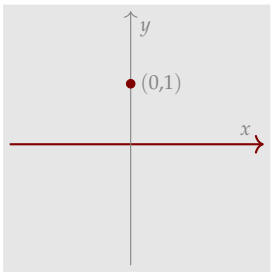
One easily convinces oneself that equality holds in Proposition 11.4 for all primes $\mathfrak{p}$ in a local catenary domain, and in fact, Louis Ratliff has shown that the converse also is true when A is Noetherian—a rather deep result with a rather involved proof.

Examples

(11.5) Let $A = k[x,y]$ with constituting relations $xy = y(y-1) = 0$. It is the coordinate ring of the algebraic subset V of $\mathbb{A}^2_k$ equal to the union of the x-axis and the point $(0,1)$. The primary decomposition of the zero ideal in A is given as $(0) = (y) \cap (x, y-1)$. Hence (y) and $(x, y-1)$ are the minimal prime ideals in A. Now, $(x-a, y)$ is a maximal ideal for any $a \in k$, so A possesses saturated chains

$$(y) \subset (x-a, y),$$

and therefore $\dim A = 1$. On other hand $(x, y-1)$ is clearly a maximal ideal, and hence it is both maximal and minimal. So V, even though it is one-dimensional, has a component of dimension zero.



(11.6) If you want an example that is a local ring, refine the previous example and consider $A = k[x,y,z]$ with constituting relations $zx = zy = 0$. Then A is the coordinate ring of the union V of the xy-plane and the z-axis in $\mathbb{A}^3_k$. Let further A be the local ring of V at the origin; that is, the localization of A at (x,y,z). In A the decomposition $(0) = (z) \cap (x,y)$ is the minimal primary decomposition of (0) so that both (z) and (y,x) are minimal prime ideals in A; and there are two maximal chains $(z) \subset (x,z) \subset (x,y,z)$ and $(x,y) \subset (x,y,z)$ in A of different length.

(11.7) *A semi-local ring:* Consider the polynomial ring $k[x,y,z]$ and let $\mathfrak{p} = (x)$ and

$\mathfrak{q} = (y, z)$. Let S denote multiplicatively closed subset consisting of elements not in the union $\mathfrak{p} \cup \mathfrak{q}$. Furthermore, let $A = S^{-1}k[x, y, z]$. We contend that A is a semi-local ring with maximal ideals $\mathfrak{m} = \mathfrak{p}A$ and $\mathfrak{n} = \mathfrak{q}A$ whose heights are respectively one and at least* two, and there are maximal chains of unequal length in A. That $\mathfrak{m}$ and $\mathfrak{n}$ are the only two maximal ideals is clear: indeed, every prime in A is of the form $\mathfrak{a}A$ for a prime ideal $\mathfrak{a} \subseteq k[x, y, z]$ not meeting S; that is, $\mathfrak{a}$ is contained in the union $\mathfrak{p} \cup \mathfrak{q}$. Prime Avoidance then ensures that $\mathfrak{a}$ either lies in $\mathfrak{p}$ or in $\mathfrak{q}$. That $\dim A_{\mathfrak{p}} = 1$ and $\dim A_{\mathfrak{q}} \geqslant 2$ are just Exercises 11.1 and 11.2 on page 289.

*In fact, the height equals two. Later, this will be crystal clear, but for the moment we do not spend energy on it

(11.8) *Chains in the polynomial ring over a* DVR*:* Recall that in Exercise 3.8 on page 78 you were asked to describe the maximal ideals in the polynomial ring $A[x]$ over a discrete valuation ring A. As in that exercise, we let π be a generator for the unique maximal ideal $\mathfrak{m}$ of A. There are two types of maximal ideals in $A[x]$: first we have those of shape $(g(x), \pi)$ with $g(x)$ being irreducible modulo π. These live in saturated chains of length two:

$$(0) \subset (\pi) \subset (g(x), \pi) \qquad \text{or} \qquad (0) \subset (g(x)) \subset (g(x), \pi).$$

Secondly, we have those which are principal; *i. e.* those shaped like $(g(x))$ where $g(x)$ is an irreducible polynomial which is constant modulo π (a typical example would be $(\pi x - 1)$). These live in chains of length one:

$$(0) \subset (g(x)).$$

So we see that $\dim A[x] = 2$, but that there are saturated chains of different lengths. But $A[x]$ is catenary, chains connecting the minimal prime (0) to a fixed maximal ideal are all of the same length.

It is noteworthy that a prime ideal $(g(x))$ with g irreducible mod π lives in a unique maximal chain: if $(g(x)) \subset (h(x), \pi)$ with $h(x)$ irreducible mod π, it is straightforward to verify that $(g(x), \pi) = (h(x), \pi)$. So the ideal $(g(x))$ is not the intersection of maximal ideals, and the closed points of closed subsets of Spec $A[x]$ are not always dense in the subset (*e.g.* the closed points are not dense in $V(g(x))$, as there is merely one).

☆

EXERCISE 11.5 Let A be a DVR whose maximal ideal is generated by π. Describe all prime ideals in $A[x]$ lying between (0) and a maximal ideal of shape $(g(x), \pi)$ (where $g(x)$ is a polynomial which is irreducible modulo π). ★

Cutting out a hypersurface

Cutting a variety X with a hypersurface is a rather common technique in algebraic geometry, which on the level of algebras corresponds to passing to the quotient $A/(f)A$ of A by a principal ideal. One expects the dimension to go down, however, it might

happen that the cutting hypersurface contains one of the components of X, and in case that component is one of maximal dimension, the dimension stays the same. To avoid such an accidental behaviour, one must assume that f does not lie in any of the minimal prime ideals of A; then one has:

PROPOSITION 11.9 *Let A be ring of finite Krull dimension and let $f \in A$ be an element not belonging to any minimal prime ideal in A, then* $\dim A/(f)A < \dim A$.

PROOF: Chains of prime ideals in $A/(f)A$ are in one-to-one correspondence with chains in A all whose members contain f. Moreover, a prime ideal $\mathfrak{p}$ that is minimal over (f), is by hypothesis not minimal in A, and therefore properly contains a minimal prime $\mathfrak{q}$ of A. Consequently any ascending chain in A emanating from $\mathfrak{p}$ can be lengthened downwards by appending $\mathfrak{q}$. ❑

11.2 *Krull's Principal Ideal Theorem*

This important theorem is also known under its German name Krull's Hauptidealsatz. It lies at the bottom of the dimension theory in commutative algebra and algebraic geometry, and in his book [**?**] Irving Kaplansky refers to it as "may be the most important single theorem in the theory of Noetherian rings". There is however the rather simple underlying intuition that the dimension of a solution space goes down by at most one when an additional equation is introduced. We recognize this from the theory of linear equations, but the principle has a much wider scope (as shows *e.g.* the Principal Ideal Theorem).

(11.10) The scene is set with a ring A and the two main players, a prime ideal $\mathfrak{p}$ and an element x of $\mathfrak{p}$ over which $\mathfrak{p}$ is minimal. That is to say, there is no prime ideal properly lying between (x) and $\mathfrak{p}$. And then the conclusion is that the height $\operatorname{ht}\mathfrak{p}$ is at most one. Of course, the prime ideal $\mathfrak{p}$ might be a minimal one and then the height would be equal to zero, but if x is non-zero-divisor, the height will be one.

EXAMPLE 11.9 As a prelude to the theorem, it might be instructive to consider the special and simple case that (x) is a prime ideal in a Noetherian domain A. Then there are no non-zero prime ideals properly contained in (x): if $\mathfrak{p} \subseteq (x)$ were one of the kind, it would follow that $x\mathfrak{p} = \mathfrak{p}$; indeed, each member $a \in \mathfrak{p}$ would be on the form $a = bx$, and because $x \notin \mathfrak{p}$, one would have $b \in \mathfrak{p}$. In view of Nakayama Extended, the ideal $\mathfrak{p}$ would be killed by an element shaped like $1 + cx$, which is absurd. ☆

(11.11) In the course of the proof we are obliged to use the symbolic powers of a prime ideal $\mathfrak{q}$ in A as a substitute for the actual powers. Recall (see Exercise 10.11 on page 275) that the symbolic power $\mathfrak{q}^{(n)}$ is defined as $\mathfrak{q}^{(n)} = \mathfrak{q}^n A_{\mathfrak{q}} \cap A$. It has the virtue of being $\mathfrak{q}$-primary, which the power $\mathfrak{q}^n$ might not be in general. But don't be panic-stricken by the appearance of these gizmos: A part from $\mathfrak{q}^{(n+1)}$ being contained in $\mathfrak{q}^{(n)}$, the only

property we shall need, is that from $cx \in \mathfrak{q}^{(n)}$, but $x \notin \mathfrak{q}$, follows that $c \in \mathfrak{q}^{(n)}$. Indeed, x is invertible in the localization $A_\mathfrak{q}$ as it does not lie in $\mathfrak{q}$, so $c \in \mathfrak{q}^n A_\mathfrak{q}$ follows from $cx \in \mathfrak{q}^n A_\mathfrak{q}$.

$(x) \subseteq \mathfrak{p}$
$\cup$
$\mathfrak{q}$
$\cup$
$\mathfrak{q}'$

THEOREM 11.12 (KRULL'S PRINCIPAL IDEAL THEOREM) *Let A be a Noetherian ring and x an element from A. Assume that $\mathfrak{p}$ is a prime ideal in A which is minimal over (x). Then $\operatorname{ht}\mathfrak{p} \leqslant 1$.*

PROOF: We are to show that there are no chain of prime ideals of length two of shape $\mathfrak{q}' \subset \mathfrak{q} \subset \mathfrak{p}$. By passing to the quotient $A/\mathfrak{q}'$ and subsequently localizing in $\mathfrak{p}/\mathfrak{q}'$, we may assume that A is a local domain with maximal ideal $\mathfrak{p}$, and our task is to prove that if $\mathfrak{q} \subset \mathfrak{p}$, then $\mathfrak{q} = 0$.

The first observation is that, since $\mathfrak{p}$ is minimal over (x), the ring A/xA has only one prime ideal. Being Noetherian, it is Artinian, and we have the opportunity to activate the descending chain condition. The chain to exploit, is the descending chain $\{(x) + \mathfrak{q}^{(n)}\}_n$, where $\mathfrak{q}^{(n)}$ is the n-th symbolic power of $\mathfrak{q}$. The chain $\{(x) + \mathfrak{q}^{(n)}\}$ corresponds to the descending chain $\{\left((x) + \mathfrak{q}^{(n)}\right)/(x)\}$ in A/xA and must eventually be stable as A/xA is Artinian. Hence there is an n so that

$$(x) + \mathfrak{q}^{(n+1)} = (x) + \mathfrak{q}^{(n)}.$$

This entails that if $a \in \mathfrak{q}^{(n)}$, one may write $a = b + cx$ with $b \in \mathfrak{q}^{(n+1)}$, so that $cx \in \mathfrak{q}^{(n)}$. From this follows that $c \in \mathfrak{q}^{(n)}$ since $x \notin \mathfrak{q}$, and consequently it holds true that $\mathfrak{q}^{(n)} \subseteq \mathfrak{q}^{(n+1)} + x\mathfrak{q}^{(n)} \subseteq \mathfrak{q}^{(n)}$. Nakayama's lemma yields that $\mathfrak{q}^{(n)} = \mathfrak{q}^{(n+1)}$. In its turn, this yields that $\mathfrak{q}^n A_\mathfrak{q} = \mathfrak{q}^{n+1} A_\mathfrak{q}$, and appealing once more to Nakayama's lemma, we may conclude that $\mathfrak{q}A_\mathfrak{q} = 0$; that is, $\mathfrak{q} = 0$. ❑

EXAMPLE 11.10 The non-Noetherian ring $A = k[t, x_1, x_2, \ldots]$ with constituting relations $x_i = tx_{i+1}$ for $i \in \mathbb{N}$, which was the subject of Exercise 9.18, gives an example underlining that the Noetherian hypothesis is necessary. The principal ideal $(t)A$ is maximal, but contains the non-zero prime ideal $\mathfrak{p} = (x_1, x_2, \ldots)$, and consequently it is of height (at least) two.

Note that $\dim A = 2$, but when t is killed, all the x_i's are killed as well, so that $A/(t)A \simeq k$. Hence $\dim A/(t)A = 0$, and killing t thus diminishes the dimension by two. Notice that this may also happens for Noetherian rings. ☆

The general version of The Principal Ideal Theorem

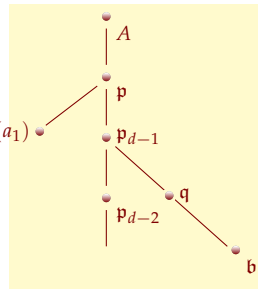
This result, often called "The height theorem", is the natural generalization of the Principal Ideal theorem to minimal primes over ideals with more than one generator. The natural guess that the height will be at most the number of generators, is actually true, and agrees with the naive intuition that imposing r constraints on a system should at most lower the dimension of the solution space by r: it diminishes at most with one

for each new condition imposed. This points to an induction argument, but a slightly more subtle one than the naive intuition suggests.

THEOREM 11.13 (THE HEIGHT THEOREM) *Let A be a Noetherian ring and let $\mathfrak{p}$ be a prime ideal minial over an ideal $\mathfrak{a}$ generated by r elements. Then* $\operatorname{ht}\mathfrak{p} \leqslant r$.

PROOF: Let $\mathfrak{a} = (a_1, \dots, a_r)$. As indicated above the proof goes by induction on r, and heading for a contradiction, we assume that there is a chain $\{\mathfrak{p}_i\}$ in A of length $d > r$ ending at $\mathfrak{p}$. Because $\mathfrak{a}$ is not contained in $\mathfrak{p}_{d-1}$, we may assume that a_1 does not lie in $\mathfrak{p}_{d-1}$, and so there is no prime lying properly between $(a_1) + \mathfrak{p}_{d-1}$ and $\mathfrak{p}$. The radical of $(a_1) + \mathfrak{p}_{d-1}$ therefore equals $\mathfrak{p}$, and a power of $\mathfrak{p}$ is contained in $(a_1) + \mathfrak{p}_{d-1}$. Since $\mathfrak{a} \subseteq \mathfrak{p}$ we may for ν sufficiently big, write

$$a_i^\nu = c_i a_1 + b_i$$



with $b_i \in \mathfrak{p}_{d-1}$, and we let $\mathfrak{b} = (b_2, \dots, b_r)$. Then $\mathfrak{b}$ is contained in $\mathfrak{p}_{d-1}$. We contend that there is prime ideal $\mathfrak{q}$ lying between $\mathfrak{b}$ and $\mathfrak{p}_{d-1}$, properly contained in $\mathfrak{p}_{d-1}$; indeed, if $\mathfrak{p}_{d-1}$ were minimal over $\mathfrak{b}$, the height of $\mathfrak{p}_{d-1}$ would be at most $r-1$ by induction, but being next to the top in a chain of length d, the ideal $\mathfrak{p}_{d-1}$ is of height at least $d-1$, and $r-1 < d-1$.

Now, the idea is to pass to the ring $A/\mathfrak{q}$. The ideal $\mathfrak{q} + (a_1)$ contains a power of $\mathfrak{a}$, hence there is no prime ideal between $\mathfrak{q} + (a_1)$ and $\mathfrak{p}$, which means that the ideal $\mathfrak{p}/\mathfrak{q}$ is minimal over the principal ideal $\mathfrak{q} + (a_1)/\mathfrak{q}$, and therefore of height one after the Principal Ideal Theorem, but there is also the chain $0 \subset \mathfrak{p}_{d-1}/\mathfrak{q} \subset \mathfrak{p}/\mathfrak{q}$. Contradiction. ❑

Some consequences

(11.14) It ensues from the Height Theorem that any local Noetherian ring A has a finite Krull dimension. Indeed, the maximal ideal $\mathfrak{m}$ is finitely generated, and by the Height Theorem the height of $\mathfrak{m}$, which is the same as $\dim A$, is bounded by the number of generators. Similarly, Noetherian rings enjoy a descending chain condition for prime ideals. Any term in a chain is finitely generated and hence is of finite height, and the length of the chain is bounded by the height of the top member. We have proved the first part of the following:

PROPOSITION 11.15 *A local Noetherian ring is of finite Krull dimension bounded by the number of generators of the maximal ideal. Noetherian rings satisfy the* DCC *for prime ideals in the strong sense that there is a uniform bound on the length of descending chains.*

Note that in any ring, Noetherian or not, the weaker property that any non-empty set of prime ideals has a minimal and a maximal member holds true. A Zorn's-lemma-argument gives this, the intersection and the union of the ideals in a chain of primes being prime.

Despite every chain of prime ideals being of finite length, Noetherian rings may be of infinite dimension; the lengths of the different chains could be unbounded. The first example of such a ring was given by Nagata and will be discussed in Example 14.1 on page 357.

(11.16) Another special feature the lattice of prime ideals in Noetherian rings has, is that in between two prime ideals, one strictly contained in the other, there will be infinitely many prime ideals if any at all.

PROPOSITION 11.17 *Let A be a Noetherian ring and let $\mathfrak{q} \subset \mathfrak{p}$ be two prime ideals. If there is a prime ideal lying strictly between $\mathfrak{p}$ and $\mathfrak{q}$, there will be infinitely many.*

PROOF: Assume there is a prime ideal strictly between $\mathfrak{p}$ and $\mathfrak{q}$ which means that the $\mathfrak{p}$ is of height at least two over $\mathfrak{q}$. If there were only finitely many prime ideals, say $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$, lying strictly between $\mathfrak{p}$ and $\mathfrak{q}$, there would by prime avoidance be an element $x \in \mathfrak{p}$ not lying in any of the $\mathfrak{p}_i$ since by assumption $\mathfrak{p}$ is not contained in any of the $\mathfrak{p}_i$'s. Then $\mathfrak{p}$ would be minimal over $\mathfrak{q} + (x)$, and by the Principal Ideal Theorem $\mathfrak{p}$ would be of height one over $\mathfrak{q}$; contradiction. ❑

Curiously enough, the proposition fails spectacularly for non-Noetherian rings. One of the simplest non-Noetherian valuation rings is of dimension two and has only three prime ideals—which of course form a chain—and more generally, for every dimension n there are similar examples of valuation rings having only $n+1$ prime ideals. Describing these example requires some theory about general valuations and will be done later in the course (Proposition 15.63 and Example 15.9 on page 400).

These examples also show that the Principal Ideal Theorem may fail without the Noetherian hypothesis; *e.g.* in the two-dimesional example just mentioned the maximal ideal is of height two, but it is minimal over any one of its elements not in the other non-zero prime ideal.

EXAMPLE 11.11 Here is another example of the principal ideal theorem failing for non-Noetherian rings. Back in Lecture 9, in Exercise 9.18, you were asked to studied the non-Noetherian ring $A = k[t, x_1, x_2, \ldots]$ with constituting relations $x_i = tx_{i+1}$ for $i \geqslant 1$. In this ring $\mathfrak{m} = (t)A$ is a maximal ideal and $\mathfrak{p} = (x_1, x_2, \ldots)$ is a prime ideal contained in $\mathfrak{m}$ (actually equal to $\bigcap_i \mathfrak{m}^i$). So in this ring the height of the principal ideal $(t)A$ equals two. ☆

Embedding dimension (embeddingsdimensjon)

(11.18) Among the different numbers associated with a ring which is reminiscent of being a dimension, is the so-called *embedding dimension* of a local ring A. If $\mathfrak{m}$ denotes maximal ideal of A, the embedding dimension of A is defined as the vector space dimension $\dim_{A/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2$ (the module $\mathfrak{m}/\mathfrak{m}^2$ is killed by $\mathfrak{m}$ and therefore is a vector space over $A/\mathfrak{m}$). Any vector space basis of $\mathfrak{m}/\mathfrak{m}^2$ is of the form $[x_1], \ldots, [x_r]$ for

members x_i of $\mathfrak{m}$. Nakayama's lemma implies that the x_i's generate $\mathfrak{m}$, and in its turn The Height Theorem yields that $\dim A \leqslant r$. We have thus proved

PROPOSITION 11.19 *Assume that A is a local Noetherian ring with maximal ideal $\mathfrak{m}$. Then $\dim A \leqslant \dim \mathfrak{m}/\mathfrak{m}^2$.*

EXERCISE 11.6 Show that among the Noetherian rings only the Artinian ones and the semi local one-dimensional ones have finitely many prime ideals. ★

EXERCISE 11.7 Assume that A is a quotient of te polynomial ring $k[x_1, \ldots, x_n]$ over the field k. Let $\mathfrak{m}$ be a maximal ideal in A. Show that the embedding dimension of $A_\mathfrak{m}$ is at most n. ★

11.3 UFD's *once more*

In Lecture 3 we showed a criterion of Kaplansky's (Proposition 3.13 on page 73) which tells us that a domain A is a UDF if and only if "every prime contains a prime". When A is a Noetherian domain, this criterion can be improved. Citing Proposition 11.15 above which asserts that prime ideals in a Noetherian ring satisfy the descending chain condition, we infer that any prime ideal in A contains a prime ideal of height one (if this is not true, one easily constructs a descending chain that does not terminate). To ensure that A is a UFD it therefore suffices that prime ideals of height one contain prime elements, but of course in that case, since prime elements generate prime ideals, the height one ideal is itself genertd by the prime element. This leads to

THEOREM 11.20 *A Noetherian domain A is a UFD if and only if every prime ideal of height one is principal.*

(11.21) A direct consequence of the theorem is that if A is Noetherian and factorial, any localization $S^{-1}A$ of A is factorial; indeed, the height one primes in $S^{-1}A$ are precisely the ideals shaped like $S^{-1}\mathfrak{p}$ for height one primes $\mathfrak{p}$ in A not meeting S, and localizations of principal ideals persist being principal.

The converse is however not true, as shows the example of the simple cusp $A = k[x, y]$ with constituting relation $y^2 = x^3$. Inverting x gives $A_x = k[t, t^{-1}]$, where $t = xy^{-1}$; this is factorial, but A is not. The element x is irreducible but not prime, and this is loosely speaking, what prohibits A from being a UFD. Inverting x makes it unit, and the problem disappears.

However, if one merely inverts prime elements, the ring will be a UFD when the localization is. The following proposition is due to Nagata:

PROPOSITION 11.22 *Let A be a domain with ascending chain condition on principal ideals (e.g. Noetherian) and S a multiplicative subset generated by prime elements. If $S^{-1}A$ is factorial, then A is factorial.*

Note that the elements from S are non-empty products on primes so that S does not contain any unit.

PROOF: By Kaplansky's criterion it suffices to exhibit a prime element in every prime ideal $\mathfrak{p}$ of A. If $\mathfrak{p} \cap S \neq \varnothing$, this is immediate, so we may well assume that $\mathfrak{p} \cap S = \varnothing$. Let x be prime element in $S^{-1}\mathfrak{p}$. We may assume that x lies in A, and by the lemma below, we may also assume that x has no factor from S. That x is a prime in $S^{-1}A$ means that if $x|ab$, it follows that x divides either a or b in $S^{-1}A$, say a; that is, it holds true that $sa = a'x$ for some $s \in S$ and some $a' \in A$. Now, s is a finite product of primes not dividing x, consequently s may be cancelled, and $a = a''x$ for some $a'' \in A$. ❑

Here comes the required lemma:

LEMMA 11.23 *Let A be a domain with ascending chain condition on principal ideals (e.g. Noetherian) and let S be a multiplicative system in A generated by primes. If $x \in A$ does not belong to S, either x has no factor from S, or one may write $x = sy$ where $s \in S$ and y has no factor from S.*

PROOF: Consider the set Σ of principal ideals (z) where z runs through elements such that $z^{-1}x \in S$. If non-empty the set Σ has a maximal element say (y). Then $x = sy$ with $s = xy^{-1}$ and $s \in S$. Assume the that $y = tz$ with $t \in S$. Then $(z) \in \Sigma$ because $xz^{-1} = st \in S$ and it follows that t is a unit, a contradiction. ❑

EXAMPLE 11.12 Introduce a grading of the polynomial algebra $B = k[x_1, \ldots, x_r]$ by assigning positive weights w_i to the variables x_i's. Moreover, let f be a polynomial, irreducible and homogeneous of degree w. Let $A = k[x_1, \ldots, x_r, z]$ with constituting relation $z^c = F(x_1, \ldots, x_r)$. Assume that $c \equiv \pm 1 \mod w$. Then A is factorial.

The trick is to consider the localization $A[z^{-1}]$. The first step is to observe that we have $A/(z)A \simeq k[x_1, \ldots, x_r]/(f)$, which is an integral domain because f is assumed to be irreducible. Hence z is a prime element of A, and it suffices to see that $A[z^{-1}]$ is factorial. Now, in $A[z^{-1}]$ one has the elements $y_i = z^{-dw_i}$ where d is the natural number so that $c = \pm 1 + dw$. One finds

$$z^c = f(x_1, \ldots, x_r) = f(z^{dw_1}y_1, \ldots, z^{dw_r}y_r) = z^{dw}f(y_1, \ldots, y_r) = z^{c\pm 1}f(y_1, \ldots, y_r),$$

and it therefore holds true that

$$z = f(y_1, \ldots, y_r) \text{ or } z^{-1} = f(y_1, \ldots, y_r)$$

according to $c \equiv 1$ or $c \equiv -1 \mod w$. But in both cases we conclude that $A[z^{-1}] = k[y_1, \ldots, y_r, f^{-1}]$, which is a factorial ring. Hence A itself is factorial by Proposition 11.22.

In particular the so-called Du Val E_8-singularity $k[x, y, z]/(x^2 + y^3 + z^5)$ is factorial. ☆

11.4 System of parameters

(11.24) Let A be a local ring with maximal ideal $\mathfrak{m}$ whose Krull dimension is n. A sequence $x_1, \ldots, x_n$ of n elements in A is called a *system of parameters* if the ideal $\mathfrak{a}$ they generate is $\mathfrak{m}$-primary. Since $\mathfrak{m}$ is maximal, this amount to the radical of $(x_1, \ldots, x_n)$ being equal to $\mathfrak{m}$ (Proposition 10.4 on page 261), or if A is Noetherian that $\mathfrak{a}$ contains some power of $\mathfrak{m}$.

System of parameters (parmetersystemer)

PROPOSITION 11.25 *Every local Noetherian ring has a system of parameters.*

PROOF: Let A be the ring, let $\mathfrak{m}$ the maximal ideal, and let $n = \dim A$. We shall, by a recursive construction, exhibit a sequence $x_1, \ldots, x_n$ of elements in $\mathfrak{m}$ so that each ideal $\mathfrak{a}_i = (x_1, \ldots, x_i)$ generated by the i first elements in the sequence has all its minimal primes of height i. Assume that $\mathfrak{a}_\nu$ has been constructed and consider the prime ideals $\{\mathfrak{p}_j\}$ minimal over $\mathfrak{a}_\nu$. They all have height ν so if $\nu < n$, none of them equals $\mathfrak{m}$. Hence their union is not equal to $\mathfrak{m}$ by Prime Avoidance (Lemma 2.32 on page 42), and we may pick an element $x_{\nu+1}$ from A so that $x_{\nu+1} \in \mathfrak{m}$, but $x_{\nu+1} \notin \bigcup_i \mathfrak{p}_i$. Then any prime ideal minimal over $\mathfrak{a}_{\nu+1} = (x_1, \ldots, x_{\nu+1})$ has height $\nu + 1$; indeed, let $\mathfrak{p}$ be one of them. It is not among the minimal prime ideals $\mathfrak{p}_i$ of $\mathfrak{a}_\nu$, and therefore must contain one of $\mathfrak{p}_i$'s properly, say $\mathfrak{p}_j$, and we infer that $\operatorname{ht}\mathfrak{p} > \operatorname{ht}\mathfrak{p}_j = \nu$. The other inequality; that is $\operatorname{ht}\mathfrak{p} \leqslant \nu + 1$, ensues from the Height Theorem. ❑

(11.26) The geometric counterpart of a system of parameters is, given a variety X and point P on X, a sequence of hyper-surfaces that locally intersect the given variety in just the point P, or expressed more precisely, that P is an isolated point in the intersection of X and the hyper-surfaces.

(11.27) One cannot in general hope that the maximal ideal itself is generated by $\dim A$ elements. The simple double point $A = k[X, Y]/(Y^2 - X^2(X + 1))$ gives an easy example. The maximal ideal $\mathfrak{m} = (x, y)$ requires both x and y as generators; indeed, no non-trivial linear combination $\alpha X + \beta Y$ with $\alpha, \beta \in k$ can for degree reasons belong to $(Y^2 - X^2(X + 1))$, and therefore x and y are linearly independent modulo $\mathfrak{m}^2 = (x^2, xy, y^2)$.

The geometric situation is as follows. Both the X-axis and the Y-axis (and in fact, any line through the origin) intersect the curve $C = V(Y^2 - X^2(X + 1))$ only at the origin, but because C has a double point there, there will always be an intersection multiplicity. Heuristically, the curve C has two branches through the origin, and each one contributes to the intersection with the line.

Noetherian local rings whose maximal ideal needs no more generators than the Krull dimension are said to *regular*. A general Noetherian ring is regular if the local rings $A_\mathfrak{p}$ are regular for all prime ideals $\mathfrak{p}$ in A.

Regular local rings (regulære lokale ringer)

EXERCISE 11.8 With notations as above, show that $\ell_{A_\mathfrak{p}}(A_\mathfrak{p}/(Y-\alpha X)A_\mathfrak{p}) = 2$ when α is a scalar and $\alpha \neq \pm 1$, but that $\ell_{A_\mathfrak{p}}(A_\mathfrak{p}/(Y+X)A_\mathfrak{p}) = \ell_{A_\mathfrak{p}}(A_\mathfrak{p}/(Y-X)A_\mathfrak{p}) = 3$. The excess length in the latter case is heuristically explain by the lines $Y = \pm X$ being tangent to C at the origin). ★

EXERCISE 11.9 *A du Val singularity.* Consider the ring $k[x,y,z]$ with the constituting relation $xy + z^{n+1} = 0$ (that is, the ring $k[X,Y,Z]/(XY + Z^{n+1})$) and let $\mathfrak{m} = (x,y,z)$.

a) Show that $x, \alpha y + \beta z$ is a system of parameters for $A_\mathfrak{m}$ whenever $(\alpha,\beta) \neq (0,0)$. Compute $\ell_{A_\mathfrak{m}}(A_\mathfrak{m}/(x,\alpha y+\beta z))$

b) Show that any pair of linearly independent linear forms w_1 and w_2 in x, y and z form a system of parameters for $A_\mathfrak{m}$. Discuss $\ell_{A_\mathfrak{m}}(A_\mathfrak{m}/(w_1,w_2))$.

★

Dimension and fibres

(11.28) One of the more important formulas taught in courses in linear algebra relates the dimension of the kernel and the cokernel of a linear map. Recall that for a given linear map $\phi\colon V \to W$ it reads as

$$\dim \operatorname{im}\phi + \dim \ker \phi = \dim V + \dim W,$$

which, since $\dim \operatorname{im}\phi \leqslant \dim V$, yields the inequality $\dim V \leqslant \dim + \dim W\phi^{-1}(0)$. For a smooth map $\phi\colon X \to Y$ between manifolds there is a similar inequality

$$\dim X \leqslant \dim Y + \dim \phi^{-1}(y),$$

for y when y belongs to the image $\phi(X)$ of ϕ; in fact, this is just the inequality from linear algebra above applied to the derivative of ϕ. When $\phi\colon X \to Y$ is a map of varieties, or between spectra of rings, there is a similar formula. We do not intend to enter any geometric discussion, but shall give a a closely related algebraic version for maps of local rings.

Maps of local rings (lokale ringavbildning)

(11.29) Recall that a *map of local rings* is ribetween two local rings which sends the maximal ideal into the maximal ideal.

PROPOSITION 11.30 *Let A and B be the two local rings having maximal ideals $\mathfrak{m}$ and $\mathfrak{n}$ respectively, and assume that $\phi\colon A \to B$ is a map of local Noetherian rings. Then it holds true that*

$$\dim B \leqslant \dim A + \dim B/\mathfrak{m}B.$$

PROOF: We begin with choosing two systems of parameters. The first will be a system of parameters $x_1,\dots,x_r$ for the maximal ideal $\mathfrak{m}$, and the second one for the ideal $\mathfrak{n}/\mathfrak{m}B$ in the ring $B/\mathfrak{m}B$. Let $y_1,\dots,y_s$ be a lifting to B of the latter. We contend that the ideal $\mathfrak{a} = (\phi(x_1),\dots,\phi(x_r),y_1,\dots,y_s)$ generated in B by the two systems is $\mathfrak{n}$-primary;

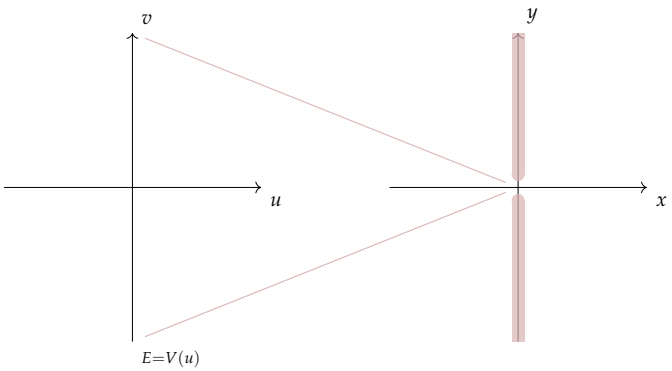
indeed, some power $\mathfrak{n}^\nu$ is contained in $(y_1, \ldots, y_s) + \mathfrak{m}B$, and consequently some higher power lies in $\mathfrak{a}$ since high powers of $\mathfrak{m}$ lie in $(x_1, \ldots, x_r)$. ❑

EXAMPLE 11.13 *Strict inequality may occur—Affine blow up:* Consider the map $\psi\colon \mathbb{C}^2 \to \mathbb{C}^2$ sending a point (u, v) to (u, uv). The fibre over $(0, 0)$ is the entire line $u = 0$, and thus of dimension strictly larger than the difference between the dimensions of the source and the target.

Transcribing this into the local algebra in the context of Proposition 11.30 we consider the map of rings $\phi\colon k[x, y] \to k[u, v]$ that sends $x \mapsto u$ and $y \mapsto uv$ (think of x and y as coordinates in the target $\mathbb{C}^2$). To set up the appropriate localizations a we let $\mathfrak{n} = (u, v)$ be the ideal of the origin in the source $\mathbb{C}^2$ and $B = k[u, v]_\mathfrak{n}$ the local ring there; similarly, we put $\mathfrak{m} = (x, y)$, the ideal of the origin in target $\mathbb{C}^2$, and let $A = k[x, y]_\mathfrak{m}$. One verifies that $\mathfrak{m}B = (u, uv) = (u)$, and that the "fibre" $B/\mathfrak{m}B$ is given as $B/\mathfrak{m}B = (k[u, v]/(u, uv))_{(u,v)} = k[v]_{(v)}$. It is of dimension one, whereas, of course, $\dim B = \dim A$.

The map ψ restricts to a bijection between $\mathbb{C}^2\backslash V(u)$ and $\mathbb{C}^2\backslash V(x)$, the complements of respectively the v-axis and he y-axis: indeed, v may be recovered from uv whenever $u \neq 0$. The most dramatic effect of ψ is to collapse the v-axis to one point, the origin in the target $\mathbb{C}^2$. For this reason ψ is called a "blow–down", or when seen in the perspective from the target $\mathbb{C}^2$, a "blow–up". ☆

EXERCISE 11.10 With the notation of Example 11.13 above, show that the map ϕ induces an isomorphism between $k[x, x^{-1}, y]$ and $k[u, u^{-1}, v]$. Show that the set-theoretical image of ψ equals $\mathbb{C}^2\backslash V(x)$. Translate this into a statement about prime ideals and the map ϕ, and verify it. Let $a \in k$ be a scalar. Give a description the inverse image $\phi^{-1}(v - a)k[u, v]$ of the principal ideal $(v - a)k[u, v]$ under ϕ (corresponding to the image of the line $v = a$ under ψ), and interpret this geometrically. ★



EXERCISE 11.11 Assume that A and B are domains that are finitely generated over a field k and let $\phi\colon A \to B$ be an injective k-algebra homomorphism.

a) Show that $\dim B \leqslant \dim A + \dim B/\mathfrak{m}B$ for any maximal ideal $\mathfrak{m}$ in A such that $\mathfrak{m}B$ is a proper ideal.
b) Show that there is an element $f \in A$ so that $\dim B = \dim A + \dim B/\mathfrak{m}B$ for all maximal ideals $\mathfrak{m} \in D(f) \subseteq \operatorname{Spec} A$.

Interpretation this in a geometric language. ★

11.5 *Dimension of polynomial rings*

We turn to the question about what the relation between the dimension of a ring A and of the polynomial ring $A[t]$ is. The intuitive, and naive, guess would be that the polynomial ring is of dimension one more than A; this is certainly supported by polynomial rings $k[x_1, \dots, x_n]$ being of dimension n, and more generally, by being true whenever A is finitely generated over a field. But nature is not that simple, and thist does not hold true universally; some non-Noetherian rings show their treacherous character in this respect.The dimension of $A[t]$ can in fact be as large as $2 \dim A + 1$. However, Noetherian rings behave well, and for those $\dim A[t] = \dim A + 1$, as we soon shall see. The proof will be a direct application of the dimension-inequality from Proposition 11.30 on page 300.

Of course, for algebras finite generated over a field we have already shown theses results, but we find this "doubling" justified since many important classes of Noetherian rings, for instance those being finitely generated over $\mathbb{Z}$ and the complete Noetherian rings, are not of finite type over a field.

We begin by studying how the prime ideals in $A[t]$ are related to the prime ideals in A, or in other words, we shall take a closer look at the fibres of the projection map $\pi\colon \operatorname{Spec} A[t] \to \operatorname{Spec} A$ induced by the inclusion $A \subseteq A[t]$.

The fibres of the projection map

(11.31) Each prime ideal $\mathfrak{p}$ in A gives rise to an ideal $\mathfrak{p}^+ = \mathfrak{p}A[t]$, which is formed by the polynomials $\sum_i a_i t^i$ having all coefficients in $\mathfrak{p}$. Clearly $\mathfrak{p}^+ \cap A = \mathfrak{p}$.

Our next observation is that the ideal $\mathfrak{p}^+$ is a prime ideal. Since the polynomial ring $A/\mathfrak{p}[t]$ is an integral domain, this follows *e.g.* from the isomorphism

$$A[t]/\mathfrak{p}A[t] \simeq A/\mathfrak{p}[t], \tag{11.1}$$

which is induced by the map $A[t] \to A/\mathfrak{p}[t]$ that sends $\sum a_i t^i$ to $\sum [a_i] t^i$. This is obviously a surjective homomorphism whose kernel is easily seen to be $\mathfrak{p}A[t]$: a polynomial $\sum a_i t^i$ is mapped to zero if and only if $[a_i] = 0$ for all i; that is, if and only all a_i lie in $\mathfrak{p}$.

(11.32) The prime ideal $\mathfrak{p}^+$ is one of the ideals that contract to $\mathfrak{p}$, or in geometric terms, one which belongs to the fibre over $\mathfrak{p}$ of the map $\pi\colon \operatorname{Spec} A[t] \to \operatorname{Spec} A$. But there are others; in fact, we shall see there are infinitely many. To begin with we consider the case that A is a domain and $\mathfrak{p} = 0$, to which the general case subsequently will be reduced.

LEMMA 11.33 *Let A be a domain with fraction field K. Sending $\mathfrak{p}$ to $\mathfrak{p}K[t]$ sets up a one a one to one correspondence between non-zero prime ideals $\mathfrak{p}$ in $A[t]$ such that $\mathfrak{p} \cap A = 0$ and the non-zero proper ideals in $K[t]$.*

PROOF: Observe that $K[t]$ equals the localisation $S^{-1}A[t]$ where S is the multiplicative set $S = A\backslash\{0\}$. The lemma is then just Proposition 7.20 on page 182 on primes in localizations. ❑

In general, if $\mathfrak{q} \cap A = \mathfrak{p}$ it holds that $\mathfrak{p}^+ \subseteq \mathfrak{q}$ so prime ideals in $A[t]$ that intersects A in $\mathfrak{p}$ are in one-to-one correspondence with prime ideals in $A[t]/\mathfrak{p}A[t]$ whose intersection with $A/\mathfrak{p}$ in the zero-ideal, or in other words prime ideals in the polynomial ring $A/\mathfrak{p}[t]$ that intersects $A/\mathfrak{p}$ in zero. Hence by replacing A by $A/\mathfrak{p}$ we place ourselves in the situation of the previous lemma, and with the *ad hoc* notation $K_\mathfrak{p}$ for the fraction field of $A/\mathfrak{p}$, we arrive at the following description of the fibre of π over $\mathfrak{p}$:

LEMMA 11.34 *Let A be a ring and let $\mathfrak{p} \subseteq A$ be a prime ideal. There is a one-to-one inclusion preserving correspondence between the lattice of primes in $A[t]$ contracting to $\mathfrak{p}$ and the lattice of prime ideals in the polynomial ring $K_\mathfrak{p}[t]$.*

Examples

(11.14) One of the simplest examples of the set-up in the lemma above, is the inclusion $\mathbb{C}[x] \subseteq \mathbb{C}[x,y]$, whose geometric incarnation we denote by π: Spec $\mathbb{C}[x,y] \to$ Spec $\mathbb{C}[x]$. The maximal ideals in $\mathbb{C}[x,y]$ are of the form $(x-a, y-b)$ and constitute a $\mathbb{C}^2$, while those in $\mathbb{C}[x]$ are shaped like $(x-a)$ and constitute a $\mathbb{C}$. Restricted to the set of maximal ideals the map π is, of course, just the first projection $\mathbb{C}^2 \to \mathbb{C}$; indeed, $(x-a, y-b) \cap \mathbb{C}[x] = (x-a)$ (trivial since $(x-a)$ is maximal in $\mathbb{C}[x]$).

The fibre over $(x-a)\mathbb{C}[x]$ consists of the point* $(x-a)^+ = (x-a)\mathbb{C}[x,y]$, whose vanishing locus is the line $x = a$, and the maximal ideals $(x-a, y-b)$ corresponding to points on that line—these are all, since the irreducible polynomials in $\mathbb{C}[y,x]/(x-a) \simeq \mathbb{C}[y]$ are linear. The fibre of π over the zero ideal (0) comprises all the principal prime ideals, *i. e.* those shaped like $(f(x,y))$ with f irreducible, and additionally the zero ideal.

*This is often called generic point of the fibre

It is noteworthy that the set Spec $\mathbb{C}[x,y]$ is not equal to the Cartesian product of the sets Spec $\mathbb{C}[x]$ and Spec $\mathbb{C}[y]$— there is no room for the principal primes in the latter—whereas the set of maximal ideals equals the product of the two sets of maximal ideals.

(11.15) Back in Paragraph 3.29 on page 76 we took a look at Spec $\mathbb{Z}[t]$. If $(p) \in$ Spec $\mathbb{Z}$ is a prime ideal generated by a prime p, then $(p)^+$ is just the principal ideal $(p)\mathbb{Z}[t]$. The fibre of π over (p) is the closed set $V((p))$ consisting of prime ideals containing p. A part from $(p)\mathbb{Z}[t]$ itself, it comprises all maximal ideals of shape $(p, f(t))$ where $f(t)$ is a polynomial in $\mathbb{Z}[t]$ which is irreducible mod p; that is, its class in $\mathbb{F}_p[t]$ generates

a maximal ideal. Hence the fibre over (p) is homeomorphic to Spec $\mathbb{F}_p[t]$. The fibre of π over the zero ideal (0) in $\mathbb{Z}$ consists of the principal ideals $(f(t))\mathbb{Z}[t]$ generated by irreducible polynomials. The polynomials persist being irreducible in $\mathbb{Q}[t]$ (*e.g.* by Gauss's Lemma) so that fibre is homeomorphic to Spec $\mathbb{Q}[t]$.

Discrete valuation rings (diskrete valuasjonsringer)

(11.16) Of quite another flavour is the ring $A[t]$ when A is a so-called *discrete valuation ring*; that is, a local Noetherian domain of dimension one whose maximal ideal $\mathfrak{m}$ is principal, say generated by π. The localizations $k[t]_{\mathfrak{p}}$ and $\mathbb{Z}_{(p)}$ are instances of such.

The space Spec A has two points $\mathfrak{m}$ and (0). The fibre over $\mathfrak{m}$ is Spec $k[t]$ where k denotes the residue field $k = A/\mathfrak{m}$, and the fibre over (0) equals Spec $K[t]$ with K being the fraction field of A.

Notice that, unlike in the two preceding examples, the "generic fibre", *i. e.* the fibre over (0), contains *maximal* ideals, namely the principal ideals $(f(t))$ that are generated by polynomials $f(t)$ which are invertible modulo π; that is, they can be brought on of the form $\pi g(t) + 1$.

See also Proposition 3.30 on page 77 and Exercise 3.8.

☆

Exercises

(11.12) Let A be a discrete valuation ring (as defined in example 11.16 above). Show that a non-zero prime ideal in $A[t]$ is of one of the four types:

i) If $\mathfrak{p}$ is maximal and belongs to the fibre over $\mathfrak{m}$; that is, $\mathfrak{p} \cap A = \mathfrak{m}$, then $\mathfrak{p} = (x, f)$ with $f \in A[t]$ being irreducible mod x.
ii) If $\mathfrak{p}$ is prime and belongs to the fibre over $\mathfrak{m}$, then $\mathfrak{p} = \mathfrak{m}^+ = \mathfrak{m}A[t]$.
iii) If $\mathfrak{p}$ is maximal lies in the generic fibre; *i. e.* $\mathfrak{p} \cap A = (0)$, then $\mathfrak{p} = (f)$ with $f \in A[t]$ primitive and irreducible and being a unit mod x.
iv) If $\mathfrak{p}$ is prime and in the generic fibre but not maximal, then $\mathfrak{p} = (f)$ with f primitive and irreducible but not a unit mod x.

(11.13) Describe the fibres of the canonical map $\pi\colon$ Spec $\mathbb{C}[\![x]\!][t] \to$ Spec $\mathbb{C}[\![x]\!]$. If $\mathbb{C}\{x\}$ denotes the ring of convergent power series with complex coeffcients, describe the fibres of $\pi\colon$ Spec $\mathbb{C}\{x\}[t] \to$ Spec $\mathbb{C}\{x\}$.

(11.14) Describe the fibres of the canonical map $\pi\colon$ Spec $\mathbb{Z}_2[t] \to$ Spec $\mathbb{Z}_2$.

★

Bounds on the dimension of the polynomial ring

(11.35) The fibre over $\mathfrak{p}$ of the projection map Spec $A[t] \to$ Spec A is thus canonically homeomorphic to Spec $K_{\mathfrak{p}}[t]$. So in addition to $\mathfrak{p}^+$, we find infinitely[1] many other prime

[1] Non-zero prime ideals in $K_{\mathfrak{p}}[t]$ are generated by irreducible polynomials of which there are infinitely many

ideals in $A[t]$ contracting to $\mathfrak{p}$. They all contain $\mathfrak{p}$, but there is no inclusion relation between any of them.

PROPOSITION 11.36 *The inequalities* $\dim A + 1 \leqslant \dim A[t] \leqslant 2\dim A + 1$ *hold true for any ring* A.

PROOF: In any chain $\{\mathfrak{p}_i\}$ in $A[t]$ contracting to a chain $\{\mathfrak{p}_i\}$ in A at most two members contract to each $\mathfrak{p}_i$. Hence the length of $\{\mathfrak{p}_i\}$ is at most $2\dim A + 1$. On the other hand, if $\{\mathfrak{p}_i\}$ is a chain in A of length r, the chain $\mathfrak{p}_i^+$ will be of length r since each $\mathfrak{p}_i{}^+$ intersects back to $\mathfrak{p}_i$. This gives $\dim A[t] \geqslant r$, but there are prime ideals in $A[t]$ strictly greater than $\mathfrak{p}_r^+$ which may be joined to the chain, and hence $\dim A[t] \geqslant r + 1$. ❑

Abraham Seidenberg gave examples of non-Noetherian rings showing that $\dim A[t]$ may take any value between $\dim A$ and $2\dim A + 1$. Below we reproduce one of these examples (Example 14.2 on page 359), which in fact goes back to Krull, describing a ring with $\dim A = 1$, but such that $\dim A[t] = 3$.

The case of Noetherian ring

Noetherian rings are in this context well behaved. One has

THEOREM 11.37 *Let* A *be a Noetherian ring, and* $\mathfrak{p} \subseteq A$ *a prime ideal. Then* $\operatorname{ht}\mathfrak{p}^+ = \operatorname{ht}\mathfrak{p}$. *In particular, it holds true that* $\dim A[t] = \dim A + 1$.

PROOF: We merely need to prove that the inequality $\operatorname{ht}\mathfrak{p}^+ \leqslant \operatorname{ht}\mathfrak{p}$ holds for all prime ideals $\mathfrak{p}$ in A; this implies the equalities $\operatorname{ht}\mathfrak{p}^+ = \operatorname{ht}\mathfrak{p}$ and $\dim A[t] \leqslant \dim A + 1$. Indeed, if $\mathfrak{m}$ is a maximal ideal in $A[t]$ and $\mathfrak{p} \subseteq \mathfrak{m} \cap A$, it holds that $\operatorname{ht}\mathfrak{m} = 1 + \operatorname{ht}\mathfrak{p}^+ = 1 + \operatorname{ht}\mathfrak{p} \leqslant \dim A + 1$. And, as shown in the proof of Proposition 11.36 above, any chain in A induces a chain in $A[t]$ of the same length, we have $\operatorname{ht}\mathfrak{p}^+ \geqslant \operatorname{ht}\mathfrak{p}$.

Replacing A by $A_\mathfrak{p}$ we may assume that A is local, and designating the maximal ideal by $\mathfrak{m}$, we are to show

$$\dim A[t]_{\mathfrak{m}^+} \leqslant \dim A.$$

Bearing the inequality (11.30) in mind (with $B = A[t]_{\mathfrak{m}^+}$) we will be trough once we prove that $\dim A[t]_{\mathfrak{m}^+}/\mathfrak{m}A[t]_{\mathfrak{m}^+} = 0$. But this holds since $A[t]_{\mathfrak{m}^+}/\mathfrak{m}A[t]_{\mathfrak{m}^+}$ is isomorphic to the rational function field $k(t)$ over the residue field $k = A/\mathfrak{m}$: reducing coefficients modulo $\mathfrak{m}$ gives an isomorphism $A[t]/\mathfrak{m}A[t] \simeq k[t]$, as in (11.1), under which the multiplicative system $S = A[t]\backslash\mathfrak{m}^+$ maps to the multiplicative set $k[t]\backslash\{0\}$ in $k[t]$. Hence the induced map between the localizations yields the desired isomorphism. ❑

Algebras finitely generated over fields again

(11.38) From Theorm 11.37 we deduce by an obvious induction that the polynomial ring $k[x_1, \ldots, x_n]$ is of dimension n, and gives another proof of this important fact. And

similarily, we see that $\dim \mathbb{Z}[x_1, \ldots, x_n] = n + 1$. More over, that $\operatorname{ht} \mathfrak{p}^+ = \operatorname{ht} \mathfrak{p}$ yields the stronger result, that maximal ideals are all of the same height, equal to the dimension.

THEOREM 11.39 *If A is a domain finitely generated over a field, or over a one dimensional Noetherian domain R with infinitely many prime ideals. Then all maximal ideals in A are of height equal to* $\dim A$.

The proof of the theorem relies on the following nice proposition from Kaplansky's book about elements generating the fraction field of a domain:

PROPOSITION 11.40 *Let A be a domain with fraction field K, and u an element in A. Then the following three statements are equivalent:*

i) The element u is contained in all non-zero prime ideals in A;
ii) Every non-zero ideal in A contains a power of u;
iii) $K = A[u^{-1}]$.

The only Noetherian rings for which the three statements hold, are the semi–local rings of Krull dimension at most one.

PROOF: That two first statements are equivalent, is clear as the radical of any ideal is the intersection of the prime ideals containing it, so let us see that *ii)* implies *iii)*. To that end, pick any non-zero element $x \in A$. By *ii)* we may find a natural number r so that $u^r \in (x)$; in other words, it holds true that $x^{-1} = yu^{-r}$ for some $y \in A$, and hence $x^{-1} \in A[u^{-1}]$. To establish the converse implication assume that $K = A[u^{-1}]$, and let $\mathfrak{a}$ be a non-zero ideal. For every non-zero element $x \in \mathfrak{a}$ it holds that $x^{-1} = yu^{-r}$ for some u and some r; that is, $u^r \in (x) \subseteq \mathfrak{a}$.

To prove the last statement, note that any prime ideal in A of height one is minimal over u, and there is only finitely many such when A is Noetherian (Proposition 9.17 on page 235). Furthermore if a prime ideal in A were of height two, it would according to Proposition 11.17 on page 296 contain infinitely many prime ideals of height one, so there are not any. Hence A is of dimension one. Then all non-zero prime ideals are maximal and of height one, and as we have seen, there are only finitely many. ❑

LEMMA 11.41 *Assume that A is a Noetherian domain and that B is a domain containing A which is finitely generated as an A-algebra. If the spectrum* Spec B *is finite, the spectrum* Spec A *is also finite.*

PROOF: By descending induction on the number of generators B requires, it suffices to do the case that $B = A[u]$. The ring $A[u]$ is Noetherian (Hilbert's Basis Theorem) and is of dimension at most one. Here the proof bifurcates:

1) The case that $A[u]$ is field. Then u is invertible in $A[u]$, and one easily sees that u^{-1} is integral over A: indeed, there is an expression

$$u^{-1} = (a_n u^n + \ldots + a_1 u + a_0)$$

with $a_i \in A$, which gives an integral dependence relation for u^{-1} when multipled with u^{-n}. Applying Proposition 11.40 above to $R = A[u^{-1}]$, which is Noetehrian by Hilbert's Basis Theorem, we deduce that Rhas finitely many prime ideals. Now R is integral over A, so by Lying–Over also A has only finitely many prime ideals.

2) The case that $A[u]$ is not a field. Being Noetherian and having just a finite number of prime ideals, it is of dimension one. The element u cannot be transcendental, for then A would be a field and in the polynomial ring over a field there are manifestly an infinite number of prime ideals. Hence $\dim A = 1$ and u is algebraic. Thus there is a relation

$$a_n u^n + \ldots + a_1 u + a_0 = 0$$

with $a_i \in A$. The coeficient a_n is contained in finitely many prime ideals, which are exactly those that disappear in the localized ring A_{a_n}. Moreover, $A[u]_{a_n}$ is integral over A_{a_n}. Then again by Lying–Over, the ring A_{a_n} has only finitely many prime ideals since this is the case for $A[u]_{a_n}$, and consequently A has only finitely many prime ideals. ❑

(11.42) Fiannly we are prepared for the proof of theorem 11.39:

Proof of Theorem 11.39: It suffices to see that conclusion holds for polynomial rings over R. And by induction on the number of variables, it will be enough to prove that if the theorem holds for A, it holds for the polynomial ring $A[t]$.

We saw above that $\operatorname{ht} \mathfrak{p}^+ = \operatorname{ht} \mathfrak{p}$, and it therefore suffices to show that for all maximal ideals $\mathfrak{m}$ in $A[t]$ the intersection $\mathfrak{p} = \mathfrak{m} \cap A$ is a maximal ideal in A; indeed, there is no other prime in the fibre of π over $\mathfrak{p}$ than $\mathfrak{m}$ which contains $\mathfrak{p}^+$. Replacing A with $A/\mathfrak{p}$ we may assume that A is a domain and $\mathfrak{m} \cap A = 0$, and the task is then to show that A is a field.

Denote the field $A[t]/\mathfrak{m}$ by K and let $u \in K$ be the inverse of the class $[t]$; that is, $u = [t]^{-1}$ (we may assume that t does not belong to $\mathfrak{m}$, for if it did, we would have $A = A[t]/\mathfrak{m}$). There is an inclusion $A \subseteq K$ and $K = A[u^{-1}]$.

The next observation is that u integral over A; indeed, since $u \in K$ and K is generated by u^{-1} over A, there must be a relation like $u = a_0 + \ldots + a_r u^{-r}$ with $a_i \in A$ and $a_r \neq 0$, and it yields an integral relation for u when multiplied by u^r. By Lemma 12.28 on page 327, it suffices therefore to show that $A[u]$ is a field.

Certainly as $A[u^{-1}]$ is a field, it equals the fraction field of $A[u]$, so when we let Lemma 11.40 above come into play, we may conclude that $A[u]$ semi-local ring of diemension at most one. If R is a field, the intersection of all maximal ideals in $A[u]$ is zero, so $A[u]$ is a field. Otherwise R, if R does not map injectively into $A[u]$ the image

is a field, $A[u]$ is a field by what just did.Finally, if R is contained in $A[u]$ Lemma 11.41 yeids that R is semi-local and it isn't. ❑

Appendix: The trace and the norm

In this appendix the stage is set by a finite extension of domains $A \subseteq B$ with A being integrally closed in its fraction field K. The fraction field L of B is then a finite extension of K whose degree we shall denote by n. In particular, L is a vector space over K of dimension n.

(11.43) With every element a in L we may associate the multiplication map $[a]\colon L \to L$ defined by the assignment $x \mapsto ax$. It certainly is K-linear, and we may apply linear algebra when studying it. As any linear endomorphism, the multiplication map $[a]$ has a *characteristic polynomial*, which is given as

The characteristic polynomial (det karakteristiske polynomet)

$$P_a(t) = \det(t \cdot I - [a]) = t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0.$$

The trace of elements (sporet til et element)
The norm of an element (normen til et element)

The coefficients a_i are interesting invariants of the element a, in particular the constant term and the subleading term which up to sign are the trace and the norm of a; more precisely, the *trace* is defined as $\mathrm{tr}_{L/K}(a) = -a_{n-1}$ and the *norm* is given as $N_{L/K}(a) = (-1)^n a_n$. Observe that the trace is K-linear in a, and the norm, which in fact is nothing but the determinant $\det[a]$, depends multiplicatively on a.

With a there is also associated another polynomial, the *minimal polynomial* $m_a(t)$ over K, which is the monic polynomial of lowest degree such that $m_a(a) = 0$; or, if you want, the monic generator of the ideal in $K[t]$ consisting of polynomials having a as root. From elementary field theory we know that $K(x) = K[t]/(m_a)$, and hence the degree of m_a coincides with $[K(a) : K]$.

(11.44) From the Cayley-Hamilton theorem ensues that a is a root of P_a, and therefore P_a will always have the minimal polynomial m_a as factor. In most cases the two differ, the exception being when L is a primitive extension of K with a as a generator; that is, when $L = K(a)$. Both polynomials will then be of degree n and must agree. However, in all cases the two are closely related; the characteristic polynomial P_a is always a power of the minimal one m_a:

LEMMA 11.45 (CHARACTERISTIC AND MINIMAL POLYNOMIALS) *The characteristic polynomial of a is a power of the minimal polynomial; more precisely, $P_a = m_a^r$ where $r = [L : K(a)]$.*

PROOF: Let $x_1, \ldots, x_r$ be a basis for $K(a)$ over K and $y_1, \ldots, y_s$ one of L over $K(a)$ chosen so that $x_1 = y_1 = 1$. The products $x_i y_j$ constitute a basis for L over K, and this leads to a decomposition of L as a K-vector space shaped like:

$$L = K(a) \oplus K(a)y_2 \oplus \ldots \oplus K(a)y_s,$$

Clearly each of the subspaces $K(a)y_i$ is invariant under multiplication by a, Even more is true: multiplication by y_i is an isomorphism from $K(a)$ to $K(a)y_i$ which commutes with the multiplication map $[a]$; hence the characteristic polynomial of $[a]$ acting on $K(a)$ and the one of $[a]$ acting on $K(a)y_i$ are equal. But as we observed just before the lemma, the characteristic polynomial of $[a]$ acting on $K(a)$ coincides with the minimal polynomial m_a.

Finally, it is a general fact that if θ is an endomorphism of a vector space having a direct sum decomposition $\bigoplus V_i$ with each summand invariant under θ, the characteristic polynomial of θ equals the product of the characteristic polynomials of θ acting on each direct summand; *i. e.* of the restrictions $\theta|_{V_i}$. In our case this gives the desired equality $P_a = m_a^r$, which closes the proof. ❑

COROLLARY 11.46 *Assume that$K \subseteq L$ is a field extension and $a \in K$ an element. Then it holds that*

i) $\operatorname{tr}_{L/K} a = [L : K(a)] \operatorname{tr}_{K(a)/K} a$;

ii) $N_{L/K}(a) = \big(N_{K(a)/K}(a)\big)^{[L:K(a)]}$.

(11.47) One of main properties of trace and norm (or any of the other symmetric functions of the eigenvalues, for that matter) is that they take integral elements to integral elements:

PROPOSITION 11.48 *Let A be a domain with fraction fied K and let $K \subseteq L$ be an extension of fields. Assume that $a \in L$ is an element integral over A, then $\operatorname{tr}_{L/K}(a)$ and $N_{L/K}(a)$ are integral over A as well.*

PROOF: In an algebraic closure of K, the polynomial P_a splits into a product of linear factors, the roots being the eigenvalues λ_i of $[a]$. If the element $a \in L$ is integral over A, it satisfies an integral dependence relation

$$f(t) = t^d + a_{d-1}t^{d-1} + \ldots + a_1 T + a_0 = 0,$$

where the coefficients a_i belongs to A. Assuming d is the minimal degree of such a relation, we infer that $p(t)$ equals the minimal polynomial of a; that is $f = m_a$.

By Lemma 11.45 above, the characteristic polynomial P_a is a power of m_a and therefore has coefficients in A. Consequently, the eigenvalues λ_i of $[a]$ being roots in P_a are integral over A as well. Now, the trace is the sum of the eigenvalues, and the norm equals their product, so these two are both integral over A (and belong to K). ❑

The trace

PROPOSITION 11.49 *Let $K \subseteq L$ be an extension of fields. Then the three properties following hold true that:*

i) *The trace is* $\mathrm{tr}_{L/K}$ *is K-linear, and* $\mathrm{tr}(1) = [L : K]$;
ii) *If* $a \in L$ *is an element one has* $\mathrm{tr}_{L/K} = [L, K(a)]\,\mathrm{tr}_{K(a)/K}$;
iii) *If* $K \subseteq L' \subseteq L$ *is an intermediate field, then* $\mathrm{tr}_{L'/K} \circ \mathrm{tr}_{L/L'} = \mathrm{tr}_{l/K}$;

(11.50) Well-known properties of linear maps translate into basic properties of the the norm and the trace:

LEMMA 11.51 *Let* $A \subseteq B$ *be an extension of domains such that the extension of their fraction fields* $K \subseteq L$ *is finite. For elements* $f, g \in L$, *it holds true that*

i) *The norm is multiplicative; that is,* $N(f \cdot g) = N(f) \cdot N(g)$ *and* $N(1) = 1$. *If* $f \in K$, $N(f) = f^{[L:K]}$;
ii) *The trace is K linear, and* $\mathrm{tr}(1) = [L : K]$;
iii) *When A is normal and* $f \in B$, *the element f is a factor of* $N(f)$; *that is,* $N(f) = fb$ *for some* $b \in B$

Be aware that the equality in *ii)* takes place in K; so that when K is a field of characteristic p and p divides the degree $[L : K]$, it holds that $\mathrm{tr}(1) = 0$.
PROOF: Since obviously $[fg] = [f] \circ [g]$ the norm is multiplicative as the determinant is, and if $f \in K$, the multiplication map $[f]$ is just f times the identity id_L. The trace is the sum the diagonal elements of the matrix of $[f]$ in any K-basis for L, from which the additivity ensues, and it also ensues that the trace of the identity equals the dimension of L over K.

For the third feature observe that if $P_f(t) = t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + (-1)^n N(f)$ is the characteristic polynomial of $[f]$, the Cayley–Hamilton theorem yields

$$N(f) = (-1)^{n+1} \cdot f \cdot (f^{n-1} + a_{n-1}f^{n-2} + \ldots + a_1). \tag{11.2}$$

and because the a_i's lie in A and f lie in B, the expression in parenthesis lies in B. ❑

Separable and inseparable polynomials

In the literature a polynomial is said to be *separable* if it has distinct roots in an algebraic closure $\bar{K}$. In a rather vague etymology the usage is some times explained by that the roots being distinct they can be separated.

There is standard way of detecting multiple roots of a polynomial is by use of the derivative. A root a of $f(t)$ is not simple precisely when the derivative $f'(t)$ vanishes at a as well. Indeed, one may write $f(t) = (t-a)^m g(t)$ where $g(a) \neq 0$. Leibnitz' rule then yields

$$f'(t) = m(t-a)^{m-1} + (t-a)^m g'(t),$$

and we see that $f'(a) = 0$ if and only if $m \geqslant 2$.

Obviously, a non-constant *irreducible* polynomial $f(t)$ cannot share a common factor with a (non-zero) polynomial of lower degree, so in particular if a were a multiple

root, $(x-a)$ would be a common factor of f and f', which would force f' to vanish identically.

Over fields of characteristic zero the only polynomials with vanishing derivative are the constants, so in that case an irreducible f has distinct roots and is hence separable. However, when K has characteristic p, this is not true any more. For instance, the power t^p has derivative pt^{p-1} which vanishes since $p = 0$ in K. More generally one finds applying the chain rule that

$$(f(t^{p^\nu}))' = f'(t^{p^\nu}) \cdot p^\nu t^{p^\nu - 1} = 0.$$

Hence polynomials shaped like $f(t^{p^m})$ have vanishing derivatives. And in fact these are all.

LEMMA 11.52 *Let $f(t)$ be a polynomial in $K[t]$ and assume that $f'(t)$ is the zero polynomial. Then K is of positive characteristic, say p, and $f(t) = g(t^{p^\nu})$ for some $g \in K[t]$ whose derivative is not identical zero, and $\nu \in \mathbb{N}_0$. If f is irreducible, g will be separable.*

PROOF: Write $f(t) = \sum_{i\in I} a_i t^i$ where $I \subseteq \mathbb{N}$ are the indices with $a_i \neq 0$. Then $f'(t) = \sum_{1\leqslant i\leqslant n} i \cdot a_i t^{i-1}$, and since powers of t are linearly independent, it follows that $ia_i = 0$ for all $i \in I$. Hence $i = 0$, that is i is divisible by p and we can write $i = p^{\nu_i} m_i$. Letting ν be the smaller of the ν_i's and $g(t) = \sum a_i t^{p^{\nu_i - \nu} m_i}$, we find $f(t) = g(t^{p^\nu})$. The exponent of the term in $g(t)$ with index i so that $\nu_i = \nu$, is not divisible by p, and hence g' is not identically zero. Finally, if g has a double root, it can not be irreducible, since $g' \neq 0$, hence neither f can be irreducible. ❑

PROPOSITION 11.53 *The trace form $\mathrm{tr}_{L/K}\, xy$ is non-degenerate if and only if all elements in L are separable over K.*

PROOF: By the Primitive Element Theorem there is an element a such that $L = K(a)$. Let $Q(t)$ be the minimal polynomial so that $L = K[t]/(Q(t))$. Separability means that the roots of $Q(t)$ in an algebraic closure $\bar{K}$ are distinct (they can be "separated"). It follows that $(Q(t)) = (t-\beta_1) \cap \ldots \cap (t-b_n)$ and The Chinese Remainder theorem gives an isomorphism a

$$L \otimes_K \bar{K} = \bar{K}[t]/(Q(t)) \simeq \prod_i \bar{K}.$$

Now, the multiplication map $[f]$ is a $\bar{K}$ map of $L\otimes_K \bar{K}$, and a basis $\{e_i\}$ induces a basis $e_i \otimes 1$ of $L\otimes_K \bar{K}$, so and the matrix of $[f]$ in the two are of course equal, and the trace $\mathrm{tr}_{L/K} \otimes id = \mathrm{tr}_{L\otimes_K \bar{K}/\bar{K}}$ is same whether f is considered a map of K or of $L\otimes_K \bar{K}$.

But the Chinese basis shows that blabla. . . . . ❑

Appendix (temporary): A geometric version of the Principal Ideal Theorem

We proceed to offer the promised alternative proof of the Hauptidealsatz valid for algebras finite type over fields, which David Mumford in his red book[**?**] on varieties contributes to John Tate. It relies on the Normalization lemma and uses the norm, so we begin the section with recalling a few properties of the norm (for details see the Appendix) together with an easy lemma about the transcendence degree.

(11.54) For any finite extension $A \subseteq B$ of domains with A integrally closed in its fraction field, one has the multiplicative norm map $N\colon B^* \to A^*$. If K and L designate the fraction fields of respectively A and B, the norm $N(x)$ of an element $x \in B$ is the determinant of the K-linear map $[x]\colon L \to L$ that is just the multiplication by x; *i. e.* it is the map that sends y to xy. We shall need three of basic properties of the norm all proved in detail in the Appendix.

i) The norm is multiplicative: $N(xy) = N(x)N(y)$;
ii) For elements $x \in A$ it holds that $N(x) = x^{[L:K]}$;
iii) The element x is a factor of $N(x)$; that is, $N(x) = yx$ for some $y \in B$.

The two first are well known properties of the determinant, and the last is a consequence of the Cayley-Hamilton theorem asserting that a linear map satisfies its characteristic polynomial.

(11.55) The crux of Tate's proof of the Principal Ideal Theorem is the following lemma:

LEMMA 11.56 *In the setting just described, it holds true that* $\sqrt{(x)} \cap A = \sqrt{\big(N(x)\big)}$ *for any element* $x \in B$.

PROOF: In view of x being a factor of $N(x)$, one inclusion is obvious, namely that $\sqrt{\big(N(x)\big)} \subseteq \sqrt{(x)} \cap A$. For the reverse inclusion, assume that $y \in \sqrt{(x)} \cap A$; that is, $y \in A$ is an element on the form $y^r = bx$ for some $b \in B$ and some number r. It then holds true that $y^{dr} = N(y^r) = N(b)N(x)$, where $d = [L : K]$, and hence y belongs to $\sqrt{\big(N(x)\big)}$. ❑

With theses preparations in place, the proof of the Principal Ideal Theorem is, after an initial reduction, reduced to a few lines, but of course, the burden of the proof is bore by the Normalization Lemma.

THEOREM 11.57 (GEOMETRIC PRINCIPAL IDEAL THEOREM) *Let k be a field and A a domain finitely generated k. If $f \in A$ is non-zero element and $\mathfrak{p}$ is a minimal prime ideal over (f), then* $\operatorname{trdeg}_k A/\mathfrak{p} = \operatorname{trdeg}_k A - 1$.

To ease the notation we have written $\operatorname{trdeg}_k A$ for the transcendence degree of the fraction field of a domain A. In view of Corollary 13.20 above, the conclusion might as well have been formulated as $\dim A/\mathfrak{p} = \dim A - 1$. Notice, that the conclusion

is significantly stronger than just saying $\operatorname{ht}\mathfrak{p} = 1$, and as we shortly shall see, almost effortlessly leads to A being catenary.

PROOF: The first part of the proof is a standard reduction to the case that $\sqrt{(f)}$ is a prime ideal. We let $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ be the minimal primes of (f), so that $\sqrt{(f)} = \mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_s$, and we may as well assume that $\mathfrak{p}_1$ is the particular one we study. Pick an element h from A that lies in $\mathfrak{p}_1$, but not in any of the others $\mathfrak{p}_i$'s. Then $\sqrt{(f)}A_h = \mathfrak{p}_1 A_h$ is a prime ideal. Now, $A_h/\mathfrak{p}_1 A_h = \big(A/\mathfrak{p}_1\big)_h$. Hence $A/\mathfrak{p}_1$ and $A_h/\mathfrak{p}_1 A_h$ have the same fraction field, and of course, A_h is finitely generated over k. We may thus replace A by A_h and assume that $\mathfrak{p} = \sqrt{(f)}$.

Let $n = \operatorname{trdeg}_k A$. By Noether's Normalization Lemma, there are algebraically independent elements $x_1, \ldots, x_n$ from A such that A is a finite module over the polynomial ring $R = k[x_1, \ldots, x_n]$, and then the norm map $N\colon A^* \to R^*$ is available. By Lemma 11.56 it holds true that $\mathfrak{p} \cap R = \sqrt{(f)} \cap R = \sqrt{\big(N(f)\big)}$. Now, $\sqrt{\big(N(f)\big)}$ is a height one prime ideal in a polynomial ring and is therefore principal, say generated by g. It follows that $A/\mathfrak{p}$ is finite over $R/(g)$ and hence has the same transcendence degree, and by Lemma 13.24 above the latter has transcendence degree $n-1$, ❑

Lecture 12

Integral extensions

From an algebraic point of view there is a huge difference between the ring of integers $\mathbb{Z}$ and the field of rationals $\mathbb{Q}$; we need only mention the primes. They are visible in $\mathbb{Z}$ as generators of the prime ideals, but in $\mathbb{Q}$ they are, at least from an algebraic point of view, on an equal footing with all the other non-zero elements, they are all units. When the exploration of number fields* began in the early 19^{th} century, an immediate want arose of subrings playing the role of the integers, and in which the deeper secrets of the field could be revealed. These rings were made up of the "integral elements" in the field, or more precisely what we soon shall be calling the elements integral over $\mathbb{Z}$.

*That is, finite field extensions of the rationals.

Integral extension are ubiquitous; they are found not only in number theory, but where ever commutative algebra is seriously used. In algebraic geometry, for instance, integrally closed (or normal) rings give rise to what is called normal varieties where the geometry of the codimension one subvarieties strongly influence the geometry of the entire space.

In topology one has the notion of "branched covering spaces"; that is, continuous maps $X \to Y$ between two topological spaces which are proper* and have discrete fibres (in particular they could be finite). The integral extensions are in some sense the algebraic counterpart of these, in that the map they induce between the spectra will have closely resembling properties, as expounded in the circle of ideas round the Cohen–Krull–Seidenberg theorems.

*That a map is proper means that inverse images of compact sets are compact

12.1 Definition and basic properties

Throughout this section we shall work with an extension of rings $A \subseteq B$. An element $x \in B$ is said to *integral* over A if it satisfies a monic relation

Integral elements (hele elementer)

$$x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = 0, \tag{12.1}$$

where the coefficients a_i are members of A. It is all-important that the leading coefficient be one, so there is heavy stress on the word *monic*, but of course, the leading coefficient

Integral dependence relation (helhetsrelasjon)

being invertible in A goes for the same. This distinguishes integral elements from their cousins, the algebraic elements, which satisfy similar equations, but with no constraint on the leadings coefficient. A relation like (12.1) is called an *integral dependence relation* for x over A.

Integral extension (hel utvidelse)
Integral closure (helavslutningen)

(12.1) If all elements in B are integral over A, one says that B is *integral over* A, or that B is an *integral extension* of A. The subset of B consisting of the elements integral over A is called the *integral closure* of A in B and is denotes by $\bar{A}$. Of course it depends on B, but to keep notation simple, we do not include a reference to B in the notation—the context will make it clear where the integral closure is taken. It is a basic, but slightly subtle fact shortly to be proven, that the integral closure is a ring. Finally, one says that A is *integrally closed* in B if $A = \bar{A}$; that is, if every element in B which is integral over A, belongs to A.

Integrally closed (helavsluttet)

Normal rings (normale ringer)
Normalization (normalisering)

(12.2) Both in algebraic geometry and algebraic number theory the integral closure of a domain A in its field of fractions is an important associate to the domain, and we'll denote it by $\widetilde{A}$ to distinguish it from all the crowd of integral closures. Domains being integrally closed in their field of fractions; that is, those satisfying $A = \widetilde{A}$, are called *normal*, and for a general domain $\widetilde{A}$ is sometimes called the *normalization* of A.

Examples

Integral closures and normalizations play an important role in algebraic geometry, which is particularly accentuated in the theory of curves. We'll illustrate this with the two simplest examples of curve singularities, or in algebraic parlance, two non-normal one-dimensional rings, an ordinary double point and a simple cusp. Typically for curves their normalizations "resolve singularities"; that is, it separates different branches of the curve passing through the multiple points (as in the case of the double point) or it resolves the vanishing of a derivative (as for the cusp).

Integral closures are of equally great significance in number theory where the ring of integers in number fields are the legendary stars of the theory. Our steadfast friends the quadratic extensions will serve as examples.

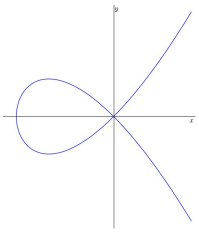
(12.1) To illustrate the difference between algebraic an integral dependence relation consider the two simple equations

$$y^2 - z = 0$$
$$(z-1)y^2 - z = 0$$

over the complex numbers. The first one "has $\sqrt{z}$ as a solution", but due to the ambiguity of the square root, it is impossible to find a continuous (yet alone analytic) solution in the entire plane. Only in simply connected domains not containing the origin can a continuous solution be found. The solutions of the second equation suffer

the same defect, but additionally they acquire a pole at $z = 1$. The difference between solutions of algebraic and integral relations is precisely the occurrence of poles in the former.

(12.2) *An ordinary double point:* We let $C \subseteq \mathbb{C}^2$ be the plane curve parameterized by $x = t^2 - 1$ and $y = t(t^2 - 1)$. It is easily seen to satisfy the equation $y^2 = x^2(x + 1)$; indeed, $(y/x)^2 = t = x + 1$. The real points of the curve is depicted in the margin.



For points on C with $x \neq 0$ the corresponding parameter value is uniquely defined, and the parameterization is one-to-one away from the origin. However, the two parameter values $t = \pm 1$ both give the origin, and the curve has a double point there.

The parameterization may be though of as the map Spec $\mathbb{C}[t] \to$ Spec $\mathbb{C}[x, y]$ induced by the ring-map $\mathbb{C}[x, y] \to \mathbb{C}[t]$ that sends $x \to t^2 - 1$ and $y \to t(t^2 - 1)$; or in the language of varieties, it is the the map $\mathbb{C} \to C = V(y^2 - x^2(x - 1)) \subseteq \mathbb{C}^2$ sending t to the point $(t^2 - 1, t(t^2 - 1))$.

This leads to considering the subring $A = \mathbb{C}[t^2 - 1, t(t^2 - 1)] \subseteq \mathbb{C}[t]$. The point of the example is that t is integral over A; indeed, almost tautologically it satisfies the equation

$$X^2 - t^2 = 0,$$

and $t^2 = (t^2 - 1) + 1 \in A$.

Moreover, the ratio between the two generators of A equals t, so that the fraction field of A equals $\mathbb{C}(t)$. And anticipating that the polynomial ring $\mathbb{C}[t]$ is normal (either Proposition 12.21 or 12.23 on page 324), we can conclude that $\mathbb{C}[t]$ is the normalization of A.

(12.3) *The simple cusp:* We also want to give an example from geometry, and the simplest example of a variety with a non-normal coordinate ring is the so-called *simple cusp*. It is the curve $C \subseteq \mathbb{C}^2$ whose equation is $y^2 = x^3$.

Simple cusp (enkel cusp eller enkel spiss)

One may parameterize C by using y/x as parameter; the parameterization being $t \mapsto (t^2, t^3)$. On the level of coordinate rings, the parameterization is reflected in the map $\mathbb{C}[x, y] \to \mathbb{C}[t]$ such that $x \mapsto t^2$ and $y \mapsto t^3$. The image is clearly the ring $A = \mathbb{C}[t^2, t^3] \subseteq \mathbb{C}[t]$. The fraction field of A equals the rational function field $\mathbb{C}(t)$ since $t = t^2 t^{-3}$ lies there. Now, A is not integrally close in $\mathbb{C}(t)$; indeed, t is integral over A being a root of the equation

$$T^2 - t^2 = 0,$$

where we note that $t^2 \in A$.

(12.4) *The golden section:* The *golden section* $(1 + \sqrt{5})/2$ is integral over $\mathbb{Z}$ since it satisfies the equation

The golden section (det gylne snitt)

$$x^2 - x - 1 = 0.$$

The integral closure of the integers $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{5})$ equals $\mathbb{Z}[(1+\sqrt{5})/2]$. Indeed, a number $\eta = a + b\sqrt{5}$ has the minimal equation

$$x^2 - 2ax + (a^2 - 5b^2) = 0, \tag{12.2}$$

and if η is integral over $\mathbb{Z}$, the coefficients of (12.2) are integral by Gauss's Lemma (Exercise 3.7 on page 78). Then $n = 2a \in \mathbb{Z}$ and $20b^2 \in \mathbb{Z}$, and hence $m = 2b \in \mathbb{Z}$ as well. Substituting back, gives $0 \equiv (n^2 - 5m^2) \equiv (n^2 - m^2) \mod 4$, which holds if and only if m and n have the same parity.

(12.5) The ring of integers in the number field $\mathbb{Q}(i\sqrt{5})$ equals $\mathbb{Z}[i\sqrt{5}]$. Indeed, the minimal equation of an element $a + ib\sqrt{5} \in \mathbb{Q}(i\sqrt{5})$ is

$$x^2 - 2ax + (a^2 + 5b^2) = 0.$$

If x is integral over $\mathbb{Z}$, the coefficients are integral, and as in the previous example, this entails that $n = 2a$ and $m = 2b$ are integral. Substituting back gives $0 \equiv n^2 + 5m^2 \equiv m^2 + n^2 \mod 4$, which occurs only if both n and m are even (squares are either equal to 1 or 0 $\mod 4$). Hence a and b are integers.

☆

Exercises

✲ (12.1) *Integers in quadratic number fields.* This is a classic from elementary number theory. The difference between the ring of integers in the two previous examples 12.4 and 12.5 illustrates a general phenomenon. Prove that if d is a square free integer, the ring of integers in $\mathbb{Q}(\sqrt{d})$ equals $\mathbb{Z}[(1+\sqrt{d})/2]$ when $d \equiv 1 \mod 4$, and $\mathbb{Z}[\sqrt{d}]$ else.

(12.2) *A not so simple cusp.* Let p be a natural number and consider the plane curve $y^2 = x^{2p+1}$. It may be parameterized by $t \mapsto (t^2, t^{2p+1})$; so the parameter is yx^{-p}. Let $A = \mathbb{C}[t^2, t^{2p+1}]$, let $\mathfrak{m}$ be the ideal (t^2, t^{2p+1}) in A, and let $\widetilde{A} = \mathbb{C}[t]$.

a) Show that $\mathbb{C}[x,y]/(y^2 - x^{2p+1}) \simeq \mathbb{C}[t^2, t^{2p+1}]$ and that $\mathfrak{m}$ is a maximal ideal;

b) Show that the fraction field of A equals the rational function field $\mathbb{C}(t)$ and that A is not normal;

c) Show that $\widetilde{A}/A$ is a cyclic A module of length p supported at the maximal ideal $\mathfrak{m}$.

★

The basic properties

(12.3) If x is an element of B which is integral over A, the subring $A[x]$ of B obtained by adjoining the element x to A is a *finitely generated* A-module. From an integral dependence relation as in (12.1) above, ensues the identity

$$x^n = -(a_{n-1}x^{n-1} + \ldots + a_1 x + a_0),$$

and a straightforward induction yields that $A[x]$ is generated by the n first powers of x. The converse of this is also true, as will be shown in the next proposition. When the A-module $A[x]$ is Noetherian, this comes almost for free; one just considers the ascending chain $M_i = A + xA + x^2A + \ldots + x^iA$ of submodules of $A[x]$, and at the point where it stabilises; that is, when $M_{\nu+1} = M_\nu$, one obtains an integral dependence relation for x since $x^{\nu+1} \in M_\nu$. The general proof requires however a little twist. Recall that a *faithful module* is one whose annihilator is the zero ideal.

PROPOSITION 12.4 (BASIC CHARACTERISATION) *Let $A \subseteq B$ be an extension of rings and let $x \in B$ be an element. The following three statements are equivalent:*

i) The element x is integral over A;
ii) $A[x]$ is a finitely generated A-module;
iii) There is a faithful $A[x]$-module which is finitely generated over A.

PROOF: The only implication showing any substantial resistance is that *i)* follows from *iii)*. So let M be a module as in *iii)*, and let $m_1, \ldots, m_n$ be generators for M over A. Each element $x \cdot m_i$ may be expressed in terms of the m_j's, and this gives relations

$$x \cdot m_i = f_{i1}m_1 + \ldots f_{in}m_n, \tag{12.3}$$

for $1 \leqslant i \leqslant n$, where all coefficient f_{ij} belong to A. We introduce an $n \times n$-matrix Φ by letting $\Phi = x \cdot I - (f_{ij})_{ij}$ where I is the $n \times n$-identity matrix (the matrix Φ for $n = 3$ is shown in a footnote)[1]. Equation (12.3) above then translate into the equality $\Phi \cdot m = 0$ where $m = (m_1, \ldots, m_n)$. Hence the determinant $\det \Phi$ kills M by the determinant trick (Lemma 4.61 on page 110), and we deduce that $\det \Phi = 0$ as M is a faithful A-module. But developing the determinant shows that $\det(x \cdot I - (f_{ij}))$ is a monic polynomial in x whose coefficients lie in A; that is, $\det \Phi = 0$ is an integral dependence relation for x over A. ❑

We notice the immediate corollary—which may also easily be proven *ad hoc*—that all elements in $A[x]$ are integral over A when x is. A faithful $A[x]$ module M which is finitely generated over A, will be faithful over any subring of $A[x]$, in particular over $A[z]$ for any $z \in A[x]$.

COROLLARY 12.5 *If x is integral over A, all elements in $A[x]$ are integral over A.*

(12.6) There is a close relationship between integral and finite extensions as unveiled in the previous proposition, but there are also significant differences. Finite extensions are integral, but in general the converse is not true. There are even examples of Noetherian

[1] For $n = 3$ the matrix Φ is shaped like $\begin{pmatrix} x - f_{11} & f_{12} & f_{13} \\ f_{21} & x - f_{22} & f_{23} \\ f_{31} & f_{32} & x - f_{33} \end{pmatrix}$

domains whose normalization $\widetilde{A}$ is not a finite module over A (we shall reproduce one in Paragraph 14.26 on page 369); they are however, rather exotic creatures, and the lions share of the rings appearing in mainstream algebraic geometry—that is, domains finitely generated over field and their localizations—have normalizations which are finitely generated as modules.

(12.7) The first conclusion to be drawn from the basic characterization of integral elements is that finitely generated algebras which are integral, are finitely generated modules; an important observations since the integral closure being finite over A or not, is an issue. One has

PROPOSITION 12.8 *Let B be an integral ring-extension of A. Any subalgebra C of B which is a finitely generated algebra over A, is a finite A-module.*

PROOF: The proof goes by induction on the number of generators of C. So let $C' \subseteq C$ be a subalgebra generated over A by the same elements as C but one, say x. By induction C' is a finitely generated module over A, and we have $C = C'[x]$. The element x being integral over A is even so more over C'. Hence C is finitely generated over C' by Proposition 12.4, and because being a finite extension is a property transitive in towers (Lemma 12.13 at the end of this subsection), it holds that C is finitely generated over A. ❑

COROLLARY 12.9 (TRANSITIVITY) *Assume that $A \subseteq B \subseteq C$ are ring-extension and that B is integral over A. Then every element in C which is integral over B is integral over A.*

PROOF: Let x be an element in C which is integral over B and satisfies the dependence relation

$$x^n + b_1 x^{n-1} + \ldots + b_{n-1} x + b_n = 0, \tag{12.4}$$

with the coefficients b_i lying in B. We let D be the sub A-algebra of B the b_i's generate. Then x is integral over D (the relation (12.4) has coefficients in D) and consequently $D[x]$ is a finite module over D. Now, D is a finite module over A after Proposition 12.8 above, and therefore $D[x]$ is finite over A as well. Hence we can conclude by the Basic Characterization 12.4 of integral elements that x is integral over A. ❑

(12.10) It is by no means obvious how to deduce a dependence relation for a product (or for a sum) from dependence relations for the factors (or the addends); that the integral closure is a ring, is a slightly subtle property. However, once the Basic Characterization (Proposition 12.4 on the preceding page) is in place, it follows readily. For a different approach, see Problem 12.7 on page 322.

PROPOSITION 12.11 (THE INTEGRAL CLOSURE IS A RING) *Assume that $A \subseteq B$ is an extension of rings. The sum and the product of two elements from B which are integral over A, are integral*

over A. The integral closure $\bar{A}$ of A in B is a subring of B. The integral closure $\bar{A}$ is integrally closed in B.

PROOF: This is just a combination of Corollary 12.5 and the transitivity property (Corollary 12.9). Indeed, let x and y be integral over A. The ring A[x] is an integral extension of A and y being integral over A, the extension $A[x,y]$ is integral over $A[x]$. Hence $A[x,y]$ is integral over A by transitivity. In particular, both the product $x \cdot y$ and the sum $x + y$ being members of $A[x,y]$ are integral over A, and $\bar{A}$ is a ring.

The last statement of the proposition might appear as a tautology, but an argument is in fact needed. We have to see that elements integral over $\bar{A}$ are integral over A, which is exactly what Corollary 12.9 above tells us since $\bar{A}$ is integral over A. ❑

(12.12) We end this subsection by proving the following lemma used in proof of Proposition 12.8 above:

LEMMA 12.13 (FINITE GENERATION IN TOWERS) *Let $A \subseteq B \subseteq C$ be a tower of rings and assume that C is a finite module over B and B is finite module over A, then C is a finite module over A.*

PROOF: Let $x_1, \ldots, x_r$ be a generating set for B as an A-module and $y_1, \ldots, y_s$ one for C over B. Then the products $x_i y_j$ will generated C over A. This is elementary: if $z = \sum b_j y_j$ with $b_j \in B$, write each coefficient b_j as $b_j = \sum_j a_{ij} x_i$ to obtain

$$z = \sum_j b_j y_j = \sum_j (\sum_i a_{ij} x_i) y_j = \sum_{i,j} a_{ij} x_i y_j.$$

❑

Integral extensions, localization and quotients

Integral extensions are well behaved in that they are compatible with the formation of localizations and quotients.

(12.14) We treat the loacalizations first:

PROPOSITION 12.15 *Let $S \subseteq A$ be a multiplicative subset and assume that B is an integral extension of A. Then $S^{-1}B$ is an integral extension of $S^{-1}A$. Forming the integral closure commutes with localization; i. e. it holds that $\overline{S^{-1}A} = S^{-1}\bar{A}$.*

PROOF: Let xs^{-1} be an element in B with $x \in A$ and $s \in S$. All elements of B are assumed integral over A, so x satisfies an integral dependence relation shaped like

$$x^n + a_{n-1}x^{n-1} + \ldots + a_0 = 0,$$

with the a_i's lying in A. Multiplying through by s^{-n} we find the relation

$$(xs^{-1})^n + a_{n-1}s^{-1}(xs^{-1})^{n-1} + \ldots + a_0 s^{-n} = 0, \tag{12.5}$$

which is a monic equation whose coefficients lie in $S^{-1}A$ and hence is an integral dependence relation for xs^{-1} over $S^{-1}A$.

For the second statement, the inclusion $S^{-1}\bar{A} \subseteq \overline{S^{-1}A}$ ensues from the first claim, and it suffices to prove the converse. To that end, assume that $xs^{-1} \in \overline{S^{-1}A}$. It satisfies an integral dependence relation as

$$(xs^{-1})^n + b_{n-1}(xs^{-1})^{n-1} + \ldots + b_0 = 0, \tag{12.6}$$

where each b_i lies in $S^{-1}A$ and hence may be written as $b_i = a_i t^{-1}$ with $a_i \in A$ and $t \in S$ (extending the fractions, we may use a common denominator for all the b_i's). Multiplying (12.6) through by $s^n t^n$ yields the relation

$$(tx)^n + a_{n-1}s(txs^{-1})^{n-1} + \ldots + s^n t^{n-1} a_0 = 0, \tag{12.7}$$

and it follows that tx is integral over A. We conclude that $xs^{-1} = (tx)s^{-1}t^{-1} \in S^{-1}\bar{A}$. ❑

(12.16) Next comes the quotients, and in addition to the usual staging of this chapter with an integral extension $A \subseteq B$, an ideal $\mathfrak{b}$ in B is given. We let $\mathfrak{a}$ be the ideal $\mathfrak{b}$ induces in A; that is, $\mathfrak{a} = A \cap \mathfrak{b}$. Then $A/\mathfrak{a} \subseteq B/\mathfrak{b}$ is an extension, which persists being integral:

PROPOSITION 12.17 *Let $\mathfrak{b} \subseteq B$ be an ideal and let $\mathfrak{a} = \mathfrak{b} \cap A$. If B is integral over A, then $B/\mathfrak{b}$ is integral over $A/\mathfrak{a}$.*

PROOF: Let $x \in B/\mathfrak{b}$ and chose an element $y \in B$ that maps to x. By assumption y is integral over A, and there is therefore a relation

$$y^n + a_{n-1}y^{n-1} + \ldots + a_1 y + a_0 = 0,$$

with the a_i's from A. Reducing that relation modulo $\mathfrak{b}$ we obtain the relation

$$x^n + [a_{n-1}]x^{n-1} + \ldots [a_1]x + [a_0] = 0,$$

where $[a_i]$ as usual denotes the classe of a_i in $A/\mathfrak{a}$. Hence x is integral over $A/\mathfrak{a}$. ❑

Exercises

(12.3) Let B be a ring and $\{B_i\}_{i\in I}$ a family of subrings of B. If each B_i is integrally closed in B, then the intersection $\bigcap_{i\in I} B_i$ is integrally closed as well.

⁂ (12.4) *New.* Let $A \subseteq B$ be an integral extension and let x be a variable. Show that the extension $A[x] \subseteq B[x]$ is integral.

(12.5) Let A be a normal domain and L an extension of the fraction field K of A. An element $x \in L$ is integral over A if and only if the minimal polynomial m_x of x has coefficients from A.

(12.6) Let $A \subseteq B$ be two domains. Show that $x \in B$ is integral over A if and only if there is a square matrix with coefficients in A having x as an eigenvalue.

(12.7) Show that if x and y are eigenvalues for Φ and Ψ then $x \cdot y$ is and eigenvalue for the Kronecker product $\Phi \otimes \Psi$ and that $x + y$ is one for the matrix $\Phi \otimes I_m + I_n \otimes \Psi$ where

the I_n are I_m are identity matrices of appropriate size. Conclude that the integral closure is a ring.

★

Being normal is a local property

As mention in the introduction to this section, a particular important situation is when A is a domain and $B = K(A)$ is the field of fractions of A. Recall that the integral closure of A in $K(A)$ is called the *normalization* of A, and in case A is integrally closed in $K(A)$, one says that A is a *normal domain*.

Normalizations (normaliseringer)
Normal domain (normalt område)

(12.18) In the previous section we showed that the two operations normalization and localization commute, and it ensues that A being normal implies that all localizations $A_\mathfrak{p}$ at prime ideals are normal. The converse also holds, so being normal is a local property:

PROPOSITION 12.19 (BEING NORMAL IS LOCAL) *Assume that A is a domain. Then A is normal if and only if $A_\mathfrak{m}$ is normal for all maximal ideals $\mathfrak{m}$ in A.*

PROOF: Notice first that all the localization $A_\mathfrak{m}$ have K as fraction field as well. Consider the inclusion $A \hookrightarrow \widetilde{A}$, which fits into the short exact sequence

$$0 \longrightarrow A \longrightarrow \widetilde{A} \longrightarrow \widetilde{A}/A \longrightarrow 0$$

of A-modules. Because localization is an exact functor, when localized at a maximal ideal $\mathfrak{m}$, the sequence gives rise to the short exact sequence

$$0 \longrightarrow A_\mathfrak{m} \longrightarrow (\widetilde{A})_\mathfrak{m} \longrightarrow (\widetilde{A}/A)_\mathfrak{m} \longrightarrow 0.$$

According to Proposition 12.15, forming integral closures commute with localization, so it holds that $(\widetilde{A})_\mathfrak{m} = (\widetilde{A_\mathfrak{m}})$. Thence $(\widetilde{A}/A)_\mathfrak{m} = (\widetilde{A_\mathfrak{m}})/A_\mathfrak{m}$, and the claim follows by the Localness of Being Zero (Proposition 7.47 on page 196). ❑

(12.20) In the friendly case that $\widetilde{A}$ is a finitely generated module over A, the quotient $\widetilde{A}/A$ is a finitely generated module over A as well, and it ensues that $\operatorname{Supp} A/\widetilde{A}$ is a closed subset of $\operatorname{Spec} A$; in fact, it equals $V(\operatorname{Ann} A/\widetilde{A})$.

Localizing at (0) (remember, A is a domain), or equivalently, tensorizing by K, we see that $\widetilde{A}/A \otimes_A K = 0$ because A and $\widetilde{A}$ both have K as fraction field. Hence $\widetilde{A}/A$ is not of global support. So in that benign case, when $\widetilde{A}$ is finite over A, for "most" primes $\mathfrak{p}$ the local ring $A_\mathfrak{p}$ is normal; that is, for primes in an open dense subset of $\operatorname{Spec} A$ the local rings $A_\mathfrak{p}$ are normal.

EXAMPLE 12.6 For example the normalization of $\mathbb{Z}[i\sqrt{5}]$ eqauls $\mathbb{Z}[(1+i\sqrt{5})/2]$, so in $\operatorname{Spec} \mathbb{Z}[i\sqrt{5}]$ the set of "non-normality" is the closed set $V((2))$. ☆

12.2 *Examples*

We indulge ourselves in a few examples of rather large classes of rings that are normal. The unique factorization domains are always normal, and the polynomial rings over normal domains are normal. The third class we shall investigate are the rings of invariants of actions of finite groups. It is a general principle that these rings are normal, at least when the ring acted upon is normal. This class includes all the so-call "quotient singularities". We shall treat a few simple examples in detail but leave the general case to the zealous students in the form of a guided exercise.

Factorial domains

A large and versatile class of domains that are normal are the UFD's:

PROPOSITION 12.21 *If the domain A is a* UFD*, then A is normal.*

PROOF: Let K be the fraction field of A, and let $z = x/y \in K$ be an element which is integral over A, and going for a contradiction, we assume that $z \notin A$. Reducing the fraction if need be, we may assume that x and y are without common factors. The element z being integral means that there is a relation

$$z^n + a_{n-1}z^{n-1} + \ldots + a_i z^i + \ldots + a_0 = 0,$$

with the a_i's lying in A. Multiplying through by y^n and rearranging the equation, gives

$$-x^n = a_{n-1}x^{n-1}y + \ldots + a_i x^i y^{n-i} + \ldots + a_0 y^n.$$

Every irreducible factor of y divides the right side, hence it divides the left side and consequently also x. Contradiction. ❑

Polynomial rings

PROPOSITION 12.22 *If $A \subseteq B$ is an intergral extension, then the extension $A[t] \subseteq B[t]$ of polynomial rings is integral.*

PROOF: Let $p \in B[t]$, and let $C \subseteq B$ be the A-subalgebra generated by the coefficients of p; it is integral over A and hence is a finite A-module. Obviously $C[t]$ is a finite $A[t]$-module having the same generators over $A[t]$ as C has over A, and by the Basic Characterization (Proposition 12.4 on page 319) p is integral over $A[t]$. ❑

PROPOSITION 12.23 *The polynomial ring $A[t]$ over a normal domain A is normal.*

PROOF: We let K be the fraction field of A. The polynomial ring $K[t]$ is normal with the same fraction field as $A[t]$, and therefore it suffices to see that $A[t]$ is integrally closed in $K[t]$. To that end, let $p(t)$ be a polynomial in $K[t]$ integral over $A[t]$, and assume that

$$p^n + f_1 p^{n-1} + \ldots + f_{n-1}p + f_n = 0,$$

where the f_i's belong to $A[t]$. We shall need a root field L of p containing K. Clearly each root of p in L is a root of f_n, so if f_n were monic, the roots would all be integral over A. The coefficients of f_n being polynomials in the roots would as well belong to A since A is normal by hypothesis; and consequently if also p were monic, we would have $p \in A[t]$. One can achieve this favourable situation simply by replacing p by $q = p - t^r$ for $r >> 0$. Indeed, a simple computation gives that $q = p - t^r$ satisfies a relation

$$q^n + g_1 q^{n-1} + \ldots + g_{n-1} q + g_n,$$

with "constant term"

$$g_n = t^{nr} + q_1 t^{r(n-1)} + \ldots + q_1 t^r + q_n = 0$$

which is monic when $r >> 0$, and hence q and thereby p, lies in $A[t]$. ❑

Rings of invariants

(12.24) Now comes the promised result on rings of invariants, and as promised, we shall proceed in a rather relaxed way merely treating the simples possible case. That is, the case of a cyclic group of order two acting on a normal domain B. Such an action is given by an *involution* on B; in other words, by a ring map $\sigma\colon B \to B$ satisfying $\sigma^2 = \mathrm{id}_B$. The map σ extends to an involution of the fraction field K of B by the obvious assignment $\sigma(xy^{-1}) = \sigma(x)\sigma(y)^{-1}$. Furthermore, we let $A = B^\sigma = \{\, x \in B \mid \sigma(x) = x \,\}$ be the ring of invariants and L its field of fractions. In this setting we have

Involutions (involusjoner)

PROPOSITION 12.25 *It holds true that $L = K^\sigma = \{\, z \in K \mid \sigma(z) = z \,\}$. Moreover B is integral over A and if B is normal, A will be normal.*

PROOF: Clearly $L \subseteq K^\sigma$. If $\sigma(x)/\sigma(y) = x/y$ it holds that $y\sigma(x) = x\sigma(y)$, and we may write $x/y = \sigma(x)x/\sigma(x)y$ with both $\sigma(x)x$ and $\sigma(x)y$ being invariant. Hence $L = K^\sigma$.

$$\begin{array}{ccccc} K^\sigma = & L & \subseteq & K \\ & \cup\!| & & \cup\!| \\ B^\sigma = & A & \subseteq & B \end{array}$$

Any element $x \in B$ satisfies the relation

$$x^2 - (\sigma(x) + x)x + \sigma(x)x = 0. \tag{12.8}$$

Both $\sigma(x) + x$ and $x\sigma(x)$ are invariant under σ and therefore belong to A, hence (12.8) is an integral dependence relation for x over A.

Finally, as B is integral over A, the integral closure of A in K equals $\widetilde{B}$ by transitivity, and hence $\widetilde{A} = \widetilde{B} \cap L$, from which ensues that $A = B \cap L = \widetilde{A}$ in the case that $B = \widetilde{B}$. ❑

EXAMPLE 12.7 *The quadratic cone again:* The coordinate ring $A = k[x, y, z]/(xy - z^2)$ of the quadratic cone is not factorial, as we have seen, but it is normal. It thus gives an example that the converse of Proposition 12.21 on the facing page is not valid. We shall exhibit A as the ring of invariants of the $\mathbb{Z}/2\mathbb{Z}$-action on $k[u, v]$ given by sign change of

both variables; a method that only works when the characteristic of k is different from two (even though the result remains true).

Sending $x \mapsto u^2$, $y \mapsto v^2$ and $z \mapsto uv$ induces a ring homomorphism from A onto the subring $k[u^2, uv, v^2]$ of the polynomial ring $k[u, v]$, and this map is an isomorphism: When x, y and z are given the weight two, A will be a graded algebra, and the map will be homogenous of degree zero. The kernel is therefore a homogeneous ideal. Assume that P is a homogeneous element in the kernel. Replacing z^2 by xy, one may write $P = Q(x, y) + R(x, y)z$, and since powers of u and v occurring in $Q(u^2, v^2)$ are even whereas those occurring in $R(u^2, v^2)uv$ are odd, it follows that both $Q(x, y)$ and $R(x, y)z$ lie in the kernel. Now, any homogeneous form in two variables splits as a product of linear forms, so if the kernel were non-zero, it would contain a linear form. But since u^2, uv and v^2 are linearly independent, this is not the case, and we can conclude that the map is injective. Obviously, it is surjective, hence it is an isomorphism. So define an action of the group $\mathbb{Z}/2\mathbb{Z}$ on $k[u, v]$ by letting the generator σ act by $u \mapsto -u$ and $v \mapsto -v$. Then $k[u^2, uv, v^2]$ will be the ring of invariants: Indeed, for a polynomial $p(u, v) = \sum a_{ij} u^i v^j$ one has

$$p(-u, -v) = \sum_{i+j \text{ even}} a_{ij} u^i v^j - \sum_{i+j \text{ odd}} a_{ij} u^i v^j,$$

and this equals $p(u, v)$ precisely when all terms with $i + j$ odd vanish. Hence if $a_{ij} \neq 0$, either both i and j are even, say $i = 2\nu$ and $j = 2\mu$, and $u^i v^j = (u^2)^\nu (v^2)^\mu$, or both are odd, in which case $u^i v^j = (u^2)^\nu (v^2)^\mu uv$ with $i = 2\nu + 1$ and $j = 2\mu + 1$. ☆

Exercises

⁂ (12.8) *Invariants under finite groups.* This exercise is a continuation of Proposition 12.25 above. Let the finite group G act on the domain B and let A denote the ring of invariants; that is, $A = B^G = \{ x \in B \mid g(x) = g \text{ for all } g \in G \}$. Let K be the fraction field of B and L that of A.

a) Show that action of G on B extends in a unique way to an action on K.

b) Assume x and y are two elements from B with $y \neq 0$ and such that $g(xy^{-1}) = xy^{-1}$ for all $g \in G$. Show that $y \prod_{g \neq e} g(x)$ is invariant under G. Conclude that L is the field of invariants in K; that is, $L = K^G = \{ xy^{-1} \mid g(xy^{-1}) = xy^{-1} \text{ for all } g \in G \}$. HINT: $x/y = \left(x \prod_{g \neq e} g(x)\right) \left(y \prod_{g \neq e} g(x)\right)^{-1}$.

c) Show that B is integral over A. HINT: Show that any symmetric polynomial in the $g(x)$'s is invariant and use that $\prod_{g \in G} (x - g(x)) = 0$.

d) Show that A is normal whenever B is.

⁂ (12.9) *New:Transitive action on fibres.* Keeping the notation from the previous exercise, show that G acts transitively on the fibre over a prime ideal $\mathfrak{p} \in \operatorname{Spec} A$; that is, it acts

transitively on the set $\{\,\mathfrak{q} \in \operatorname{Spec} B \mid \mathfrak{q} \cap A = \mathfrak{p}\,\}$.

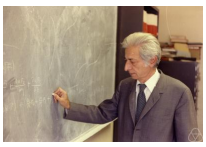
(12.10) *The cone over rational normal curves.* Let $u^i v^{d-i}$ with $0 \leqslant i \leqslant d$ be a basis for the monomials of degree d in $\mathbb{C}[u,v]$. Show that the ring $\mathbb{C}[u^i v^{d-i} | 0 \leqslant i \leqslant d]$ is normal. HINT: Let μ_d be the group of d-th roots of unity and let $\zeta \in \mu_d$ act on $\mathbb{C}[u,v]$ by $u \mapsto \zeta u$ and $v \mapsto \zeta v$.

★

12.3 The Cohen–Krull–Seidenberg Theorems

There is cluster of results proven in full generality by I. S. Cohen and A. Seidenberg and dubbed "Going–Up", "Going–Down" and "Lying–Over" by them. These results were all first found by Krull, whose proofs, however, were valid only for integral domains.

The results relate prime ideals in one ring to prime ideals in another ring which is integral over the first. So let $A \subseteq B$ be the two rings. Every prime ideal $\mathfrak{q} \subseteq B$ intersects A in a prime ideal $\mathfrak{p} = A \cap \mathfrak{q}$ (one says $\mathfrak{q}$ *lies over* $\mathfrak{p}$ or *contracts* to $\mathfrak{p}$) and as we know, this sets up the corresponding map $\pi\colon \operatorname{Spec} B \to \operatorname{Spec} A$ between the spectra. The Cohen–Seidenberg theorems are basically results about this map, approaching questions like when is it surjective, and what are the fibres? What about chains of prime ideals, can they extended? Is it a close map? Or an open map?



Abraham Seidenberg (1916–1988) American mathematician

(12.26) Notice that B is just assumed to be integral over A and is not necessarily a finitely generated A-module. So for example, the highly infinite extension $\mathbb{Z} \subseteq \overline{\mathbb{Z}}$ where $\overline{\mathbb{Z}}$ denotes the integral closure of $\mathbb{Z}$ in the field of algebraic integers $\overline{\mathbb{Q}}$ will satisfies the hypothesis.

A basic lemma—the case of fields

(12.27) This pivotal lemma treats the special case of fields:

LEMMA 12.28 *Let $A \subseteq B$ be an integral extension of domains. If one of the rings is a field the other one is a field as well.*

PROOF: Assume first that B is a field. If $y \in A$ is a non-zero element, the inverse y^{-1} is integral over A and satisfies a dependence relation

$$y^{-n} + a_{n-1}y^{-(n-1)} + \ldots + a_1 y^{-1} + a_0 = 0,$$

with the a_i's being elements from A. Multiplying through by y^n gives

$$1 + y(a_{n-1} + \ldots + a_1 y^{n-2} + a_0 y^{n-2}) = 0$$

which shows that y is invertible in A. Next assume that A is a field, and let $x \in B$ be a given non-zero element. It satisfies a relation

$$x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = 0,$$

with coefficients a_i from A, and assuming that the degree n is minimal, it holds that $a_0 \neq 0$. Then a_0 will be invertible, and we have

$$x \cdot a_0^{-1}(x^{n-1} + a_1 x^{n-2} + \ldots + a_1) + 1 = 0.$$

❑

COROLLARY 12.29 *Assume that $A \subseteq B$ is an integral extension. A prime ideal $\mathfrak{n}$ in B is maximal if and only if $\mathfrak{n} \cap A$ is maximal.*

PROOF: The extension

$$A/\mathfrak{n} \cap A \subseteq B/\mathfrak{n}$$

is integral by Proposition 12.17 on page 322, and the corollary ensues since the quotient by an ideal is a field if and only if the ideal is maximal. ❑

The Lying–Over Theorem

The Lying-over Theorem (Lying–over-teoremet)

The first theorem of the cluster—*the Lying-Over Theorem*—describes the structure of the fibres of the map π; that is, a qualitative description of the set of prime ideals in B intersecting A in a given fixed prime ideal. The fibres are non-empty; in other words, all prime ideals $\mathfrak{p}$ in A are of the form $\mathfrak{p} = \mathfrak{q} \cap A$, and there are no inclusion relations between members of a fibre.

(12.30) Here it comes:

PROPOSITION 12.31 (LYING–OVER) *Assume that $A \subseteq B$ is an integral extension. For each prime ideal $\mathfrak{p} \subseteq A$ there is at least one prime ideal $\mathfrak{q} \subseteq B$ such that $\mathfrak{q} \cap A = \mathfrak{p}$. Moreover, if $\mathfrak{q}$ and $\mathfrak{q}'$ are prime ideals in B with $\mathfrak{q} \cap A = \mathfrak{q}' \cap A$ and $q \subseteq \mathfrak{q}'$, then $\mathfrak{q} = \mathfrak{q}'$.*

In geometric terms the Lying–Over theorem asserts that the induced map $\pi\colon \operatorname{Spec} A \to \operatorname{Spec} B$ between the spectra is surjective and has discrete fibres, additionally it will also be a closed map; *i. e.* images of closed sets are closed, as Proposition 12.11 below shows.

PROOF: We begin with treating the local case, and subsequently we'll reduce the general situation to that case by localizing.

Assume then that A is local with maximal ideal $\mathfrak{m}$. Let $\mathfrak{n}$ be any maximal ideal in B; there are such according to the The Fundamental Existence Theorem for Ideals (Theorem 2.49 on page 49). By Corollary 12.29 above the contraction $\mathfrak{n} \cap A$ is maximal, hence equal to $\mathfrak{m}$ since $\mathfrak{m}$ is the only maximal ideal in A.

To see that no inclusion relations holds among ideals in a fibre of π, assume that $\mathfrak{q} \subseteq \mathfrak{q}'$ are two primes in B both intersecting A in $\mathfrak{m}$ (we are still in the local situation). Again by Corollary 12.29 both $\mathfrak{q}$ and $\mathfrak{q}'$ are maximal and must consequently be equal as one is contained in the other.

Let then $\mathfrak{p}$ be a prime ideal in A. To reduce to a local situation we pass to the localized extension $A_\mathfrak{p} \subseteq B_\mathfrak{p}$, which persists being integral in view of Proposition 12.15. Lying–Over being true in the local case, implies there is a prime ideal in $B_\mathfrak{p}$, which as all prime ideals in $B_\mathfrak{p}$ is of the form $qB_\mathfrak{p}$ with $\mathfrak{q}$ a prime ideal in B, such that $\mathfrak{q}B_\mathfrak{p} \cap A_\mathfrak{p} = \mathfrak{p}A_\mathfrak{p}$. But then the equality $\mathfrak{q} \cap A = \mathfrak{p}$ follows *e.g.* because

$$\mathfrak{q} \cap A = \mathfrak{q}B \cap A = (A_\mathfrak{p} \cap \mathfrak{q}B_\mathfrak{q}) \cap A = \mathfrak{p}A_\mathfrak{p} \cap A = \mathfrak{p}.$$

If $q \cap A = q' \cap A = \mathfrak{p}$ for two prime ideals in B, one included in the other, it holds by the local case that $\mathfrak{q}B_\mathfrak{p} = \mathfrak{q}'B_\mathfrak{p}$, and hence $\mathfrak{q}$ and $\mathfrak{q}'$ are equal. ❑

PROPOSITION 12.32 *If $A \subseteq B$ is an integral extension, the map $\pi\colon \operatorname{Spec} B \to \operatorname{Spec} A$ induced between the spectra will be a closed map*

PROOF: Let $\mathfrak{a} \subseteq B$ be an ideal. We shall show that $\pi(V(\mathfrak{a})) = V(\mathfrak{a} \cap A)$. To that end, we apply Lying–Over to the inclusion $A/\mathfrak{a} \cap A \subseteq B/\mathfrak{a}$ and conclude that every $\mathfrak{p} \in V(\mathfrak{a} \cap A)$ is of the form $\mathfrak{q} \cap A$ with $\mathfrak{q} \in V(\mathfrak{a})$. The other inclusion being trivial, we are through. ❑

EXAMPLE 12.8 It might well happen that π is surjective and has finite fibres without B being integral over A. A cheap example can be the extension

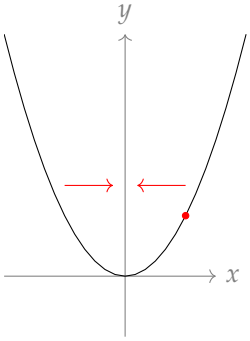
$$A = k[x^2] \subseteq k[x, (x-1)^{-1}] = B$$

where we assume that k is algebraically closed and not of characteristic two. The geometric interpretation is the parabola C given as $y = x^2$ with a hole punched in it: the point $(1,1)$ is removed. The map π is just projection $C\backslash\{(1,1)\}$ onto the y-axis.

The ring $k[x^2]$ is isomorphic to the polynomial ring $k[y]$ (rebaptize x^2 to y). Every maximal ideal $\mathfrak{m}$ in A is of therefore the form $x^2 - a^2$ (all elements in k have a square-root), and it holds true that $(x-a)B \cap A = (x^2-a^2)A$ since $x^2 - a^2 = (x+a)(x-a)$ in B. However, $(x-1)^{-1}$ is not integral over A; indeed, if it were, multiplying an integral dependence relation of degree n by $(x-1)^n$, would have given a relation

$$1 + p_{n-1}(x-1) + \ldots + p_1(x-1)^{n-1} + p_0(x-1)^n = 0,$$

where the coefficients p_i's are elements on $k[x^2]$. Putting $x = 1$ gives an obvious contradiction. As of the fibres of $\mathfrak{p}$, a nice exercise would be to check that if $a \neq 1$, the two prime ideals $(x-a)B$ and $(x+a)B$ are the ones lying over $(x^2-a^2)A$, but the sole prime ideal lying over $(x-1)A$ is $(x+1)B$. Figuring out what the fibres are when k is of characteristic two would as well be instructive. ☆



The Going–Up Theorem (Going–Up teoremet)

Going–Up

The *Going–Up Theorem* is about extending, or lifting as one also says, ascending chains of prime deals in A to chains in B by climbing them—the lifted chain ascends from a

given extension of the smallest prime ideal in the chain in A—which is contrary to the Going–Down Theorem where chains are lifted by a downwards "movement". If every one-step chain in A may be lifted, an easy induction ensures that every finite ascending chain may be lifted.

$$\begin{array}{ccccc} \mathfrak{q}_0 & \subseteq & \mathfrak{q}_1 & \subseteq & B \\ \cup| & & \cup| & & \cup| \\ \mathfrak{p}_0 & \subseteq & \mathfrak{p}_1 & \subseteq & A \end{array}$$

(12.33) The one step case is what is usually called the Going–Up Theorem:

THEOREM 12.34 (GOING–UP) *Let $A \subseteq B$ is an integral extension of rings, and let $\mathfrak{p}_0 \subseteq \mathfrak{p}_1$ be two prime ideals in A. Furthermore assume that $\mathfrak{q}_0$ is a prime ideal in B lying over $\mathfrak{p}_0$. Then there is a prime ideal $\mathfrak{q}_1$ in B containing $\mathfrak{q}_0$ and lying over $\mathfrak{p}_1$.*

PROOF: Consider the extension $A/\mathfrak{p}_0 \subseteq B/\mathfrak{q}_0$ which is integral by Proposition 12.17. By the Lying-Over Theorem, there is a prime ideal in $B/\mathfrak{q}_0$ lying over $\mathfrak{p}_1/\mathfrak{p}_0$. As all prime ideals in $B/\mathfrak{q}_0$ are, it is shaped like $\mathfrak{q}_1/\mathfrak{q}_0$ for some $\mathfrak{q}_1$ in B. Then $\mathfrak{q}_1 \cap A = \mathfrak{p}_1$. ❑

COROLLARY 12.35 (GOING–UP II) *Assume that $A \subseteq B$ is an integral extension and that $\mathfrak{q}_0 \subseteq B$ is a prime ideal. Let $\mathfrak{p}_0 = \mathfrak{q}_0 \cap A$. Any saturated chain $\{\mathfrak{p}_i\}$ of prime ideals in A ascending from $\mathfrak{p}_0$ lifts to a saturated chain $\{\mathfrak{q}_i\}$of prime ideals in B ascending from $\mathfrak{q}_0$.*

PROOF: The proof goes by induction on the number of prime ideals in the chain in A, and one should find the proof completely transparent pondering the following display:

$$\begin{array}{ccccccccc} \mathfrak{q}_0 & \subset & \mathfrak{q}_1 & \subset & \dots & \subset & \mathfrak{q}_{n-1} & & \\ \cup| & & \cup| & & & & \cup| & & \\ \mathfrak{p}_0 & \subset & \mathfrak{p}_1 & \subset & \dots & \subset & \mathfrak{p}_{n-1} & \subset & \mathfrak{p}_n. \end{array}$$

The upper chain exists by induction, an one just fils in the upper right corner citing the Going–Up Theorem.

A chain $\{\mathfrak{q}_i\}$ in B that lifts the chain $\{\mathfrak{p}_i\}$, will be saturated whenever $\{\mathfrak{p}_i\}$ is; indeed, any prime strictly in between $\mathfrak{q}_i$ and $\mathfrak{q}_{i+1}$ would either meet A in $\mathfrak{p}_i$ or $\mathfrak{p}_{i+1}$ since $\{\mathfrak{p}_i\}$ is saturated, but this can not happen since Lying–Over guarantees there are no inclusions among primes in same the fibre. ❑

Exercises

⁂ (12.11) *New.* Assume that $A \subseteq B$ is an integral ring extension. Show that $J(B) \cap A = J(A)$.

⁂ (12.12) Let $A \subseteq B$ be an extension of rings which is *purely inseparable*; that is, A is of characteristic p and for each element $x \in B$ there is an exponent p^ν so that $x^{p^\nu} \in A$. Show that the induced map $\pi\colon \operatorname{Spec} B \to \operatorname{Spec} A$ is a homeomorphism. ★

Going–Down

The Going–Up Theorem asserts that the larger of two prime-ideals has an extension when the smaller one has one as well, in the Going–Down Theorem the order is reversed,

if the larger can be lifted the smaller can be lifted and in such a way that lifted ideal is contained in the lifting of the larger. Contrary to Going–Up Going–Down does not hold for general integral extensions, the smaller ring A must be normal, and the larger B must be a faithful A-module; most frequently one finds formulations with the weaker requirement that B be a domain. We shall not go into the proof of the general Going–Down, but confine ourself to state the theorem and give a proof in the special case that the extension is finite and separable (the inseparable case is treated in an exercise). However, there is a simple and nice example explaining why the smaller ring must be normal, which we can not resist giving.

(12.36) As indicated we content ourself to formulate the general Going–Down theorem, here is the core statement:

$$\begin{array}{ccccc} \mathfrak{q}_1 & \subseteq & \mathfrak{q}_0 & \subseteq & B \\ \cup\!| & & \cup\!| & & \cup\!| \\ \mathfrak{p}_1 & \subseteq & \mathfrak{p}_0 & \subseteq & A \end{array}$$

THEOREM 12.37 (GOING–DOWN) *Let $A \subseteq B$ be an integral extension of integral domains and assume that A is normal. Given prime ideals $\mathfrak{p}_1 \subseteq \mathfrak{p}_0$ in A and $\mathfrak{q}_0$ in B lying over $\mathfrak{p}_0$. Then there exists a prime ideal $\mathfrak{q}_1$ in B contained in $\mathfrak{q}_0$ and lying over $\mathfrak{p}_1$.*

By a repeated application of the theorem one readily shows that every descending chain of prime ideals in A can be lifted to one in B which extends downwards from a given lifting of the top member of the chain in A.

COROLLARY 12.38 *Let $A \subseteq B$ be an integral extension of integral domains and assume that A is normal. Let $\mathfrak{p}_n \subseteq \ldots \subseteq \mathfrak{p}_0$ be a chain of prime ideals in A and let $\mathfrak{q}_0$ be a prime ideal in B lying over $\mathfrak{p}_0$. Then there exists a chain $\mathfrak{q}_n \subseteq \ldots \subseteq \mathfrak{q}_0$ in B such that $\mathfrak{q}_i \cap A = \mathfrak{p}_i$.*

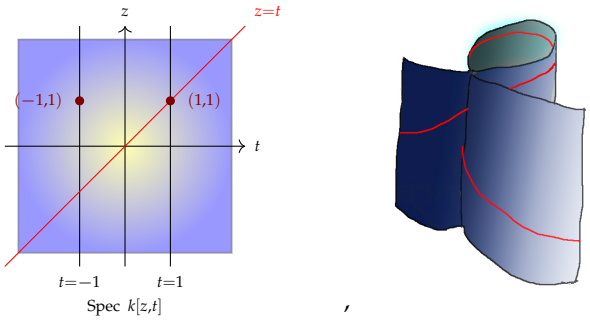
(12.39) As promised, we shall provide a simple proof when B is the integral closure of A in a finite and separable extension of the fraction field of A. Notably, in characteristic zero any finite extension is separable so many cases met in practical work is covered by this version (if you wonder about the inseparable case, do Exercise 12.13 below). The proof relies on Exercises 12.8 and 12.9 about group actions (for which solutions are provided).

PROOF OF GOING-DOWN IN THE FINITE AND SEPARABLE CASE: We assume first that we are in the "Galois-situation" where there is a finite group acting on B such that $A = B^G$. By the Lying–Over Theorem there is a prime $\mathfrak{q}_1'$ lying over $\mathfrak{p}_1$ (which not necessarily is contained in $\mathfrak{q}_0$). However, by Going–Up it is contained in a prime ideal $\mathfrak{q}_0'$ that lies over $\mathfrak{p}_0$. Now, by Exercise 12.9 the group G acts transitively on the fibres, so there is a $g \in G$ such that $\mathfrak{q}_0 = g(\mathfrak{q}_0')$. Then $g(\mathfrak{q}_1')$ is our man.

In the general situation when B is the integral closure of A in a finite and separable extension L of the fraction field K of A, there is an extension E of L which is Galois over K, say with Galois group G. Then if C denotes the integral closure of A in E, it holds

that $A = C^G$. Thus by the beginning of the proof Going–Down holds for the extension $A \subseteq C$, an *a fortriori* for $A \subseteq B$ since $A \subseteq B \subseteq C$. ❑

EXERCISE 12.13 *The unseparable case.* Extend the above proof to the case where $A \subseteq B$ is merely assumed to be finite; that is, it is not necessarily separable (but still A and B are domains, and A is still integrally closed in its fraction field). HINT: With the notation of the proof: let E be a finite extension of L normal over K that contains E. If G is the Galois group of L over K, then C^G is a purely inseparable extension of A. Then use Exercise 12.12 on page 330. ★



EXAMPLE 12.9 The example is built on one of the simplest non-normal rings, namely the coordinate ring of the so-called *ordinary double point*; the plane curve C with equation $y^2 = x^2(x+1)$. The curve C was already studied in Example 12.2 on page 317 where it was parameterized by the assignments $x = (t^2-1)$ and $y = t(t^2-1)$. To fix the ideas, we shall work over the complex numbers although things go through over any field whose characteristic is not two.

The variety we have in mind is the cylinder D over C with the z-axis as generator; so it is given by the same equation $y^2 = x^2(x+1)$, but in the three-dimensional space $\mathbb{C}^3$. The coordinate ring A of D equals $\mathbb{C}[x,y,z]$ with constituting relation $y^2 - x^2(x+1) = 0$. This ring also equals the subring $\mathbb{C}[(t^2-1), t(t^2-1), z]$ of the polynomial ring $\mathbb{C}[t,z]$; the parameterization tells us that.

A heuristic description of the geometry is as follows: The map $\psi\colon \mathbb{C}^2 \to \mathbb{C}^3$ that sends (t,z) to $(t^2-1, t(t^2-1), z)$ is bijective onto D except that both $(1,z)$ and $(-1,z)$ are mapped to $(0,0,z)$; so the two lines in $\mathbb{C}^2$ defined respectively by $t=1$ and $t=-1$, are both sent to the z-axis.

Consider the line L in the parameter plane $\mathbb{C}^2$ whose equation is $z = t$. It passes by $P_1 = (1,1)$, but not by $P_2 = (-1,1)$. The crucial point is that an irreducible curve in $\mathbb{C}^2$ with the same image as L must coincide with L where ψ is injective; that is, off the two lines $t = \pm 1$, and hence is must equal L. Now $Q = (0,0,1) \in \psi(L)$, and P_2 maps to Q,

but no irreducible curve mapping to $\psi(L)$ passes by P_2. It is straightforward to check that the image $\psi(L)$ is given in $\mathbb{C}^3$ by the two equations $z^2 = x+1$ and $y^2 = xz$.

The algebraic (and precise) picture is as follows: There are two maximal ideals in $\mathbb{C}[t,z]$ lying over $\mathfrak{m} = (x,y,z-1)$ (which is the ideal of Q), namely $(t-1,z-1)$ and $(t+1,z-1)$ (which are the ideals of P_1 and P_2 respectively); indeed, one easily finds

$$\mathfrak{m}\cdot\mathbb{C}[t,z] = (t^2-1, t(t^2-1), z-1) = ((t-1)(t+1), z-1) = (t-1,z-1)\cap(t+1,z-1).$$

The ideal $\mathfrak{p} = (z^2-(x+1), y-zx)$ (which is the ideal in A of $\psi(L)$) is prime , since the map $\mathbb{C}[x,y,z] \to \mathbb{C}[x,z]$ that sends y to xz and lets x and z be untouched, transforms it to the prime ideal (z^2-x-1). One has $\mathfrak{p}\subseteq\mathfrak{m}$. We contend that $(z-t)\cap A = \mathfrak{p}$ and that $z-t$ is the only prime ideal in $\mathbb{C}[t,z]$ extending $\mathfrak{p}$. Indeed, in $\mathbb{C}[t,z]$ one finds the primary decomposition

$$\mathfrak{p}\cdot\mathbb{C}[t,z] = (z^2-t^2, (t^2-1)(z-t)) = (z-t)\cap(z-1,t+1)\cap(z+1,t-1),$$

from which the claim follows. (The curve $\psi(L)$ hits the z-axis in the two points $(0,0,1)$ and $(0,0,-1)$, which explains the occurrence of the component $(z+1,t-1)$.)

Now the point is that the extension $(t+1,z-1)$ of $\mathfrak{m}$ does not contain the only extension $(z-t)$ of $\mathfrak{p}$.

☆

Consequences for dimension and height

A most useful consequence of the Going–Up Theorem is that any chain in an integral extension B of a ring A, when intersected with A becomes a chain* in A; this ensures that the dimension is preserved in integral extensions: We have

*Remember that inclusions in chains are supposed to be strict.

PROPOSITION 12.40 (GOING–UP III) *If $A\subseteq B$ is an integral extension of rings, then* $\dim A = \dim B$.

PROOF: This is a direct consequence of the Going-Up theorems. As formulated in Going–Up II (Corollary 12.35 on page 330) every chain in A has a chain in B lying over it, which means that $\dim A \leqslant \dim B$. On the other hand, Lying–Over (Proposition 12.31 on page 328) entails that each saturated chain in B remains saturated when intersected with A. Hence $\dim B \leqslant \dim A$. ❑

(12.41) For example, if $\mathbb{Q}\subseteq K$ is any field extension (finite or not) the ring of integers in K; that is, the integral closure A of $\mathbb{Z}$ in K, is of course integral over $\mathbb{Z}$ and consequently is of dimension one. In particular, this applies to the quadratic extensions $K = \mathbb{Q}(\sqrt{d})$ we have seen, but also to the more impressive extension $K = \overline{\mathbb{Q}}$, the field of algebraic numbers. The ring of algebraic integers $\overline{\mathbb{Z}}$ is therefore of dimension one, but recall, it is not a Noetherian ring.

(12.42) An easy corollary of the Going–Down Theorem may be expressed by the slogan: "height is preserved in integral extensions where Going–Down holds". The precise statement is as follows:

COROLLARY 12.43 *Assume that $A \subseteq B$ is an integral extension of integral domains with A normal. If $\mathfrak{q}$ is a prime in B and $\mathfrak{p} = \mathfrak{q} \cap A$, then $\operatorname{ht}_A \mathfrak{p} = \operatorname{ht}_B \mathfrak{q}$.*

PROOF: Each strictly decreasing chain $\{\mathfrak{q}_i\}$ descending from $\mathfrak{q}$ intersects A in a chain descending from $\mathfrak{p}$ which is strictly decreasing by Lying-Over. Hence $\operatorname{ht}\mathfrak{p} \geqslant \operatorname{ht}\mathfrak{q}$. Each strictly decreasing chain of prime ideals $\{\mathfrak{p}_i\}$ descending from $\mathfrak{p}$ lifts according to Going-Down to a strictly decreasing chain of prime ideals descending from $\mathfrak{q}$, so the inequality $\operatorname{ht}\mathfrak{p} \leqslant \operatorname{ht}\mathfrak{q}$ holds as well. ❑

12.4 *Then finiteness issue*

Domains that are integrally closed in their fraction fields, have several very good properties which makes them easier to work with, and a very natural technique is to relate a domain, or a variety, to its normalization.

It suffices to mention the particular case of one-dimensional domains finitely generated over field; they are precisely the coordinate rings of the affine non-singular curves, so normalizing a curve is a way to desingularize it; that is, to exhibit a non-singular curve with the same function field—a so-called *non-singular model* of the curve. In the general case, for varieties of higher dimension, the normalizations are not necessarily non-singular anymore, but still they have lots of good properties.

One dimensional normal domains finite over the integers $\mathbb{Z}$, are the classical Dedekind rings for which the extended fundamental theorem of arithmetic is valid (and which we shall come back to). One desires that any "number domain" can be embedded in a Dedekind ring A, with the two differing only at finite many places, and moreover, finite extensions of the number field gives rise, passing to the integral closure, to extensions of the rings of integers.

It is thus desirable that the domains one works with have an integral closure in their fraction field, or in a finite extension of the fraction field, that is a finite module. But, alas, this is not even true for Noetherian domains, (the first examples were found by Nagata). For domains finitely generated over fields, however, it holds true, and this will be sufficient for many geometric applications.

We intend to give a very short account of this story—without complete proofs—but hopefully indicating enough to give the reader a feeling of the inherent mechanisms of the stuff. The *trace* will be a valuable tool, and we have included a short (and incomplete) description of its most important properties in an appendix

The first finiteness theorem

(12.44) We attack the first finiteness theorem:

THEOREM 12.45 *Let A be a normal Noetherian domain with fraction field K and let L be a finite separable field extension of K. Then the integral closure $\bar{A}$ of A in L is a finite A-module.*

PROOF: Let $x_1, \ldots, x_n$ be elements in $\bar{A}$ which form a K-basis for L. After Proposition 12.50 above we know that the trace gives a non-degenerate quadratic form $\mathrm{tr}_{L/K}\, xy$ on the field L, and the basis $x_1, \ldots, x_n$ has a dual basis $y_1, \ldots, y_n$ with respect to this form; that is, the y_i's satisfy the relations $\mathrm{tr}\, x_i y_j = \delta_{ij}$. We contend that

$$\bar{A} \subseteq Ay_1 + \ldots + Ay_n, \tag{12.9}$$

and consequently it will hold that $\bar{A}$ is finite over A: indeed, the right hand sum in (12.9) is Noetherian being finitely generated over the Noetherian ring A, so any submodule will be finitely generated over A.

To establish the inclusion in (12.9) we express any member a of $\bar{A}$ in the dual basis as $a = a_1 y_1 + \ldots + a_n y_n$. Multiplying by x_i and taking traces we find that $a_i = \mathrm{tr}_{L/K}\, x_i a$. But $x_i a$ belongs to $\bar{A}$, and as the trace of an integral element is integral, $\mathrm{tr}_{L/K}\, x_i a$ belongs to A as A is normal. ❑

The second finiteness theorem

(12.46) Combined with Emmy Noether's normalization lemma, the First Finiteness Theorem gives as a corollary the result—important for algebraic geometers—that normalizations of domains finitely generated over a field are finite modules over the domain. When the ground field is of positive characteristic there is a separability issue which we'll unscrupulously sweep under the carpet. This shouldn't bother you if you primary are interested in algebraic geometry over fields of characteristic zero; *e.g.* over the complex numbers, but if unresolved issues in positive characteristic disturb your sleep, you should complete exercise xxx (not yet written, sorry).

THEOREM 12.47 *Let A be a domain which is a finitely generated k-algebra and let K be its field of fractions. Then the normalization $\widetilde{A}$, that is integral closure of A in K, is a finite A-module.*

PROOF: According to the normalization lemma we may find algebraically independent elements $w_1, \ldots, w_n$ in A so that A is a finite module over the polynomial ring $B = k[w_1, \ldots, w_n]$. Thence the fraction field K is finite over $L = k(w_1, \ldots, w_n)$, being a polynomial ring B is integrally closed in L, and $\widetilde{A}$ equals the integral closure of B in K. In view of Theorem 12.45 above, this finishes the proof in the case that k is of characteristic zero, since then the extension K of L is automatically separable. And as warned, the inseparable case is done by under-the-carpet-sweeping. ❑

Exercises

(12.14) Let A be a normal domain containing the rationals $\mathbb{Q}$. Assume that $A \subseteq B$ is an extension of domains and that B is finite over A. Show that A is a direct summand in B. HINT: Use the trace $\mathrm{tr}_{L/K}$ with L and K the fraction fields of respectively A and B .

(12.15) Let $A = k[x_0, x_1, x_2, \ldots] = k[x_i | i \in \mathbb{N}]$ where k is is a field whose characteristic is not 2, and let $B = k[x_i x_j | i, j \in \mathbb{N}]$. Let further K and L be the fraction fields respectively of A and B.

a) Show that A is not a finite B-module.
b) Show that L is finite and separable over K, and that $[L : K] = 2$. In fact, the extension is Galois with group $\mathbb{Z}/2\mathbb{Z}$.
c) Show that A is normal, and conclude that A is the integral closure of B in K.

★

12.5 *Appendix: Trace and separability*

The trace of an endomorphism (sporet til en endomorfi)

(12.48) Recall that any endomorphism θ of a finite dimensional vector space over a field k has a *trace*, denoted $\mathrm{tr}\,\theta$. It is defined as the the negative of the subleading coefficient of the characteristic polynomial of θ; that is, by the formula

$$\det(t \cdot \mathrm{id} - \theta) = t^n - (\mathrm{tr}\,\theta) \cdot t^{n-1} + \ldots + (-1)^n \det\theta.$$

The characteristic polynomial splits into linear factors in the algebraic closure of k—the roots are the *eigenvalues* or *characteristic roots* of θ—and the trace equals the sum of these roots. By developing the determinant one sees that the trace equals the sum of the diagonal elements in any matrix representing θ. This shows that the dependence on θ is linear; that is, $\mathrm{tr}(a\theta + a'\theta') = a\,\mathrm{tr}\,\theta + a'\,\mathrm{tr}\,\theta'$.

(12.49) In the staging of this section, where $K \subseteq L$ is a finite field extension, each element $x \in L$ has a trace $\mathrm{tr}_{L/K} x$, which equals the trace of the endomorphism $[x] \colon L \to L$ that sends y to xy.

The trace form (sporformen)

The trace gives rise to a K-bilinear form on L, namely the form $\mathrm{tr}_{L/K} xy$, which is called the *trace form*. It is obviously symmetric and K-linear since the trace is. In the present context the following property of the trace form is all important, but the proof relies on properties which are relegated to an appendix.

PROPOSITION 12.50 (TRACE AND SEPARABILITY) *A finite extension L of the field K is separable if and only if the trace form is non-degenerate.*

PROOF: The main observation is that the trace form is either identically equal to zero or non-degenerate since if $\mathrm{tr}_{L/K} x_0 \neq 0$, it follows that $\mathrm{tr}_{L/K}\, y(y^{-1}x_0) \neq 0$ for any non-zero $y \in L$.

To begin with suppose that the extension $K \subseteq L$ is inseparable so there is an element $x \in L$ not in K with $x^p = a \in K$. One of the properties of the trace not yet established, but which we anyhow shall use, is the functoriallity in towers: it holds that $\mathrm{tr}_{L/K} = \mathrm{tr}_{K(x)/K} \circ \mathrm{tr}_{L/K(x)}$. It thus suffices to see that $\mathrm{tr}_{K(x)/K}$ vanishes identically. The minimal irreducible polynomial of each power x^i with $1 \leqslant i < p$ is given as

$$T^{ip} - a^i,$$

hence $\mathrm{tr}_{K(x)/K}\, x^i = 0$ when $0 < i$, and $\mathrm{tr}_{K(x)/K}\, 1 = [K(x) : K] = p = 0$ as the characteristic K equals p.

Let us prove that a separable extension has a non-degenerated trace form (which is the implication we shall need); this follows from the primitive element theorem, which says that $L = K(x)$ for some $x \in L$: The minimal polynomial of x is of degree $n = [L : K]$, and it is separable* (by the definition of the extension being separable). The Cayley–Hamilton theorem tells us that x is a root of the characteristic polynomial P_x, hence P_x has the minimal polynomial as a factor, and both being of degree n, they are equal. Consequently P_x is separable, and Lemma 12.51 below produces an element of non-zero trace. ❑

*Recall that a polynomial is separable *if all the roots in an algebraic closure of the ground field are simple roots.*

LEMMA 12.51 *Let θ be an endomorphism of an n-dimensional vector space over the field k whose characteristic polynomial is separable. Then $\mathrm{tr}\,\theta^r \neq 0$ for some r with $0 \leqslant r < n$.*

PROOF: This gives us the opportunity to retrieve our good old acquaintance the famous Van der Monde determinant down from the loft. It is the determinant of the $n \times n$-matrix

$$D = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{pmatrix},$$

and it has the virtue of being non-zero when the λ_i's are distinct, as they are in our case. Recall that $\mathrm{tr}\,\theta^r$ is the power sum $\mathrm{tr}\,\theta^r = \sum_i \lambda_i^r$ of the eigenvalues λ_i, so that

$$(\mathrm{tr}\,\theta^0, \mathrm{tr}\,\theta, \mathrm{tr}\,\theta^2, \dots, \mathrm{tr}\,\theta^{n-1}) = D \cdot (1, \dots, 1),$$

and since D is invertible, the left hand vector is non-zero, which means that for at least one exponent r it holds that $\mathrm{tr}\,\theta^r \neq 0$. ❑

Appendix: transcendence degree

We have relegated a few simple results of preparatory character to this appendix. They do not take part in the main battle, but are merely skirmishes on the flanks, though of

significant importance for the progress. At least the two firsts are easy and elementary. The concept of transcendence degree might be a little more involved. It should be known to mosts students from earlier courses, but we include it for the benefit of the others.

Transcendence degree

(12.52) Let $k \subseteq K$ be a field extension and let $x \in K$ be an element not lying in k. The elements in K can be parted in to two classes, the algebraic elements and the transcendental ones. The algebraic ones are those x that satisfies a relation like

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = 0 \tag{12.10}$$

where the a_i's are elements from k and $a_n \neq 0$, and the transcendental ones are the rest; that is, those for which $p(x) \neq 0$ for any non-zero-polynomial in $k[X]$. Since K is field, we may as well assume that $a_n = 1$ in (12.10) and being algebraic over k is the same as being integral. The elements in K that are algebraic over k, thus form a subfield $\bar{k}$, the *algebraic closure* of k in K.

The algebraic closure (den algebraiske tillukningen)

Algebraically dependent elements (algebraisk avhengige elementer)

Algebraically independent elements (algebraisk uavhengige elementer)

transcendence basis (transcendence basis)

More generally a collection $x_1, \ldots, x_r$ of element from the bigger field K is said to be *algebraically dependent* over k if for some non-zero polynomial p in r variables with coefficients from k it holds true that $p(x_1, \ldots, x_r) = 0$, and of course, if no such polynomial can be found, the collection is said to be *algebraically independent* over k. A collection of algebraically independent elements $x_1, \ldots, x_n$ is a *transcendence basis* of K over k if it is maximal. In other words, the x_i's are algebraically independent and K is algebraic over $k(x_1, \ldots, x_n)$.

PROPOSITION 12.53 *Every finitely generated field extension K of k that has a finite transcendence basis.*

PROOF: Let $x_1, \ldots, x_n$ generate K over k. If K is not algebraic over k, at least one of the x_i's is transcendental, and we may well assume it is x_1. Then K is generated by $x_2, \ldots, x_n$ over $k(x_1)$ and by induction on n this field extension has a transcendence basis, which together with x_1 yields a transcendence basis for K over k. ❑

(12.54)

PROPOSITION 12.55 *If K is a finitely generated field extension of k, then all transcendent bases for K over k have the same number elements.*

Transcendence degree (transcendensgrad)

The common number is called the *transcendence degree* of K over k and denoted by $\mathrm{trdeg}_k(K)$. In case K is not finitely generated, a similar statement holds true; transcendence bases are then no longer necessarily finite, but they the will still be of the same cardinality.

LEMMA 12.56 (EXCHANGE LEMMA) *Let $k \subseteq L \subseteq K$ be a tower of fields and let a and b be two elements of K. Assume that b is transcendent over L, but algebraic over $L(a)$. Then a is algebraic over $L(b)$.*

PROOF: Since b is algebraic over $L(a)$, there is a polynomial $f(x,y)$ with coefficients from L such that $f(a,y) \neq 0$, but $f(a,b) = 0$. Expanding f in powers of x we obtain

$$0 = f(a,b) = \sum q_i(b)a^i,$$

where $q_i(y) \in L[y]$. This looks very much like a dependence relation for a over $L(b)$; it only remains to see that $f(x,b)$ is not identically zero, but since b is transcendent over L, and at least one of the polynomials $q_i(y)$'s is non-zero, this holds true. ❑

PROOF OF PROPOSITION 12.55: Let $a_1, \ldots, a_n$ be a transcendence basis for K over k of shortest length, and let $b_1, \ldots, b_m$ be another one.

If the two bases have a common element, induction on n will finish off the proof; indeed, if $a_1 = b_1$, replacing k by $k(a_1) = k(b_1)$, we may conclude that $b_2, \ldots, b_m$ and $a_2, \ldots, a_n$ have the same number of elements.

Now, the role of the Exchange Lemma is to permit us to exchange a_1 with one of the b_i's: In order to do that, introduce the auxiliary field $L = k(a_2, \ldots, a_n)$. Not all the b_i's cannot be algebraic over L, and we may assume that b_1 is transcendental over L. Since b_1 is algebraic over $L(a_1)$, by the Exchange Lemma a_1 will be algebraic over $L(b_1) = k(b_1, a_2, \ldots, a_n)$. As being an algebraic extension is a property transitive in towers, we may conclude that K, being algebraic over $L(a_1, b_1)$, is algebraic over $L(b_1) = k(b_1, a_2, \ldots, a_n)$, and so $b_1, a_2, \ldots, a_n$ is a transcendence basis for K over k. ❑

Lecture 13

Algebras of finite type over fields

The algebras of finite type over a field form the bedrock of the theory of varieties, and therefore the whole algebraic geometry depends on their properties. This chapter is devoted to the two most basic results about such algebras and some the central corollaries.

First comes Noether's Normalization lemma, a most valuable technically flavoured result linking algebras of finite type over fields to polynomial rings, and on which rests, in our approach, the proof of the next basic result, the Nullstellensatz.

Hilbert's Nullstellensatz is the bridge between algebra and geometry, to which algebraic geometry owes its very existence. We already met the Nullstellensatz in low dimensions. In dimension one it is just the definition of a field being algebraically closed (or, in the complex case one would rather say, the good old Fundamental Theorem of Algebra), and in dimension two it boils down to a corollary of the classical Gauss's lemma (Theorem 3.32 on page 78).

Finally, we close the chapter by giving some consequences of the two big results. Domains of finite type over a field has the agreeable property that all maximal chains are of the same length; a property we have termed to be of *uniform altitude*. Moreover their Krull dimension equals the transcendence degree of the their fraction field; in particular, a polynomial ring in n variables is of the highly expected dimension n. And at the very end of the chapter we skirts the question whether the tensor product of two domains over a field is a domain, gives a few examples and prove it holds true when the ground field is algebraically closed.

13.1 Noether's normalization lemma

Once more a lemma that has become a theorem and once more an important result due to Emmy Noether. The lemma, or should one say the theorem, states that a domain A of finite type over k can be realized as a finite algebra over a polynomial ring. In other

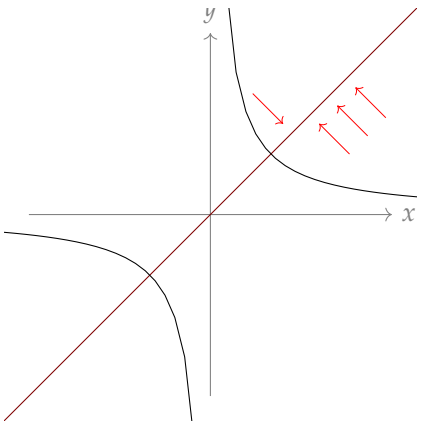
words, one may find elements $w_1, \dots, w_n$ in A which are algebraically independent over k—and hence $k[w_1, \dots, w_n]$ is isomorphic to a polynomials ring —such that A is a finite module over $k[w_1, \dots, w_n]$.

Since the w_i's are algebraically independent, and since A is a finite module over $k[w_1, \dots, w_n]$, they form a transcendence basis over k for the fraction field K of A. The number n will therefore be equal to the transcendence degree $\operatorname{trdeg}_k K$.

A heuristic sketch of the proof

(13.1) We find it worthwhile to give a preliminary sketch of the proof, so let A be an algebra of finite type over k. Starting out with the subfield $k \subseteq A$ and adjoining elements, we will sooner or later exhaust the entire ring A since A is finitely generated over k. In the beginning we may add new elements which are algebraically independent of those already added, but at a certain point, when the maximal number of algebraically independent elements is reached, new elements are forced to be algebraically dependent on the previous. If a new element is *integrally dependent* on the old ones, we are happy, if not, we have to go back and perturb the already added elements to make the new-comer integral; and the crux of the proof is to see that this perturbation is possible.

EXAMPLE 13.1 A non-integral new-comer will typically have a pole, and to illustrate the perturbation process, we consider the simplest way of adding a function with a pole, namely the extension $k[x, 1/x]$ of $k[x]$. The geometric counterpart is the projection of the classical hyperbola, $xy = 1$ onto the x-axis, the hyperbola just being the graph of the function $1/x$.



The ring $k[x, 1/x]$ is not finite over $k[x]$, but perturbing x slightly, we obtain a subring over which $k[x, 1/x]$ is finite. The subring $k[x + 1/x]$ will do the job; indeed, $k[x, 1/x] = k[x, x + 1/x]$ is generated by x as an algebra over $k[x + 1/x]$, and one has the integral dependence relation

$$x^2 - x(x + 1/x) + 1 = 0.$$

Geometrically, the inclusion of $k[x+1/x]$ into $k[x,1/x]$ corresponds to the projection sending (x,y) to $y+x$, which serves a coordinate on the line $y=x$.

It is worth noticing that almost *any* perturbation of x will work; that is, $k[x,1/x]$ is finite over $k[ax+b/x]$ as long as both the scalars a and b are non-zero; of course, the reason lies in the two asymptotes of the hyperbola. ☆

EXERCISE 13.1 Show that $k[x,1/x]$ is a finite module over $k[ax+b/x]$ for any scalars a b both being different from zero. ★

EXERCISE 13.2 *(Exam MAT4200 2020).* Let $B=k[x,y,z,w]$ with constituting relation $zw-xy=0$, and let A be the polynomial ring $A=k[x,y,z]$. Show that $A\subseteq B$ and that the extension is not integral. Describe perturbations x', y' and z' of x, y and z that make A integral over $k[x',y',z']$. ★

Statement and proof

(13.2) As indicated in the preliminary sketch, the proof of Noether's Normalization Lemma goes by induction on the number of generators A requires as an algebra over k, and the basic ingredient in the induction step is the following lemma:

LEMMA 13.3 *Let k be a field and let $A=k[x_1,\dots,x_m]$ be an algebra over k generated by algebraically dependent elements $x_1,\dots,x_m$. Then there are elements $y_1,\dots,y_{m-1}$ in A such that A is a finite module over $k[y_1,\dots,y_{m-1}]$.*

In the case that $m=1$ the conclusion should be understood as x_1 being algebraic over k; that is, it satisfies a polynomial equation $p(x_1)=0$ with coefficients from k.

PROOF: By assumption the m elements $x_1,\dots,x_m$ are algebraically dependent and thus satisfy an equation

$$p(x_1,\dots,x_m)=0,$$

where p is a non-zero polynomial in m variables with coefficients in k. The trick is to perturb the variables X_i's by putting $Y_i=X_{i+1}-X_1^{s^i}$ where s is a sufficiently large natural number. Then letting q be the polynomial given by

$$q(X_1,Y_1,\dots,Y_{m-1})=p(X_1,Y_1+X_1^s,Y_2+X_1^{s^2},\dots,Y_{m-1}+X_1^{s^{m-1}}) \tag{13.1}$$

we observe that

$$q(x_1,y_1,\dots,y_{m-1})=0, \tag{13.2}$$

where y_i is the element in A given as $x_{i+1}-x_1^{s^i}$. Moreover, and this is the important point, the coefficient of the highest power of X_1 occurring in q is a non-zero scalar; that is,

$$q(X_1,Y_1,\dots,Y_{m-1})=p(X_1,Y_1+X_1^s,Y_2+X_1^{s^2},\dots,Y_{m-1}+X_1^{s^{m-1}})=\alpha X_1^d+\text{lower terms},$$

with $a \in k$ and $a \neq 0$, and (13.2) is therefore an integral dependence relation for x_1 over $k[y_1, \ldots, y_{m-1}]$.

Indeed, the substitutions above transform a monomial $X_1^{\alpha_1} \cdot \ldots \cdot X_m^{\alpha_m}$ into a polynomial whose term of highest weight in X_1 is X_1^d where $d = \alpha_1 + \alpha_2 s + \ldots + \alpha_m s^{m-1}$, and whose leading coefficient is one. The crux is that all these combinations of α_i's and powers of s are different when s is large enough: only finitely many can arise from (non-zero) terms in q, and equating two can only be done in finitely many ways; hence there are only finitely many values of s for which two combinations coincide. ❏

(13.4) We are now well-prepared to attack the full version of the normalization lemma:

THEOREM 13.5 (NOETHER'S NORMALIZATION LEMMA) *Let k be a field and suppose given an algebra $A = k[x_1, \ldots, x_m]$ generated over k by elements $x_1, \ldots, x_m$. Then there are algebraically independent elements $w_1, \ldots, w_n$ in A so that A is a finite module over the polynomial ring $k[w_1, \ldots, w_n]$.*

PROOF: As announced, we proceed by induction on m. If the elements $x_1, \ldots, x_m$ are algebraically independent, we are through because then A itself will be a polynomial ring. If not, there is a non-zero polynomial p such that $p(x_1, \ldots, x_m) = 0$, and by Lemma 13.3 above, there are elements $y_1, \ldots, y_{m-1}$ such that A is finite over $B = k[y_1, \ldots, y_{m-1}]$. By induction there are algebraically independent elements $w_1, \ldots, w_n$ in B so that B is finite over $k[w_1, \ldots, w_n]$. But then A will be finite over $k[w_1, \ldots, w_n]$ as well, and we are through. ❏

COROLLARY 13.6 *Assume that A is a domain which is finitely generated over the field k, and whose fraction field is of transcendence degree n over k. Then there are n algebraically independent elements $w_1, \ldots, w_n$ in A such that A is a finite module over $k[w_1, \ldots, w_n]$.*

PROOF: By the Normalization Lemma we may find algebraically independent elements $w_1, \ldots, w_n$ in A such that A is finite over $k[w_1, \ldots, w_n]$. Then K will be a finite extension of the rational function field $k(w_1, \ldots, w_n)$, and hence $n = \operatorname{trdeg}_k K$. ❏

EXERCISE 13.3 The proof of Lemma 13.3 given above is a slightly simplified version of the proof Nagata gives in his book. In her original proof Emmy Noether used linear substitutions. They are of a simpler nature than those above, which involve high powers, but require that the ground field be infinite.

a) Show that setting $y_i = x_{i+1} - \alpha_i x_1$ with $\alpha_i \in k$ for each index i, transforms the relation $p(x_1, \ldots, x_m) = 0$ into an equality

$$p(x_1, y_1 + \alpha_1 x_1, \ldots, y_{m-1} + \alpha_{m-1} x_1) = p_d(1, \alpha_1, \ldots, \alpha_{m-1}) x_1^d + \text{lower terms},$$

where d is the degree of p and p_d the homogenous component of p of degree d. HINT: Taylor expansion.

b) Show that for any non-zero polynomial $f(t_1, \dots, t_m)$ with coefficients in an infinite field k, there are infinitely many choices of elements α_i's from k so that $f(\alpha_1, \dots, \alpha_m) \neq 0$. HINT: Induction on the number of variables.

c) Conclude that if $A = k[x_1, \dots, x_m]$ with $x_1, \dots, x_m$ algebraically dependent, there are elements $y_1, \dots, y_{m-1}$ in A such that A is finite module over the k-algebra $k[y_1, \dots, y_{m-1}]$.

★

13.2 *Hilbert's Nullstellensatz*

This key result in the fruitful synthesis of algebra and geometry that algebraic geometry is, comes in about half a dozen different formulations placing different weight on the geometric and the algebraic aspects. Some are called "weak" and are apparently weaker than those called "strong", but of course, in the end they will all be equivalent.

The version of the Nullstellensatz that best elucidates the bridging role between algebra and geometry, asserts when the ground field k is algebraically closed, that all maximal ideals in $k[x_1, \dots, x_n]$ are of the form $(x_1 - a_1, \dots, x_n - a_n)$ where $(a_1, \dots, a_n)$ is a point in k^n. It thus establishes a one-to-one-correspondence between points in k^n; that is, geometry, and maximal ideals in $k[x_1, \dots, x_n]$; that is, algebra.

And there are scores of different proofs. The one we offer relies on Noether's Normalization Lemma and, when the Normalization Lemma is available, is the shortest path to the Nullstellensatz. However, there are simpler proofs which only use elementary algebra, and one is offered as an exercise (Exercise 13.7 on page 349 below).

(13.7) After a short paragraph about affine varieties, we begin with we stating the Nullstellensatz in its original form—as formulated by Hilbert, but in the modern style of contemporary mathematics—which also goes under the name of the Strong Nullstellensatz. Subsequently we formulate and prove some of the weaker avatars, and at the end of the section, we present a proof—the so–called and now classical, Rabinowitsch trick—that the Strong Nullstellensatz follows from the weak ones.

Varieties

Before the wonderful world of schemes was discovered, the basic geometric objects in algebraic geometry were the varieties. A drawback varieties suffer compared to schemes, is that their theory is fully developed only over algebraically closed fields, but being true to the original Cartesian idea they certainly appeal to the geometric intuition.

Affine varieties (affine varieteter)

(13.8) The building blocks in scheme theory are the spectra of rings, and the similar role in the theory of varieties is played by the so-called *affine varieties*. These are zero loci in k^n (where k may be any field, but for the most it will be algebraically closed) of prime ideals in the polynomial ring $k[x_1, \dots, x_n]$. Slightly more general, for any ideal $\mathfrak{a}$ in

Closed algebraic subset (lukket algebraisk mengde)

$k[x_1, \dots, x_n]$ the *zero locus* of $\mathfrak{a}$—or the *closed algebraic subset* defined by $\mathfrak{a}$—is the subset $Z(\mathfrak{a}) \subseteq k^n$ of points where all the polynomials from $\mathfrak{a}$ vanish, or expressed in formulae:

$$Z(\mathfrak{a}) = \{\, (a_1, \dots, a_n) \in k^n \mid f(a_1, \dots, a_n) = 0 \text{ for all } f \in \mathfrak{a} \,\}.$$

The subsets $Z(\mathfrak{a})$ of k^n and $V(\mathfrak{a})$ in Spec $k[x_1, \dots, x_n]$ are closely related, but be careful not to confuse them: a point $(a_1, \dots, a_n) \in k$ belongs to $Z(\mathfrak{a})$ if and only if the maximal ideal $(x_1 - a_1, \dots, x_r - a_r)$ lies in $V(\mathfrak{a})$, however $V(\mathfrak{a})$ is a considerable larger set with all the prime ideals containing $\mathfrak{a}$ as members. (The admittedly bad *ad hoc* notation $Z(\mathfrak{a})$ is not very common, and is chosen just to avoid confusion with $V(\mathfrak{a})$. When fluent in the language of schemes, one would use V for both and say that $Z(\mathfrak{a})$ is the set of k-points in $V(\mathfrak{a})$) .

Another source of notational confusion is the different incarnations of the space k^n: it is a vector space and as such is denoted k^n; it has an incarnation as the spectrum of the polynomial ring, that is $\mathbb{A}^n_k = \text{Spec } k[x_1, \dots, x_n]$ (which is a different set from k^n); and it is the variety of k-points in $\mathbb{A}^n_k$ and then is denoted by $\mathbb{A}^n(k)$ or simply by k^n as the two are equal.

The Strong Nullstellensatz

Recall the converse of the Z-construction: for any subset $S \subseteq k^n$ the polynomials that vanish along S, form an ideal $I(S)$ in the polynomial ring, and the Nullstellensatz describes the relation between these to constructs. A simple but basic observation is that polynomials belonging to the radical $\sqrt{\mathfrak{a}}$ of $\mathfrak{a}$ all vanish along $Z(\mathfrak{a})$, and therefore one has $\sqrt{\mathfrak{a}} \subseteq I(Z(\mathfrak{a}))$. The Nullstellensatz tells us that this inclusion is an equality. This is also called the *Strong Nullstellensatz* since it is easily seen to imply the other versions.

THEOREM 13.9 (HILBERT'S NULLSTELLENSATZ) *Let k be an algebraically closed field and $\mathfrak{a}$ an ideal in $k[x_1, \dots, x_n]$. Then one has $I(Z(\mathfrak{a})) = \sqrt{\mathfrak{a}}$.*

(13.10) The Nullstellensatz has the following consequence, which in fact, is the wording of one of the weak versions: The locus $Z(\mathfrak{a})$ is empty if and only if the ideal $\mathfrak{a}$ is not a proper ideal; that is, if and only if $\mathfrak{a} = k[x_1, \dots, x_n]$ or equivalently, if and only if $1 \in \mathfrak{a}$. Indeed, a non-zero constant never vanishes, and for the other implication, requiring a function to vanish at all points in the empty set imposes no restriction, so $1 \in I(\varnothing)$.

Weak versions

(13.11) The first version we shall be discussing is slightly out of the line with the others. It has the virtue of being valid over any field k—also fields which are not algebraically closed—and is well adapted to Grothendieck's marvelous world of schemes. It is formulated purely in algebraic terms, and is readily deduced from the Normalization Lemma.

(13.12) As a motivating example consider an irreducible polynomial $f(x)$ in $k[x]$. It generates a maximal ideal, and it is well-known that the quotient $k[x]/(f(x))$ is a finite field extension of k; powers of the class of x with exponent less than the degree of f form a basis. This property, that quotients by maximal ideals are finite extensions of the ground field, is the one that generalizes to the case of several variables.

THEOREM 13.13 (GENERAL NULLSTELLENSATZ) *Let A be a finitely generated algebra over a field k and let $\mathfrak{m}$ be a maximal ideal in A. Then $A/\mathfrak{m}$ is a finite field extension of k.*

PROOF: The field $K = A/\mathfrak{m}$ is finitely generated as an k-algebra since A is. If it is not algebraic, it has transcendence degree at least one over k, say r, and by Noether's Normalization Lemma it is a finite module over a polynomial ring $k[w_1, \dots, w_r] \subseteq K$. By Lemma 12.28 on page 327 it ensues that $k[w_1, \dots, w_r]$ is a field, which is impossible since polynomial rings are not fields (if w is a variable, $1/w$ is certainly not a polynomial). Hence K is finite over the ground field k. ❑

(13.14) We then turn our attention to the two other weak versions of the Nullstellensatz and return to a situation where the ground field k is algebraically closed. The first version to be treated, is the one alluded to in the introduction; it describes the maximal ideals in the polynomial ring $k[x_1, \dots, x_n]$.

THEOREM 13.15 (WEAK NULLSTELLENSATZ II) *Let k be an algebraically closed field and let $\mathfrak{m}$ be a maximal ideal in the polynomial ring $k[x_1, \dots, x_n]$. Then $\mathfrak{m}$ is of the form $\mathfrak{m} = (x_1 - a_1, \dots, x_n - a_n)$ for a point $a = (a_1, \dots, a_n)$ in k^n.*

PROOF: By the general version of the Nullstellensatz above, the field $k[x_1, \dots, x_r]/\mathfrak{m}$ is a finite extension of k, and is therefore equal to k since k is assumed to be algebraically closed. Let $\pi\colon k[x_1, \dots, x_n] \to k$ be the ensuing quotient homomorphism. To retrieve the point a let $a_i = \pi(x_i)$. Then obviously all the polynomials $x_i - a_i$ lie in the kernel $\mathfrak{m}$ of π, and since *a priori* $(x_1 - a_1, \dots, x_n - a_n)$ is a maximal ideal, it must be equal to $\mathfrak{m}$, and we are through. ❑

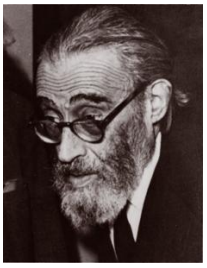
The second weak version asserts that the zero-locus $Z(\mathfrak{a})$ is non-empty whenever $\mathfrak{a}$ is a proper ideal. Note that it is crucial that the ground field be algebraically closed; if not, one easily finds examples that $Z(\mathfrak{a})$ is empty. For instance, just take $\mathfrak{a} = (x^2 + 1)$ in $\mathbb{R}[x]$.

THEOREM 13.16 (WEAK NULLSTELLENSATZ III) *Let k be algebraically closed and let $\mathfrak{a}$ be an ideal in the polynomial ring $k[x_1, \dots, x_n]$. Then $Z(\mathfrak{a})$ is non-empty if and only if the ideal $\mathfrak{a}$ is a proper ideal.*

PROOF: Since $\mathfrak{a}$ is a proper ideal, there is a maximal ideal in $k[x_1, \dots, x_n]$ containing $\mathfrak{a}$, which by the weak version II is of the form $(x_1 - a_1, \dots, x_n - a_n)$. And consequently we deduce that $(a_1, \dots, a_n) \in Z(\mathfrak{a})$. ❑

It is also quite easy to see that this last version implies the previous one: if $\mathfrak{m}$ is a maximal deal in $k[x_1,\dots,x_n]$, the last version implies that $Z(\mathfrak{m})$ is non-empty, say $(a_1,\dots,a_n) \in Z(\mathfrak{m})$. The maximal ideal $(x_1-a_1,\dots,x_n-a_n)$ will then contain $\mathfrak{m}$, but $\mathfrak{m}$ being maximal, the two must be equal.

The Rabinowitsch trick



George Yuri Rainich alias J.L. Rabinowitsch (1886–1968) Ukrainian–American mathematical physicist

We proceed to present the trick found by J. L. Rabinowitsch and published in a thirteen lines long paper in 1929 (shown *in extento* at the end of the chapter!) which proves that the third weak version of the Nullstellensatz (Theorem 13.16 above) implies the full version. Thus it accomplishes the proof of the Nullstellensatz.

LEMMA 13.17 *The Weak Nullstellensatz* III *implies that* $I(Z(\mathfrak{a})) \subseteq \sqrt{\mathfrak{a}}$ *for all ideals in polynomial rings* $k[x_1,\dots,x_n]$.

PROOF: The task is to demonstrate that $I(Z(\mathfrak{a})) \subseteq \sqrt{\mathfrak{a}}$ for any proper ideal $\mathfrak{a}$ in the polynomial ring $k[x_1,\dots,x_n]$, under the assumption that the zero-locus $Z(\mathfrak{b})$ is non-empty whenever $\mathfrak{b}$ is a proper ideal in a polynomial ring.

The crux of the trick is to introduce an auxiliary variable x_{n+1} and for each element $g \in I(Z(\mathfrak{a}))$ to consider the ideal $\mathfrak{b}$ in the polynomial ring $k[x_1,\dots,x_{n+1}]$ given by

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

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

According to the third version of the Weak Nullstellensatz the ideal $\mathfrak{b}$ is therefore not proper, so it holds that $1 \in \mathfrak{b}$, and there are polynomials f_i in $\mathfrak{a}$ and h_i and h in $k[x_1,\dots,x_{n+1}]$ satisfying a relation like

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

Substituting $x_{n+1} = 1/g$ in this relation and multiplying through by a sufficiently high power g^N of g (for instance, the highest power of x_{n+1} that occurs in any of the h_i's will suffice) we obtain

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

where each $H_i(x_1,\dots,x_n) = g^N\cdot h_i(x_1,\dots,x_n,g^{-1})$ is an element in $k[x_1,\dots,x_n]$. Hence $g \in \sqrt{\mathfrak{a}}$. ❑

Exercises

⁂ (13.4) Let k be a field not necessarily algebraic closed. Show that each maximal ideal in the polynomial ring $k[x,y]$ is of the form $(f(x),g(y))$ where f and g are irreducible polynomials whose *stem fields** E_f and E_g are linearly disjoint HINT: Recall Proposition 3.30 about polynomial rings over a PID. The concept of linearly disjoint extensions is discussed in Exercise 13.13 on page 355.

*Recall that the stem field of an irreducible polynomial f in $k[t]$ is the quotient $E_f = k[t]/(f)$

(13.5) Let k be a field not necessarily algebraically closed, and let $\mathfrak{m}$ be a maximal ideal in $k[x_1,\dots,x_n]$. Furthermore let $K = A/\mathfrak{m}$. Show that there are elements $b_1,\dots,b_n$ in K so that $\mathfrak{m} = (x_1 - b_1,\dots,x_n - b_n) \cap k[x_1,\dots,x_n]$. ★

There are other proofs of the Nullstellensatz much simpler than the one we have presented, simpler in the sense that they do not rely on substantial results in algebra, but uses only elementary algebra and standard field theory. It is worthwhile to ponder over some these proofs, and the two subsequent exercises guide you through two. The first is a classical proof valid over the complex numbers, and it is based on the complex field $\mathbb{C}$ being of infinite transcendence degree over $\mathbb{Q}$. It exhibit so called "generic points" in $Z(\mathfrak{p})$. The second is one of the simplest proofs around and uses barely more than there being infinitely many irreducible polynomials over any field.

(13.6) *A classic proof over the complex numbers* $\mathbb{C}$. This exercise is a guide through a classical proof of the Nullstellensatz in the weak form III, which basically is only valid for algebras over $\mathbb{C}$. It relies on the fact that $\mathbb{C}$ is of infinite transcendence degree over $\mathbb{Q}$ (you can take this for granted). For simplicity the exercise is confined to showing that *prime* ideals have non-empty zero loci, which is not a severe restriction as any ideal is contained in a maximal ideal. So let $\mathfrak{p}$ be a prime ideal in the polynomial ring $\mathbb{C}[x_1,\dots,x_n]$.

a) Prove that every field K of finite transcendence degree over $\mathbb{Q}$ can be embedded in $\mathbb{C}$. HINT: If $y_1,\dots,y_r$ is a transcendence basis for K over $\mathbb{Q}$, the field K will be algebraic over $\mathbb{Q}(y_1,\dots,y_r)$. Use that $\mathbb{C}$ is algebraically closed and of infinite transcendence degree over $\mathbb{Q}$.

b) Assume a finite set of generators $\{f_i\}$ for $\mathfrak{p}$ is given. Show that there is a finitely generated field extension k of $\mathbb{Q}$ such that each f_i lies in $R = k[x_1,\dots,x_n]$.

c) Let $\mathfrak{p}' = \mathfrak{p} \cap R$ and let K be the fraction field of $R/\mathfrak{p}'$. Show that K is of finite transcendence degree over $\mathbb{Q}$, and that there are embeddings of K in $\mathbb{C}$.

d) Conclude that there is a point in $Z(\mathfrak{p})$.

⁂ (13.7) *A most elementary proof.* This exercise describes the steps in one of the simples proof of the Nullstellensatz. It is completely elementary and relies only on elementary algebra and rudimentary field theory.

a) Let k any field, finite or not. Prove there are infinitely many irreducible polynomials in $k[x]$. HINT: Revive Euklid's good old proof that there are infinitely many

primes: assume $f_1, \ldots, f_r$ are the only irreducibles and consider $f_1 \cdot \ldots \cdot f_r + 1$.

b) Let k be a field. Show that the rational function field $k(t_1, \ldots, t_n)$ is not a finitely generated algebra over k. HINT: If it were, only finitely many factors would appear in denominators.

c) Assume that $k \subseteq K \subseteq L$ is a tower of field extension. Assume that L is a finite algebra over k, and at the same time is a finite dimensional vector space over K. Prove that K is finitely generated as an algebra over k. HINT: Let $\{f_j\}$ be algebra generators for L over k and $\{e_i\}$ a basis for L over K, both finite. Expand the f_j's and all the products $e_i e_j$ as combinations $f_j = \sum_i a_{ij} e_i$ and $e_i e_j = \sum_l b_{ijl}\, e_l$ with coefficients from K. Then the a_{ij}'s and the b_{ijl}'s will generate K as a k-algebra.

d) Deduce the General Nullstellensatz (Theorem 13.13).

★

13.3 *Consequences*

The dimension of polynomial rings

(13.18) One might be tempted to consider it intuitively evident that the Krull dimension of a polynomial ring in n variables is of equal to n. This is true, although astonishingly subtle to establish. However, mobilizing some of the heavier artillery in the arsenal, the proof will be straightforward.

PROPOSITION 13.19 (DIMESION OF POLYNOMIAL RINGS) *Let k be a field. The Krull dimension of the polynomial ring $k[x_1, \ldots, x_n]$ in n variables equals n.*

PROOF: We are going to prove that each maximal ideal $\mathfrak{m}$ in $k[x_1, \ldots, x_n]$ has height n. When the ground field is algebraically closed, the Weak Nullstellensatz II (Theorem 13.15 on page 347) says that $\mathfrak{m}$ is generated by n elements; indeed, $\mathfrak{m} = (x_1 - a_1, \ldots, x_n - a_n)$, and Krull's Height Theorem (Theorem 11.13) then yields that $\operatorname{ht} \mathfrak{m} \leqslant n$. On the other hand, there is the obvious chain of prime ideals

$$(0) \subset (x_1 - a_1) \subset (x_1 - a_2, x_2 - a_2) \subset \ldots \subset (x_1 - a_1, \ldots, x_n - a_n), \tag{13.3}$$

whose length is n, so that $\operatorname{ht} \mathfrak{m} \geqslant n$, and we may conclude that $\operatorname{ht} \mathfrak{m} = n$.

In the general case, let K be an algebraic closure of k and consider the extension $k[x, \ldots, x_n] \subseteq K[x_1, \ldots, x_n]$, which is integral by Proposition 12.22 on page 324 . By the Lying–Over Theorem each maximal ideal in $k[x_1, \ldots, x_n]$ is the contraction of one in $K[x_1, \ldots, x_n]$ and hence of height n by Corollary 12.43 on page 334. ❑

COROLLARY 13.20 *Let A be a domain finitely generated over the field k whose field of fractions is K. Then* $\dim A = \operatorname{trdeg}_k K$

PROOF: If $n = \operatorname{trdeg}_k K$, there is by Noether's Normalization Lemma a polynomial ring $k[x_1, \dots, x_n] \subseteq A$ over which A is finite, hence $\dim A = \dim k[x_1, \dots, x_n] = n$ by Going–Up III on page 333. ❑

(13.21) The Krull dimension of a ring cannot increase when the ring is localized—terms of chains can only disappear—and frequently it will drop. For example, the fraction field of a domain is of dimension zero regardless of what the dimension of the domain is. However, in some distinguished cases it stays the same:

COROLLARY 13.22 *Let A be a domain finitely generated over the field k and let $f \in A$ be a non-zero element. Then $\dim A_f = \dim A$.*

PROOF: The algebra A_f is finitely generated over k and has the same fraction field as A. ❑

EXERCISE 13.8 Let A be finitely generated domain over k and let $S \subseteq A$ be a multiplicatively closed set. Assume that $\operatorname{Spec} S^{-1}A$ (identified with the subset of $\operatorname{Spec} A$ of primes not meeting S) is dense in $\operatorname{Spec} A$. Show that $\dim S^{-1}A = \dim A$. ★

✲ EXERCISE 13.9 Given an example of an algebra A of finite type over a field k with a non-zero elements f such that $\dim A_f < \dim A$. HINT: Consider A's such that $\operatorname{Spec} A$ has components of different dimensions. ★

Uniform altitude

Not being catenary is a kind of pathology for, but fortunately most of the rings one meets when practicing algebraic geometry do not suffer from that shortcoming. In this section we shall show that the favourites of this chapter, the domains of finite type over a field, are of uniform altitude; that is, all maximal chains of prime ideals have the same length. The proof is a reduction to the case of polynomial rings, using Noether's Normalization Lemma, combined with an induction argument on the dimension.

(13.23) The induction step requires a little lemma about the dimension of hypersurfaces in affine space. Examples show that in general $\dim A/(f)$ may drop by two or more compared to $\dim A$ even when A is noetherian. However, the transcendence degree behaves better, as the following lemma indicates, and combined with Corollary 13.20, this tells us that the dimension of a hypersurface in affine space is what it should be:

LEMMA 13.24 *Let k be a field and $f \in k[x_1, \dots, x_n]$ an irreducible polynomial. It then holds true that $\dim k[x_1, \dots, x_n]/(f) = n - 1$.*

PROOF: If K denotes the fraction field of $A = k[x_1, \dots, x_n]/(f)$, it will, according to Corollary 13.20, be sufficient to see that $\operatorname{trdeg}_k K = n - 1$. After renaming the variables, if need be, we may assume that f is of positive degree in x_1. We claim that the classes of the remaining variables $x_2, \dots, x_n$ then will be algebraically independent over k. If not, there would be a polynomial $g(x_2, \dots, x_n)$ with coefficients in k belonging to the

ideal (f), so that we might write $g = hf$ for some polynomial h. But this is absurd since g is of degree zero in x_1 whereas f is not. It follows that K is algebraic over the rational function field $k(x_2, \ldots, x_n)$, and we are through. ❑

THEOREM 13.25 *Let A be an algebra which is finitely generated over a field k. Then all maximal chains of prime ideals ascending from a given minimal prime are of the same length. Moreover, if A is a domain, it is of uniform altitude.*

PROOF: If $\mathfrak{q}$ is the given minimal prime ideal in A, there is a one-to-one correspondence between chains in A with smallest member $\mathfrak{q}$ and chains in $A/\mathfrak{q}$; hence replacing A by $A/\mathfrak{q}$, we may assume that A is a domain, and the minimal prime is then of course the zero ideal.

The proof goes by induction on the dimension of A, so let $n = \dim A$. By the Normalization Lemma we my find a polynomial ring $B = k[x_1, \ldots, x_n]$ contained in A and such that A is finite over B. Consider a maximal chain

$$0 \subset \mathfrak{p}_1 \subset \ldots \subset \mathfrak{p}_r \tag{13.4}$$

of prime ideals in A. By the Going–Down Theorem (or more precisely its Corollary 12.43), heights are preserved in the extension $B \subseteq A$ (B is normal being a polynomial ring, and both rings are domains), and the ideal $\mathfrak{p}_1 \cap B$ is therefore a height one ideal. Thus, since B is a UFD, the ideal $\mathfrak{p}_1 \cap B$ is principal; say $\mathfrak{p}_1 \cap B = (f)$. The extension $B/(f) \subseteq A/\mathfrak{p}_1$ is integral (Proposition 12.17 on page 322) so that $\dim A/\mathfrak{p}_1 = \dim B/(f)$ by Going–Up III. Now, the lemma above tells us that $\dim B/(f) = n - 1$, and hence $\dim A/\mathfrak{p}_1 = n - 1$. By induction $\dim A/\mathfrak{p}_1$ is of uniform altitude, and since the chain

$$0 \subset \mathfrak{p}_2/\mathfrak{p}_1 \subset \ldots \subset \mathfrak{p}_r/\mathfrak{p}_1$$

induced by the chain (13.4) is a maximal chain, it follows that $r - 1 = n - 1$; in other words $r = n$. ❑

EXERCISE 13.10 Show that if a ring is catenary so is any localization. Show that if a ring is of uniform altitude so is any localization ★

Density and Jacobson rings

An all important consequence of the Nullstellensatz is that when A is an algebra of finite type over a field, the maximal ideals in Spec A form a dense subset; and the even (apparently) stronger property holds true: for any closed subset $Z \subseteq$ Spec A the maximal ideals that belong to Z, form a dense subset of Z. This turns out be extremely useful and accounts for a lot of the good properties of algebras of finite type over a field. Heuristically, one is tempted by the explanation that two "functions" (*i. e.* elements in A) assuming the same "values*" at all closed points are equal.

*with the understanding that values at different points are taken in varying fields

This density property is an important geometric feature shared by many rings not of finite type over a field. The rings for which each closed set $V(\mathfrak{a})$ is the closure its closed point;t hat is, of set of the maximal ideals containing $\mathfrak{a}$, are termed *Hilbert rings* by some and *Jacobson rings* by others, but to day Jacobson rings seems to be the most commonplace usage.

Hilbert or Jacobson rings (Hilbert eller Jacobson ringer)

LEMMA 13.26 *A ring A is a Jacobson ring if and only one of the two following equivalent conditions is satisfied:*

i) Every radical ideal is the intersection of the maximal ideals which contain it;
ii) Every prime ideal is the intersection of the maximal ideals which contain it.

PROOF: Every radical ideal equals the intersection of the prime ideals containing it (Proposition 2.58 on page 51) so the two conditions are equivalent.

For any subset $S \subseteq \operatorname{Spec} A$ it holds true that the closure of S equals $V(\bigcap_{\mathfrak{p}\in S} \mathfrak{p})$; indeed, saying that $\mathfrak{a} \subseteq \mathfrak{p}$ for all $\mathfrak{p} \in S$, is equivalent to saying that $\mathfrak{a} \subseteq \bigcap_{\mathfrak{p}\in S} \mathfrak{p}$. Hence the closure of the set of maximal ideals containing $\mathfrak{a}$ equals $V(\mathfrak{a})$ if and only if $V(\mathfrak{a}) = V(\bigcap_{\mathfrak{a}\subseteq \mathfrak{m}} \mathfrak{m})$; that is, if and only if $\bigcap_{\mathfrak{a}\subseteq \mathfrak{m}} \mathfrak{m} = \sqrt{\mathfrak{a}}$. ❑

(13.27) Of course, rings in general are not Jacobson; an obvious counter example would be any non-Artinian local ring, it has just one maximal ideal, which it is not the only prime ideals, so that the spectrum is not reduced to the sole closed point. Algebras that are finitely generated over a field, however, are all Jacobson. At the bottom this is a corollary of the Nullstellensatz, but some technical help of the Going-Up results is needed in the reduction to the case of algebraically closed ground fields.

PROPOSITION 13.28 *Let A be an algebra finitely generated over the field k and let $\mathfrak{a}$ be an ideal in A. Then the radical $\sqrt{\mathfrak{a}}$ equals the intersection of the maximal ideals containing $\mathfrak{a}$; in other words, A is Jacobson.*

PROOF: The algebra A is by assumption a quotient of a polynomial ring $k[x_1, \dots, x_n]$, and replacing $\mathfrak{a}$ by the inverse image in $k[x_1, \dots, x_n]$, we may well assume that A is a polynomial ring.

We begin with doing the case when k is algebraically closed. A maximal ideal, which is shaped like $(x_1 - a_1, \dots, x_n - a_n)$, contains $\mathfrak{a}$ if and only if the point $(a_1, \dots, a_n)$ belongs to $Z(\mathfrak{a})$. So if f lies in all these maximal ideals, it vanishes along $Z(\mathfrak{a})$, and by the Nullstellensatz, it lies in $\sqrt{\mathfrak{a}}$.

Proceeding with the general case, we let K denote an algebraic closure of k, and we may well assume that $\mathfrak{a}$ is a prime ideal. The extension of polynomial rings $A = k[x_1, \dots, x_n] \subseteq K[x_1, \dots, x_n] = B$ is integral because $k \subseteq K$ is (Proposition 12.22 on page 324), and citing the Lying–Over Theorem (Proposition 12.31 on page 328) we may chose a prime ideal $\mathfrak{q}$ lifting $\mathfrak{p}$. By the case when the ground field is algebraically closed,

$\mathfrak{q}$ equals the intersection of the maximal ideals containing it; *i. e.* it holds true that $\mathfrak{q} = \bigcap_{\mathfrak{q} \subseteq \mathfrak{m}} \mathfrak{m}$. Consequently

$$\mathfrak{p} = \mathfrak{q} \cap A = (\bigcap_{\mathfrak{q} \subseteq \mathfrak{m}} \mathfrak{m}) \cap A = \bigcap_{\mathfrak{q} \subseteq \mathfrak{m}} (\mathfrak{m} \cap A),$$

and we are through since by Corollary 12.29 on page 328 each $\mathfrak{m} \cap A$ is a maximal ideal in A. ❑

Let us remark by the way that the last part of the proof shows that integral extensions of Jacobson rings are Jacobson.

Tensor products of domains

A most useful feature varieties over algebraically closed fields have is the unrestricted possibility to form cartesian products. Schemes share this property, and when both factors are spectra, say Spec A and Spec B, the product in the category of schemes turns out to be Spec $A \otimes_k B$. Affine varieties (when considered as schemes) are spectra of integral domains finitely generated over a field k, and so if Spec B and Spec A are two of the kind, the pertinent question arises wether the tensor product $A \otimes_k B$ is a domain or not (it is clearly finitely generated over k). The answer is the subject of this section; when k is algebraically closed, the answer is yes, but for general k's the matter is subtle. (13.29) The simplest example one can imagine, shows that the answer is no in general. Just consider $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$, which is easily shown to be isomorphic as an $\mathbb{R}$-algebra to the product $\mathbb{C} \times \mathbb{C}$. More generally, any separable field extension $K = k[t]/f(t)$ of k suffers the same fate: $K \otimes_k K$ is not a domain. Indeed, for each root α of $f(t)$ in a root field E of f, there is an evaluation map $K \to k(\alpha)$, and these together induce an isomorphism $K \otimes_k K \simeq K[t]/f(t) \simeq K \times \ldots \times K$ with as many factors as f has roots.

Even nilpotents may appear, and this typically happens when the extension $k \subseteq K$ is inseparable. For instance, if the characteristic of k equals p and $K = k(b)$ where b is a p^{th}-root of some $a \in k$ (which has no p^{th}-root in k), one finds

$$K \otimes_k K \simeq K[t]/(t^p - a) = K[t]/(t^p - b^p) = K[t]/(t - b)^p.$$

(13.30) The cases above are all built on a polynomial getting new roots in an extension, and indeed, this is at the root of the phenomenon, and is hindered when the ground field is algebraically closed:

PROPOSITION 13.31 *Let k be an algebraically closed field and A and B two finitely generated k-algebras.*

i) If both A and B are reduced, the tensor product $A \otimes_k B$ is reduced as well;

ii) If both A and B are domains, then $A \otimes_k B$ is a domain as well.

PROOF: By the Nullstellensatz every quotient $B/\mathfrak{m}$ of B by a maximal ideal $\mathfrak{m}$ is equal to k. We shall temporarily adopt the following suggestive notation and write $b(x)$ for the class in $B/\mathfrak{m}$ of an element $b \in B$, where x stands for the point $\mathfrak{m}$ in Spec A. Note further that every element $f \in A \otimes_k B$ may be expressed as $f = \sum_{1 \leqslant i \leqslant t} a_i \otimes b_i$ with $a_i \in A$ and $b_i \in B$ and with the a_i's being linearly independent over k (see Exercise 6.12 on page 160).

Proof of *i)*: Assume that f is a nilpotent element in $A \otimes_k B$, say $f^n = 0$, and write $f = \sum a_i \otimes b_i$ as above. Let $\mathfrak{m}$ be a maximal ideal in B and consider the image $\bar{f} = \sum_i a_i b_i(x)$ in $A \otimes_k B/\mathfrak{m} = A$. Then $\bar{f}^n = 0$, and since A is reduced and the a_i's are linearly independent, it ensues that all $b_i(x) = 0$. In other words, each $b_i \in \mathfrak{m}$, and since this holds for all $\mathfrak{m}$, it follows that each $b_i = 0$; indeed, according to Proposition 13.28 above $\bigcap_{\mathfrak{m} \subseteq A} \mathfrak{m} = \sqrt{(0)}$, and $\sqrt{(0)} = 0$ as B is assumed to be reduced.

Proof of *ii)*: We keep our element f; it will not be nilpotent anymore, but we assume that $fg = 0$ for another element g and write $g = \sum_{1 \leqslant i \leqslant s} c_i \otimes d_i$ with the c_i's linearly independent over k. For every maximal ideal $\mathfrak{m}$ in B it then holds that $0 = \bar{f}\bar{g}$ in $A \otimes_k B/\mathfrak{m} = A$. Hence, either $\bar{f} = 0$ and all $b_i(x)$ vanish, or $\bar{g} = 0$ and all the $d_i(x)$'s do; that is, either $\mathfrak{b} = (b_1, \ldots, b_t) \subseteq \mathfrak{m}$ or $\mathfrak{d} = (d_1, \ldots, d_s) \subseteq \mathfrak{m}$. It follows that $\mathfrak{b} \cap \mathfrak{d}$ lies in every maximal of B, and because B is Jacobson, we infer that $\mathfrak{b} \cap \mathfrak{d} = (0)$. Thus either $\mathfrak{b} = (0)$ and $f = 0$, or $\mathfrak{d} = (0)$ and $g = 0$ because (0) is a prime ideal in B. ❑

Exercises

(13.11) This exercises fills in the details of the first example in Paragraph 13.29. Let k be a field and $f(t)$ a separable and irreducible polynomial in $k[t]$. Denote by K the extension $K = k[t]/(f(t))$ of k. Let E be a root field of f and let $\alpha_1, \ldots, \alpha_n$ be the roots of f in E. Define $\phi_i \colon K[t]/(f(t)) \to k(\alpha_i)$ by sending a polynomial $p(t)$ to $p(\alpha_i)$.

a) Show there is an isomorphism $K \otimes_k K \simeq K[t]/(f(t))$;

b) Show that the maps ϕ_i are well defined and together yield an isomorphism $K[t]/f(t) \to k(\alpha_1) \times \ldots \times k(\alpha_n)$, which sends a polynomial p to the tuple $(\phi_1(p), \ldots, \phi_n(p))$.

✲ (13.12) The hypothesis in Proposition 13.31 that A and B be finitely generated over k is not necessary. HINT: Show first that if $A' \subseteq A$ and $B' \subseteq B$ are subalgebras, then $A' \otimes_k B' \subseteq A \otimes_k B$, then reduce to the finite type case.

✲ (13.13) *Linearly disjoint field extensions.* If you wonder when the tensor product of two fields is a field, you probably will appreciate this exercise. Let $k \subseteq E$ and $k \subseteq F$ be two algebraic field extension and assume they appear in towers $k \subseteq E \subseteq K$ and $k \subseteq F \subseteq K$ in a field K. The *compositum* EF of E and F is the smallest subfield of K containing both. One says E and F are *linearly disjoint* if every set of elements from E which are linearly independent over k, stay linearly independent over F when considered elements in EF. The condition is *a priori* asymmetric in the two fields, but will in fact turn out to be

Compositum (kompositum)

Linearly disjoint fields (lineært ukoblede kropper)

symmetric, and it will even be independent of the embeddings in K, as follows from subproblem *c*).

a) Show that the assignment $e \otimes f \mapsto ef$ extends to a k-algebra homomorphism $\phi\colon E \otimes_k F \to EF \subseteq K$.

b) Show that ϕ is always surjective, and is an isomorphism if and only if E and F are linearly disjoint.

c) Show that $E \otimes_k F$ is a field if and only if E and F are linearly disjoint.

d) Assume E and F to be finite extensions of k. Show that E and F are linearly disjoint if and only if $[EF : F] = [E : k][F : k]$.

★

Zum Hilbertschen Nullstellensatz.

Von

J. L. Rabinowitsch in Moskau.

Satz. *Verschwindet das Polynom $f(x_1, x_2, \ldots, x_n)$ in allen Nullstellen — im algebraisch abgeschlossenen Körper — eines Polynomideals $\mathfrak{a}$, so gibt es eine Potenz f^ϱ von f, die zu $\mathfrak{a}$ gehört.*

Beweis. Es sei $\mathfrak{a} = (f_1, f_2, \ldots, f_r)$, wo f_i die Variablen $x_1, \ldots, x_n$ enthalten. x_0 sei eine Hilfsvariable. Wir bilden das Ideal $\bar{\mathfrak{a}} = (f_1, f_2, \ldots, f_r, x_0 f - 1)$. Da der Voraussetzung nach $f = 0$ ist, sobald alle f_i verschwinden, so hat das Ideal $\bar{\mathfrak{a}}$ keine Nullstellen.

Folglich muß $\bar{\mathfrak{a}}$ mit dem Einheitsideal zusammenfallen. (Vgl. etwa bei K. Hentzelt, „Eigentliche Eliminationstheorie", § 6, Math. Annalen 88 [1].) Ist also $1 = \sum_{i=1}^{i=r} F_i(x_0, x_1, \ldots, x_n) f_i + F_0 \cdot (x_0 f - 1)$ und setzen wir in dieser Identität $x_0 = \frac{1}{f}$, so ergibt sich:

$$1 = \sum_{i=1}^{i=r} F_i\left(\frac{1}{f}, x_1, \ldots, x_n\right) f_i = \frac{\sum_{i=1}^{i=r} \bar{F}_i f_i}{f^\varrho}.$$

Folglich ist $f^\varrho \equiv 0\,(\mathfrak{a})$, w. z. b. w.

[1]) Folgt auch schon aus der Kroneckerschen Eliminationstheorie.

(Eingegangen am 8. 5. 1929.)

Lecture 14

Examples of unexpected rings

This chapter is devoted to a series of examples—more of less of a pathological kind—that show what peculiar phenomena even Noetherian rings can experience. All these rings are denizens dwelling in the underworld of non-geometric rings, and you wouldn't meet them a bright day doing algebraic geometry, but be aware, they lurk near the boarder, so take care. To be serious: we think it is important and that it improves the understanding to work out some of the borderline cases in detail; remember the words of Sun Tzu about knowing your adversary. Moreover constructing these examples gives us the opportunity to practise many of the techniques taught in the course, and in the end, they are beautiful pieces of algebra.

The first of the examples we are about to give, is a classic: Nagata's example that Noetherian rings may be of infinite dimension. His famous book "Local rings" [**?**] ends with the description of a series of peculiar rings having unexpected properties, and this is the first one. The second example is due to Krull who in [**?**] gave it as an example of a normal domain with just one non-zero prime ideal that is not a valuation ring, and it also appears as the first in a series of examples constructed by Seidenberg ([**?**, **?**] that the dimension of $A[t]$ can take any value in the admissible range. Our third example is a domain that is not catenary, and finally, we construct a peculiar Noetherian ring as an infinite limit of affine blowups; in characteristic p it yields an example of the Finiteness Theorems (Theorems 12.45 and 12.47 on page 335) not being valid without the separability condition.

14.1 A Noetherian ring of infinite dimension

(14.1) We proceed to describe Nagata's example of a Noetherian ring A of infinite Krull dimension. Each maximal ideal is of course of finite height, but there are maximal ideals of arbitrary high height.

The starting point is the ring $k[x_1, x_2, \ldots]$ of polynomials in countably many variables over a field k, and the construction depends on a decomposition $\mathbb{N} = \bigcup_i I_i$ of the natural numbers into a disjoint union of finite subsets I_i so that the cardinality $\#I_i$ tends to infinity when i does. For instance, the decomposition can be as simples as the one with $I_1 = \{1\}$, $I_2 = \{2, 3\}$, $I_3 = \{4, 5, 6\}$ an so forth, with I_i consisting of i consecutive numbers.

For each natural number i we let $\mathfrak{q}_i$ be the prime ideal generated by the variables x_j with $j \in I_i$; so in the example just mentioned $\mathfrak{q}_1 = (x_1)$, $\mathfrak{q}_2 = (x_2, x_3)$ etc. Moreover, S will be the set of elements in $k[x_1, x_2 \ldots]$ not belonging to any of the $\mathfrak{p}_i$'s; in other words, it is the set of polynomials with a non-zero constant term. Clearly S is multiplicatively closed, and we let A be the localization of $k[x_1, x_2, \ldots]$ in S; that is, $A = S^{-1}k[x_1, x_2, \ldots]$. Furthermore, we put $\mathfrak{p}_i = \mathfrak{q}_i A$; they constitute all the maximal ideals in A.

The crucial observation is that when fixing an index i and localizing $k[x_1, x_2, \ldots]$ in the set* T of all none-zero polynomials in the variables x_j with $j \notin I_i$, we obtain a polynomial ring over a certain field* K_i, namely the rational function field $K_i = K(x_j | j \notin I_i)$ in the (infinite many) variables x_j with $j \notin I_i$. And the variables of the polynomial ring are of course the x_j's for which $j \in I_i$. We further observe that among all the $\mathfrak{p}_\mathfrak{j}$'s the ideal $\mathfrak{p}_\mathfrak{i}$ is the only one that survives as a proper ideal in $T^{-1}k[x_1, x_2 \ldots]$ since all the others meet T. A subsequent localization in S yields

*which obviously is multiplicatively closed

*see also Problem 7.32 on page 189

$$A_{\mathfrak{p}_i} = S^{-1}T^{-1}k[x_1, x_2, \ldots] = K_i[x_j | j \in I_i]_{\mathfrak{m}_\mathfrak{i}},$$

where $\mathfrak{m}_i$ is the maximal ideal $\mathfrak{m}_\mathfrak{i} = (x_j | j \in I_i)$ in $K_i[x_j | j \in I_i]$. This shows that $A_{\mathfrak{p}_i}$, being the localization of a polynomial ring at a maximal ideal, is Noetherian, and moreover, its dimension equals the cardinality $\#I_i$. It follows that A has infinite Krull dimension since $\#I_i$ tends to infinity with i.

It remains to see that A is Noetherian, which ensues from the following lemma:

LEMMA 14.2 *Assume that A is a ring such that all localizations $A_\mathfrak{m}$ at maximal ideals are Noetherian and that any non-zero element in A is contained in only finitely many maximal ideals. Then A is Noetherian.*

PROOF: Let $\mathfrak{a}$ be an ideal in A. For each maximal ideal $\mathfrak{m}$ containing $\mathfrak{a}$, the ideal $\mathfrak{a}A_\mathfrak{m}$ is finitely generated (because $A_\mathfrak{m}$ is Noetherian), and the generators may be chosen to lie in A. Since $\mathfrak{a}$ is only contain in finitely many maximal ideals, recollecting all such generators we get a finite set $\{f_i\}$. Now, consider a non-zero element $a \in \mathfrak{a}$. It lies in finitely many maximal ideals, some of which contain $\mathfrak{a}$ and some of which do not. Let $\{\mathfrak{m}_j\}$ be the ones that do not contain $\mathfrak{a}$, and for each j, chose an element $a_j \in \mathfrak{a}$ not in $\mathfrak{m}_j$. We contend that the f_i's together with a and the a_j's generated $\mathfrak{a}$. Indeed, these elements

were picked so that they generated $\mathfrak{a}A_{\mathfrak{m}}$ for every maximal ideal $\mathfrak{m}$: if $\mathfrak{a} \subseteq \mathfrak{m}$, the f_i's will generate, if $\mathfrak{a} \nsubseteq \mathfrak{m}$ and $a \notin \mathfrak{m}$, the element a will generate, and finally, if $\mathfrak{a} \nsubseteq \mathfrak{m}$ and $a \in \mathfrak{m}$, the a_j's will do. We may thus conclude since surjectivity is a local property. ❑

⁂ EXERCISE 14.1 This exercise is a generalization of Lemma 14.2 above due to William Heinzer and Jack Ohm. We are given a ring A and a family $\{A_i\}_{i\in I}$ of flat A-algebras.

a) Show that if B as flat A-algebra, then $(\mathfrak{a} : x)B = (\mathfrak{a}B : x)$ for all ideals $\mathfrak{a}$ and all elements x in A .

b) Assume that for each maximal $\mathfrak{m}$ in A, the ideal $\mathfrak{m}A_i$ is proper for at least one i. Show that if $\mathfrak{b}$ is an ideal in A such that $\mathfrak{b}A_i \subseteq \mathfrak{a}A_i$ for all i, then $\mathfrak{b} \subseteq \mathfrak{a}$.

c) Assume in addition to the assumption in b) that there is a finitely generated ideal $\mathfrak{b} \subseteq \mathfrak{a}$ such that $\mathfrak{b}A_i \neq \mathfrak{a}A_i$ for at most finitely many i, and that $\mathfrak{a}A_i$ is finitely generated for all i. Show that $\mathfrak{a}$ is finitely generated.

d) Assume that each A_i is Noetherian. Assume further that for each proper ideal $\mathfrak{a}$ the ideal $\mathfrak{a}A_i$ is proper for at least one and for at most finitely many i. Show that A is Noetherian.

★

14.2 *A polynomial ring of excess dimension*

The following construction of Krull's furnishes an example of (a necessarily non-Noetherian) domain having a polynomial ring of pathological large dimension: we exhibit a one-dimensional local ring A such that the dimension of $A[t]$ equals three; that is $\dim A[t] = \dim A + 2$. This also is an example of a normal domain with just one non-zero prime ideal that is not a valuation ring; that is, it is not Noetherian.

(14.3) The ring A is no more exotic than the ring of rational functions $f(x, y)$ in two variables over a field k which are defined and constant on the y-axis. The elements of A, when written in lowest terms, have a denominator not divisible by x, and $f(0, y)$, which then is meaningful, lies in k.

A crucial observation when trying to understand A, is that no rational function $g(y)$ depending only on y, belongs to A unless it is constant. On the other hand, for every rational function $g(y)$ the product $xg(y)$ lies in A since it vanishes on the y-axis.

The ring A is not Noetherian; for instance, the principal ideals (xy^{-i}) with $i \in \mathbb{N}$ form an ascending chain which does not stabilize; indeed, a relation $xy^{-(i+1)} = f(x, y) \cdot xy^{-i}$ would give $f(x, y) = y^{-1}$, which is not constant along the y-axis (neither is it well defined) and hence does not belong to A.

The elements of A are described in the following lemma:

LEMMA 14.4 *Every $f \in A$ may be written as $x^\nu g(y)\alpha$ where ν is a non-negative integer, $\alpha \in A$ is a unit and $g(y)$ is a rational function in y alone which is constant if $\nu = 0$.*

PROOF: One has the equality

$$f = x^\nu \cdot \frac{a(y) + xb(x,y)}{c(y) + xd(x,y)} = x^\nu \cdot \frac{a}{c} \cdot \frac{1 + xba^{-1}}{1 + xdc^{-1}}, \tag{14.1}$$

where a, b, c and d are polynomials with a and c being non-zero and depending on y alone, and where ν is a non-negative integer. Such a representation is indeed possible: when f is written in lowest terms, the denominator cannot have x as factor because f is well defined on the y-axis; this accounts for the exponent ν being non-negative and the function $c(y)$ being non-zero. Furthermore, $a(y)$ is non-zero when the maximal power of x is extracted from the enumerator, and the right most fraction in (14.1) will then be a unit in A. ❑

Evaluating functions in A on the y-axis (they are by definition well-defined and constant there) we obtain a ring homomorphism $A \to k$ whose kernel is a maximal ideal $\mathfrak{m}$ in A, and it turns out that this is the sole non-zero prime ideal in A:

LEMMA 14.5 *The maximal ideal $\mathfrak{m}$ is generated by the elements $xg(y)$ with $g(y)$ a rational function, and it is the only non-zero prime ideal in A. Consequently, A is of dimension one.*

PROOF: That $\mathfrak{m}$ is generated by the elements shaped like $xg(y)$ follows immediately from the previous lemma. Assume then that $\mathfrak{p}$ is a non-zero prime ideal, and let $x^i h(y)$ be a non-zero element in $\mathfrak{p}$ with $i \geqslant 0$ (there are such according to the previous lemma). Since $x \cdot h(y)^{-1}$ belongs to A, we infer that $x^{i+1} = (x^i h(y)) \cdot (xh(y)^{-1}) \in \mathfrak{p}$, and $\mathfrak{p}$ being prime, it ensues that $x \in \mathfrak{p}$. It follows that $(xg(y))^2 = x \cdot xg(y)^2 \in \mathfrak{p}$ for each $g(y)$, and again as $\mathfrak{p}$ prime, we infer that $xg(y) \in \mathfrak{p}$, and by consequence it holds that $\mathfrak{m} = \mathfrak{p}$. ❑

(14.6) The polynomial ring $A[t]$ has dimension three. It has the prime ideals 0, $\mathfrak{m}A[t]$ and $(t) + \mathfrak{m}A[t]$, but there is also a forth one, namely the ideal $\mathfrak{p}$ consisting of the polynomials $F(t)$ such that $F(y) = 0$. This ideal is not the zero ideal as $xt - xy$ lies there, and it is contained in $\mathfrak{m}A[t]$: assume namely that $F(t) = \sum_i r_i(x,y)t^i \in \mathfrak{p}$. Then substituting y for t, gives $0 = F(y) = \sum_i r_i(x,y)y^i = 0$, which with $x = 0$ yields $\sum_i r_i(0,y)y^i = 0$. By definition of A the functions $r_i(0,y)$ are constants belonging to k, and since the different powers y^i are linearly independent over k, it ensues that $r_i \in \mathfrak{m}$; that is, $F(t) \in \mathfrak{m}A[t]$.

(14.7) Let us close the subsection by showing that A is integrally closed in the fraction field $k(x,y)$, and this is where the hypothesis that k be algebraically closed comes into play.

LEMMA 14.8 *The ring A is integrally closed in $k(x,y)$ when k is algebraically closed.*

PROOF: Let $f \in k(x,y)$ be an element integral over A which satisfies the dependence relation

$$f^n + a_{n-1}f^{n-1} + \ldots + a_1 f + a_0 = 0. \tag{14.2}$$

Assume first that f is well defined on the y-axis. With $x = 0$ the relation (14.2) becomes a dependence relation for $f(0,y)$ over the field of constants k, but from k being algebraically closed it then follows that $f(0,y) \in k$. Hence $f \in A$.

It remains to see that f is well defined on the y-axis. So let $f = x^{-i}g$ with g well defined and not identically zero on the y-axis; multiply through by x^{ni}, the relation (14.2) becomes transformed into

$$g^n + x^i a_{n-1} g^{n-1} + \ldots + x^{ni} a_0 = 0$$

which shows that $g(0,y) = 0$ if $i > 0$, contradiction. Hence $f = g$, and it is well defined on the y-axis. ❑

14.3 *A Noetherian ring that is not catenary*

There is an easy way to construct such rings, which we shall follow. It is related to Masayoshi Nagata's original approach when he gave the first example of Noetherian non-catenary domains in [**?**], but with a simplifying twist. The price to pay for simplicity, however, is that the resulting rings will be pretty weird. There are several other examples of a much more geometric flavour albeit none of finite type over fields, but they are Jacobson rings— Tetsushi Ogoma in [**?**] even gave examples of normal such rings—but constructing these examples is a much more involved business than we are about to undertake.

(14.9) The crucial part of the construction is a kind of "pincer movement". One starts with a semi local domain R with two maximal ideals $\mathfrak{m}_1$ and $\mathfrak{m}_2$ whose residue fields are isomorphic, say to k, and constructs a subring A of R such that $\mathfrak{m}_1 \cap A = \mathfrak{m}_2 \cap A = \mathfrak{n}$, and such that A will be local with maximal ideal $\mathfrak{n}$. The construction depends (seriously) on the isomorphisms $R/\mathfrak{m}_i \simeq k$, so once and for all we fix these, and we let π_i be the composition of the canonical reduction map and the fixed isomorphism: that is, the composition $\pi_i \colon R \to R/\mathfrak{m}_i \xrightarrow{\simeq} k$.

The induced map Spec $R \to$ Spec A identifies the two points corresponding to the maximal ideals, both are mapped to $\mathfrak{n}$, and importantly, nothing much more happens: the map will be bijective away from the maximal ideals. The staging is shown by the diagram

$$\begin{array}{ccc} R & \longrightarrow & k \times k \\ \uparrow & & \uparrow \\ A & \longrightarrow & k \end{array}$$

where the uppermost map is the sum $\pi_1 + \pi_2$, and where $k \to k \times k$ is the diagonal. The ring A with the two remaining maps are defined by the diagram being Cartesian, which means A is the subring of R given as $A = \{\, x \in R \mid \pi_1(x) = \pi_2(x) \,\}$. The main properties of A is described in the next lemma:

LEMMA 14.10 *In the situation above the following two statements hold true*

i) The ring A is local with maximal ideal $\mathfrak{n} = \mathfrak{m}_1 \cap \mathfrak{m}_2$;

ii) The map Spec $R \to$ Spec A *takes each $\mathfrak{m}_i$ to $\mathfrak{n}$, and for each finite set of prime ideals $\{\mathfrak{p}_i\}$ in A, all different from $\mathfrak{n}$, there is an $f \in \mathfrak{n}$ avoiding all the $\mathfrak{p}_i$'s, so that $A_f = R_f$;*

PROOF: The key observation is that the intersection $\mathfrak{n} = \mathfrak{m}_1 \cap \mathfrak{m}_2$ is entirely contained in A since both maps π_i vanish there. It also is the kernel of each restriction $\pi_i|_A$ (which coincide) and is therefore a maximal ideal. One thus has

$$\mathfrak{m}_1 \cap \mathfrak{m}_2 = \mathfrak{m}_1 \cap A = \mathfrak{m}_2 \cap A.$$

And $\mathfrak{n}$ is the only maximal ideal in A since if a is invertible in R and $\pi_1(a) = \pi_2(a)$ obviously $\pi_1(a^{-1}) = \pi_2(a^{-1})$ and $a^{-1} \in A$. This proves *i)*.

We proceed with the proof of *ii)*: by Prime Avoidance there is an element f in $\mathfrak{n}$ not belonging to any of the $\mathfrak{p}_i$'s. It lies in both the maximal ideals $\mathfrak{m}_i$, and therefore it kills each quotient $R/\mathfrak{m}_i$. Now, one has the exact sequence of A-modules

$$0 \longrightarrow A \longrightarrow R \xrightarrow{\rho} k \longrightarrow 0. \tag{14.3}$$

Here ρ equals $\pi_1 - \pi_2$; that is, the composition $R \to k \oplus k \to k$ of the map $\pi_1 + \pi_2$ and the map that sends a pair to the difference. Note that this map is A-linear, but not R-linear. Since each $R/\mathfrak{m}_i$ is annihilated by f, the sequence (14.3) when localized shows that $A_f = R_f$. The rest follows immediately: there are one-to-one correspondences between prime ideals in A (respectively in R) not containing f and prime ideals in A_f (respectively in R_f). ❏

(14.11) The statement *ii)* is particularly pertinent in our context. It entails that maximal chains in R ascending to one of the maximal ideals survive unaltered in A, but in A they all abut at the sole maximal ideal. So if the two maximal ideals in R are of different heights (both superior to one), the ring A will not be catenary. Remains the question of Noetherianess of A (as we shall see, it is easy to find appropriate semi local rings).

Of course one must start out with a Noetherian R, subrings however do not always inherit Noetherianess, but R is a finite module over A. Once this is established, it ensues from a general theorem independently found by Paul Eakin and Masayoshi Nagata that A is Noetherian. This result is not covered by this course, so we shall give a simple *ad hoc* proof.

LEMMA 14.12 *In the setting of this section, R is a finitely generated A module, and when R is Noetherian, A will be Noetherian.*

PROOF: Let us prove that R is a finite A-module: any two elements e_1 and e_2 in R with $e_i \equiv \delta_{ij} \mod \mathfrak{m}_j$ generate R over A together with the unity; indeed, their classes in $R/\mathfrak{m}_1 \times R/\mathfrak{m}_2 \simeq R/\mathfrak{n} \simeq k \oplus k$ form a basis over $k \simeq A/\mathfrak{n}$, hence given $a \in R$, there are elements $a_i \in A$ so that $a - a_1 e_1 - a_2 e_2 \in \mathfrak{n}$.

Let now $\mathfrak{p}$ be a prime ideal in R, it by Cohen's Criterion it suffices to prove that $\mathfrak{p} \cap A$ is finitely generated; indeed, every prime ideal in A is shaped like that by the Lying–Over Theorem.

There are two cases: firstly, if $\mathfrak{a}$ is any ideal in R such that $\mathfrak{a} \subseteq \mathfrak{m}_1 \cap \mathfrak{m}_2$, then $\mathfrak{a}$ is entirely contained in A. By assumption R is Noetherian so that $\mathfrak{a}$ is finitely generated over R, and since R is finitely generated over A, the ideal $\mathfrak{a}$ is finitely generated over A as well.

Secondly, assume that $\mathfrak{p} \subseteq \mathfrak{m}_1$ but not contained in $\mathfrak{m}_2$ (by symmetry this case will suffice). There is an inclusion

$$\mathfrak{p} \cap A/\mathfrak{p} \cap \mathfrak{m}_2 \subseteq \mathfrak{p}/\mathfrak{p} \cap \mathfrak{m}_2,$$

and $\mathfrak{p}/\mathfrak{p} \cap \mathfrak{m}_2$ is a finitely generated as a module over $R/\mathfrak{n} \simeq k \times k$ and therefore also as a vector space over $A/\mathfrak{n}$; hence $\mathfrak{p} \cap A/\mathfrak{p} \cap \mathfrak{m}_2$ is finitely generated over A. Now $\mathfrak{p} \cap \mathfrak{m}_2$ is contained in $\mathfrak{m}_1 \cap \mathfrak{m}_2$, which we already know is a finitely generated ideal in A; so we are through. ❑

(14.13) With these two lemmas up in the sleeve, we summarize and give the example. It is easy to find semi-local rings of the sought kind: the simplest example is the localization $R = S^{-1}k[x_1, x_2, y_1, y_2, y_3]$ where the multiplicative system S is the complement of the union $(x_1, x_2) \cup (y_1, y_1, y_3)$ of the two prime ideals (x_1, x_2) and (y_1, y_1, y_3). This ring is semi-local with the two maximal ideals $\mathfrak{m}_1 = (x_1, x_2)R$ and $\mathfrak{m}_2 = (y_1, y_2, y_3)R$.

One easily verifies that $\operatorname{ht} \mathfrak{m}_1 = 2$ and $\operatorname{ht} \mathfrak{m}_2 = 3$, and that the residue fields are the function fields respectively given as $R/\mathfrak{m}_1 = \mathbb{C}(y_1, y_2, y_3)$ and $R/\mathfrak{m}_2 = \mathbb{C}(x_1, x_2)$. These residue fields are of course not isomorphic as $\mathbb{C}$-algebras, but luckily they are as fields! This is rooted in the fact that $\mathbb{C}$ is of infinite transcendence degree over the field $\overline{\mathbb{Q}}$ of algebraic integers. So in the end, all function fields over $\mathbb{C}$ are isomorphic to $\mathbb{C}$. Choosing a fixed field k in the isomorphism class as well as isomorphisms of $\mathbb{C}(x_1, x_2)$ and $\mathbb{C}(y_1, y_2, y_3)$ with k, our construction results in a domain A which will be Noetherian but not catenary. The isomorphisms between function fields and $\mathbb{C}$ are really weird, they mix up infinitely many complex numbers (so for instance, they might interchange e and π), and the resulting ring we construct will be quirky as well.

14.4 *A Noetherian bubble space*

Many a young geometer has struggled with the subtleties in the examples of Nagata and Akizuki and finally written them off as belonging to the deepest darkness in the kingdom of non-geometry. But indeed, many examples are as close to being of finite type over fields as they can be without being, and have a pronounced geometric flavour.

*A more official and serious name is "Riemann–Zariski" spaces.

The term* "bubble spaces" is used by some mathematicians about some horrendously large spaces constructed for studying sequences of blow ups of projective surfaces. Being the limit of (infinitely many) affine blow-ups of the simplest sort our space is just a tiny string of bubbles rising to the surface, but it lies at the heart of several of the Japanese examples, it is geometric and gives an easy instance of the hypothesis of separability in the First Finiteness Theorem (Theorem 12.45 on page 335) being needed. In fact, the amazing thing is that these infinite constructions may yield Noetherian rings—which also is of independent interest.

In view of the First Finiteness Theorem, an integral closure that is not finite must involve Frobenius maps in some way, and given the infinite nature of our construct, it is not surprising that the Frobenius map is not finite.

(14.14) The example A can be constructed as an algebra over any field k. It is a Noetherian normal domain of dimension two all whose local rings are regular and it is even a Jacobson ring, but it has a maximal principal ideal $\mathfrak{m} = (x)A$. The corresponding point in Spec A is given by one equation, and there is no curve passing by it! The Krull-dimension $\dim A_{\mathfrak{m}}$ does not equal the height $\operatorname{ht}\mathfrak{m}$, and there is strict inequality in the dimension formula in Proposition 11.4 on page 290.

(14.15) We shall be working with a sequence of polynomial rings $A_i = k[x, z_i]$ indexed by non-negative integers $\mathbb{N}_0$. The elements z_i will all be members of a rational function field $k(x, z)$ in two variables, where for the moment k is any field, and they will depend on a given sequence $\{a_i\}_{i\in\mathbb{N}_0}$ of elements from the ground field k. The definition is recursive: when $i = 0$, we set $z_0 = z$, and when $i > 0$, the element z_{i+1} is given by the relation

$$z_i = x(z_{i+1} + a_i). \tag{14.4}$$

Thus each ring $A_i = k[x, z_i]$ is identified with the subring $k[x, x(z_{i+1} + a_i))]$ of the next ring $A_{i+1} = k[x, z_{i+1}]$, and in this way the A_i's form an ascending chain of subrings of $k(x, z)$:

$$k[x, z_0] \subset k[x, z_1] \subset \ldots \subset k[x, z_i] \subset k[x, z_{i+1}] \subset \ldots \subset k(x, z). \tag{14.5}$$

Note that $z_{i+1} = x^{-1}z_i - a_i$ for each $i \in \mathbb{N}_0$, so for every r it holds that $k[x, x^{-1}, z_{i+r}] = k[x, x^{-1}, z_i]$; indeed, it readily follows by induction that $z_{i+r} \in k[x, x^{-1}, z_i]$ for all $r \in \mathbb{N}_0$.

(14.16) To describe the geometric counterpart of this picture, we introduce the notation $\mathbb{A}^2_i = \operatorname{Spec} A_i$. These spectra are of course all isomorphic, but they enter the process as different members of the chain

$$\ldots\ldots \longrightarrow \mathbb{A}^2_{i+1} \longrightarrow \mathbb{A}^2_i \longrightarrow \mathbb{A}^2_{i-1} \longrightarrow \ldots \longrightarrow \mathbb{A}^2_1 \longrightarrow \mathbb{A}^2_0.$$

All the maps are so-called "affine blow-ups" which we met in Example 11.13 on page 301. Heuristically, over the complex numbers each $\mathbb{A}^2_i$ may be thought of as a $\mathbb{C}^2$ equipped with coordinates (x, z_i), and the maps in the chain send $(x, z_i) \to (x, x(z_i + a_i))$. Thus the z_i-axis collapses to the origin (hence the name "blow up", as they mostly are seen from the target), but due to the parameters a_i the inverse image of the x-axis is not the x-axis (as it is in Example 11.13), but a translate, namely the line $z_i = -a_i$.

(14.17) The star of the game will be the union $A = k[x, z_0, z_1, \ldots]$ of all these subrings. It does not look very Noetherian, and certainly is not in many instances. If all the a_i's vanish, for instance, it is not Noetherian (as you were asked to show in Exercise 9.18 on page 245), but amazingly enough it will be when the power series $\zeta = \sum_{i\geqslant 1} a_i x^i$ is transcendental over k; that is, ζ is not algebraic over the polynomial ring $k[x]$. This condition may also be expressed as all the "tails" $\zeta_r = \sum_{i\geqslant 1} a_{i+r} x^i$ with $r \geqslant 0$ being algebraically independent:

LEMMA 14.18 *If ζ is transcendental over k, all the tails will be algebraically independent.*

PROOF: It holds for any $r \geqslant 0$ that $\zeta_r = (\zeta - p_r)x^{-r}$ where p_r is the polynomial $p_r = \sum_{i\leqslant r} a_i x_i$. Hence, if $F(x, \zeta_0, \ldots, \zeta_r)$ is a polynomial, for some sufficiently large natural number N one has $x^N F(x, \zeta_0, \ldots, \zeta_r) = G(x, \zeta)$ where G is a polynomial. Therefore an algebraic relation between the ζ_i gives one for ζ. ❑

The condition that ζ be transcendental enters the scene by way of a ring homomorphism $A \to k[\![x]\!]$. The power series $\zeta_i = \sum_{j\geqslant 1} a_{j+i} x^j$ fulfills the same relations as the z_i's do; indeed, one easily verifies that

$$x(\zeta_{i+1} + a_{i+1}) = xa_{i+1} + \sum_{j\geqslant 1} a_{i+1+j} x^{j+1} = \zeta_i,$$

and so $z_i \mapsto \zeta_i$ is a legitimate definition of a homomorphism. The lemma tells us that it will be injective if (and only if) the power series ζ is transcendental.

(14.19) Our first observation when starting to investigate the algebra A, is that

$$(x)A \cap k[x, z_i] = (x, z_i).$$

The element z_i belongs to $(x)A$ because $z_i = x(z_{i+1} + a_{i+1})$, which shows the inclusion $(x, z_i) \subseteq (x)A \cap k[x, z_i]$, and this suffices since (x, z_i) is a maximal ideal. A consequence

is that the principal ideal $(x)A$ is a maximal ideal: indeed, if $\mathfrak{a}$ was an ideal strictly larger than $(x)A$ any element in $f \in \mathfrak{a} \setminus (x)A$ would lie in some $k[x, z_i]$, but as $(x)A \cap k[x, z_i]$ is maximal, this is impossible.

(14.20) Here comes the salient point of the construction:

PROPOSITION 14.21 *If the power series $\sum_{i\geqslant 1} a_i x^i$ is transcendental over k, then A is Noetherian.*

PROOF: We intend to show that any prime ideal $\mathfrak{p}$ is finitely generated. Cohen's criterion (Proposition 9.35 on page 241) will then imply the lemma. There are two cases to consider according to wether $\mathfrak{p} \subseteq (x)A$ or $\mathfrak{p} \nsubseteq (x)A$.

We begin with the latter which is done by a standard argument. Notice, that since $(x)A$ is a maximal ideal, it ensues that $x \notin \mathfrak{p}$. For any i it holds that $A_x = k[x, x^{-1}, z_i]$, so A_x is Noetherian, and we may find $f_1, \ldots, f_r$ in $\mathfrak{p}$ such that $(f_1, \ldots, f_r)A_x = \mathfrak{p}A_x$. Furthermore, there is an element $g \in \mathfrak{p}$ not in $(x)A$. We contend that $\mathfrak{p} = (g, f_1, \ldots, f_r)$: let $\mathfrak{m}$ be any maximal ideal in A. When $x \notin \mathfrak{m}$, the f_i's will generate $\mathfrak{p}A_\mathfrak{m}$, and if $\mathfrak{m} = (x)A$, the element g will. Since being surjective is a local property, it follows that $\mathfrak{p} = (g, f_1, \ldots, f_r)$.

The more serious part is the case when $\mathfrak{p} \subset (x)$, which requires that ζ is transcendental. In fact, we contend that this case does not materializes unless $\mathfrak{p} = 0$. Now, if $\mathfrak{p}$ is non-zero, for each i the intersection $\mathfrak{p} \cap k[x, z_i]$ will a prime ideal properly contained in (x, z_i), and it is therefore a principal ideal (Proposition 3.30 on page 77), say generated by an irreducible polynomial $f_i(x, z_i)$, different from x since $x \notin \mathfrak{p}$. Substituting for z_i we find

$$f_i(x, z_i) = f_i(x, x(z_{i+1} + a_i)) = xg(x, z_{i+1}) = h(x, z_{i+1}) f_{i+1}(x, z_{i+1}),$$

for polynomials g and h in $k[x, z_{i+1}]$. Since f_{i+1} is irreducible and different from x, it ensues that x divides h and consequently that $f_i \in (xf_{i+1})$. By induction we infer that $f_i \in (x^r f_{i+r})$ for all r so that $f_i \in \bigcap_{r\geqslant 1}(x^r)A$. The salient point is now that $A \subset k[\![x]\!]$, and since the ring $k[\![x]\!]$ of power series is Noetherian, we can appeal to Krull's Intersection Theorem and conclude that $\bigcap_{r\geqslant 1}(x^r)A \subseteq \bigcap_{r\geqslant 1}(x^r)k[\![x]\!] = 0$. ❏

Exercises

(14.2) Show that the reverse implication in Proposition 14.21 holds: If ζ is not transcendental, then A is not Noetherian. HINT: Show that kernel of $A \to k[\![x]\!]$ is contained in (x^i) for all i.

⁂ (14.3) Let k be a field of characteristic zero and let $d \geqslant 2$ be an integer. Show that the power series $f(z) = \sum_{i\geqslant 0} z^{d^i}$ is transcendental. This is an example of a so-called *lacunary series*; most of the terms are zero, and the gaps between the non-zero terms tend (rapidly) to infinity. It persists being transcendental in positive characteristic p

(for most p when d is fixed), but the argument is involved. HINT: Use the relation $f(z^d) = f(z) - z$.

(14.4) The notation is as above. Let P denote the closed point in Spec A corresponding to the maximal ideal $(x)A$.

a) Show that the distinguished open subset $D(x)$ of Spec A is isomorphic to Spec A_x and therefore to $\mathbb{A}^2_k \backslash V(x)$.

b) Show that the Zariski closed sets in Spec A are those of the form Z or $Z \cup \{P\}$ when Z runs though the closed sets $Z \subset D(x)$; so $D(x)$ is not closed, but of course, the entire Spec $A = D(x) \cup \{P\}$ is.

★

Integral closures in characteristic p

We have come to the example of a domain with an integral closure not being a finite module; to be precise: we shall construct a Noetherian domain A, in fact a normal one, whose integral closure in a finite extension of the fraction field is not a finite module over A. Naturally, the extension must be inseparable, and we shall enter the realm of rings of positive characteristic. Finally, a slight twist of the construction gives a Noetherian domain whose integral closure in the fraction field is not finite.

(14.22) The field k will be of characteristic $p > 0$, and in the construction we need only a sequence $\{a_i\}$ of elements from k so that the power series associated to $\{a_i\}$ and $\{a_i^p\}$ both are transcendental over k. One way to achieve this is to assume that k is perfect so that the Frobenius map* on k is bijective; if $\sum_i a_i x^i$ is transcendental, $\sum_i a_i^p x^i$ will then be as well. There are amazingly many amazing transcendental power series over the finite field $\mathbb{F}_p$; for instance $\sum_i \chi_{\mathbb{P}}(i) x^i$ where $\chi_{\mathbb{P}}$ is the characteristic function of the set of primes*, and another good example is $\sum_i \mu(i) x^i$ where μ is the Moebius function (see *e.g.* **[?]** for proofs and many other examples). So, in fact, there is no restriction on the field since these series coincide with their p^{th}-power series their coefficients being either 0 or ± 1.

*In plain language the p-th power map $x \mapsto x^p$ which in characteristic p is a ring homomorphism.

*Geometrically: you start blowing up and each time the number of blowing ups executed is prime you move centre a meter; then you get a Noetherian limit-ring!

(14.23) The tactics of the construction are to introduce the Frobenius map in the above "bubble-space-setting", and the salient point is that this Frobenius map is not a finite map.

Consider for each i the subring $B_i = k[x^p, z_i^p]$ of $A_i = k[x, z_i]$. The gist of the example is that the inclusion map $A_i \subseteq A_{i+1}$ takes B_i into B_{i+1}; indeed, since we are in characteristic p it holds true that

$$z_i^p = (x(z_{i+1} + a_i))^p = x^p(z_{i+1}^p + a_i^p).$$

An even more holds true, the resulting chain of the B_i's is of the same shape as the chain (14.5) save being built with the sequence $\{a_i^p\}$ in stead of $\{a_i\}$ and the variables x^p

and z_i^p. The situation is summarized by the diagram:

$$\begin{array}{ccccccc} \ldots & \longrightarrow & k[x^p, z_i^p] & \longrightarrow & k[x^p, z_{i+1}^p] & \longrightarrow & \ldots \\ & & \downarrow & & \downarrow & & \\ \ldots & \longrightarrow & k[x, z_i] & \longrightarrow & k[x, z_{i+1}] & \longrightarrow & \ldots \end{array}$$

Note that $\zeta' = \sum_i a_i^p x^{ip}$ is transcendental when $\zeta = \sum_i a_i x^i$ is. Indeed, assume that $G(x^p, \zeta') = 0$. This implies that $G(x^p, \zeta^p) = 0$ and since k is perfect every coefficient of G has a p^{th} root and there is polynomial H such that $H(x, \zeta)^p = G(x^p, \zeta^p)$. The latter vanishes, and we conclude that $H(x, \zeta) = 0$. The games played in the upper chain is thus exactly the same as the one played in the bottom one, just with slightly altered players; in particular, $B = k[x^p, z_0^p, \ldots]$ is normal and most importantly in our context, it will be Noetherian.

In the end, the unions $B = \bigcup_i k[x^p, z_i^p] = k[x^p, z_0^p, z_i^p, \ldots]$ and $A = \bigcup_i k[x, z_i] = k[x, z_0, z_1, \ldots]$ are normal Noetherian rings and $B \subseteq A$. The fraction field of B equals $k(x^p, z^p)$ and that of A is $k(x, z)$. Since A is normal, it is contained in the integral closure $\overline{B}$ of B in the field $k(x, z)$ in fact, they coincide: each extension $B_i = k[x^p, z_i^p] \subseteq k[x, z_i] = A_i$ is integral (A_i is generated by the monomials $x^\mu z_\nu^j$ with $0 \leqslant \mu < p$ and $0 < \nu < p$ and so is finite over A_i) and every element in A belongs to some A_i.

PROPOSITION 14.24 *The ring B is Noetherian with quotient field $k(x^p, z^p)$. The integral closure $\overline{B} = A$ in $k(x, z)$ is not a finite module over B.*

PROOF: We assume for simplicity that $p = 2$. Aiming for a contradiction, let us assume that A is finitely generated over B. It will then be a Noetherian B-module since B is Noetherian, and the ascending chain $A_n = B[x, z_0, \ldots, z_n]$ of B-modules will be stationary. Hence for some $n >> 0$ it holds that $z_{n+1} \in A_n$. Now, A_n is generated as a ring over $k[x^2, z_n^2]$. by the polynomials 1, x, z_n and xz_n. The coefficients in a relation expressing z_{n+1} as a combination of these generators, involve only finitely many of the z_i's, and they will all lie in $k[x^2, z_r^2]$ for $r >> 0$; hence we may write

$$z_{n+1} = p_1(x^2, z_r^2) + xp_2(x^2, z_r^2) + z_n p_3(x^2, z_r^2) + xz_n p_4(x^2, z_r^2)$$

where the coefficients belong to $k[x^2, z_r^2]$. Now, $z_{n+1} = x^{r-n-1} z_r + s(x)$ and $z_n = x^{r-n} z_r + t(x)$ where $s(x)$ and $t(x)$ are polynomials. Merging p_1 and xp_2 into f, p_3 and xp_4 into g and incorporating s and t, we arrive at an identity

$$x^{r-n-1} z_r = f(x, z_r^2) + x^{r-n} z_r g(x, z_r^2)$$

between polynomials in $k[x, z_r]$. But such a relation is impossible: since the powers of z_r appearing in $f(x, z_r^2)$ all are even and those in all the other terms are odd, we infer

that $f = 0$, thus arriving at the identity:

$$x^{r-n-1}z_r = x^{r-n}z_r g(x, z_r^2).$$

Cancelling $x^{r-n-1}z_r$ gives $1 = xz_r g(x, z_r^2)$, which is absurd. ❑

(14.25) If one insists on having an example where the rings involved are DVR's just observe that $(x, z) \cap B = (x^p, z_p)$ so denoting this maximal ideal by $\mathfrak{m}$, we see that the discrete valuation ring $B_{\mathfrak{m}}$ does not have a finite integral closure in the field $k(x, z)$.

A ring with non-finite normalization

In the previous paragraphs we constructed rings B whose integral closure A in a finite extension of the fraction field is not a finite module. Almost for free, they give rise to rings with a normalization that is not a finitely generated module; just take any ring C lying properly between B and A which is a finite B module and which have the same fraction field as A (of course, it is not totally for free that such creatures are about). Indeed, the normalization of C equals A, and A cannot be finite over C as that would entail that A was finite over B. Moreover, C is Noetherian being a finite module over the Noetherian ring B.

(14.26) To construct the ring C as described above, we elaborating the example slightly and adjoin x and z_0 to B. In that way we obtain the ring $C = k[x, z_0, z_1^p, \ldots]$: the fraction field is $k(x, z)$ and it is a finitely generated module over B—it is generated by the monomials $x^\mu z_0^\nu$ with $0 \leqslant \mu < p$ and $0 \leqslant \nu < p$.

(14.27) Warning: This paragraph presupposes some knowledge of the completion of a ring. Krull showed that if A is a local Noetherian domain withdim $A = 1$, then the normalization $\widetilde{A}$ is finite over A if and only if the completion $\widehat{A}$ is without nilpotent elements; and there are generalisation to other dimension of the same flavour although not as clean cut. As an illustration of Krull's result, let us point to a nilpotent element in $\widehat{C}$, the completion of C with respect to the maximal ideal $\mathfrak{m} = (x, z_0)$. Not to create unnecessary confusion, and to underline that z_i^p is not a p^{th}-power in C when $i > 0$, we introduce the notation $w_i = z_i^p$. In C the relation

$$z_0^p = x^p(w_1 + a_0^p)$$

holds. Note that if w_1 were a p^{th}-power—which it is not—say $w_1 = \omega^p$, we would have a nilpotent element since then $(z_0 - x(\omega + a_1))^p = 0$. However in the completion $\widehat{C}$ the element w_1 have a p^{th}-root. In view of the relations (14.23) a straightforward induction yields the following equality which is valid for all $r \geqslant 1$:

$$w_1 = x^{pr}w_{r+1} + \sum_{j=1}^{r} x^{jp}a_j^p.$$

Now $x^{pr}w_{i+r} \in \mathfrak{m}^{pr}$, so the right hand side converges to $\sum_{j=1}^{\infty} x^{jp}a_j^p$, and allows us to exhibit w_1 as a p^{th}-power:

$$w_1 = (\sum x^j a_j)^p.$$

Lecture 15

Dedekind rings

The work of Richard Dedekind on ideals was the beginning of abstract algebra. And the rings named after him are central both in algebraic number theory and in algebraic geometry. In number theory they are the foremost players being the rings of integers in algebraic number fields, and in algebraic geometry they appear as the coordinate rings of affine regular curves.

In this chapter we shall first treat the local variant of the Dedekind rings, which are the so called *discrete valuation rings*, and as the name indicates, they are members of the larger family of *valuation rings*. Valuation rings are closely related to certain functions on fields with values in ordered groups (which will be the integers $\mathbb{Z}$ for the discrete ones) called *valuations*. The discrete valuation rings are omnipresent in both algebraic geometry and algebraic number theory, but the others are seldom met in contemporary algebraic geometry, and we relegated them to a rudimentary discussion in an appendix. They are however important in other approach to algebraic geometry than Grothendieck's schemey way; the hub in Zariski's development of algebraic geometry is the valuation rings in function fields and the so-called *places*.

The prominent place discrete valuation rings have in mathematics relies on they being the local one-dimensional integrally closed Noetherian rings. One meets them, for instance, as the local rings of rational functions regular at a point on any non-singular curve (*e.g.* an open subsets of a compact Riemann surface when the ground field is $\mathbb{C}$), with the valuation being the order of the function at the point. In the study of varieties of higher dimension, they play a central role in the description of so-called divisors, which are gadgets built from codimension one subvarieties (note that points on curves are of codimension one), and in number theory they are omnipresent; rings of integers in algebraic number fields have local rings all being discrete valuation rings.

A common technique in number theory, which also from time to time is seen used in algebraic geometry, is to pass back and forth between characteristic zero and characteristic p, and this is most often done by working with algebras (or schemes) over a discrete valuation ring whose residue field is of characteristic p while the fraction

field ha characteristic zero.

15.1 *Discrete valuation rings*

There are two traditional examples of valuations found in elementary mathematics. One is the order of vanishing, $\operatorname{ord}_z(f)$, of a holomorphic function f at a point z in the complex plane (or a rational function to make it more elementary), and the other is the so-called *p-adic valuation* $v_p(x)$ defined on the integers. The latter is the exponent to which a fixed prime p occurs in the prime factorization of the integer x; in other words, one has $x = p^{v_p(x)}y$ where the integer y is relatively prime to p. Both functions are easily extended to fractions—that is, respectively to function meromorphic at the point and to rational numbers—simply by taking the difference of the values of the enumerator and the denominator.

The notion of *discrete valuations*, which we are about to introduce, is built on features these two examples have in common; how they behave with respect to products and sums. The order of vanishing, for instance, obey the two well-known rules

$$\operatorname{ord}_z(f \cdot g) = \operatorname{ord}_z(f) + \operatorname{ord}_z(g) \qquad \operatorname{ord}_z(f+g) \geqslant \min(\operatorname{ord}_z(f), \operatorname{ord}_z(g)),$$

and the p-valuation shows exactly the same behaviour in that

$$v_p(xy) = v_p(x) + v_p y \qquad v_p(x+y) \geqslant \min(v_p(x), v_p(y)).$$

With every discrete valuation is associated a *discrete valuation ring*. In the function case it will be the ring of functions holomorphic at z and in the p-adic case the ring $\mathbb{Z}_{(p)}$ of rational numbers which when written in lowest terms, do not have p as factor in the denominator. We recognize this ring as the ring of integers $\mathbb{Z}$ localized at the prime ideal (p).

Discrete valuations

Discrete valuations (diskrete valuasjoner)

(15.1) In what follows we shall be working with a field K. A *discrete valuation* on K is a non-zero function $v\colon K^* \to \mathbb{Z}$ that obeys the two following rules:

- $v(xy) = v(x) + v(y)$;
- $v(x+y) \geqslant \min(v(x), v(y))$,

where $x, y \in K$, and where we must assume that $x + y \neq 0$. The fist requirement may be rephrased as v being a group homomorphism from the multiplicative group K^* to the additive group $\mathbb{Z}$. Moreover, since $1^2 = 1$, it immediately follows that $v(1) = 2v(1)$, so that $v(1) = 0$, and consequently it holds that $v(x^{-1}) = -v(x)$.

If v is surjective, the valuation said to be *normalized*. The image $v(K^*)$ is a subgroup of $\mathbb{Z}$ and has, as every subgroup of $\mathbb{Z}$, a unique positive generator e. Thus all the values $v(x)$ have e as factor, and $e^{-1}v$ will be a normalized valuation which is canonically associated with v.

Normalized valuations (normaliserte valuasjoner)

It is convenient to introduce a symbol ∞ and extend the addition of integers to $\mathbb{Z} \cup \{\infty\}$ by the rules $\alpha + \infty = \infty + \alpha = \infty$, and of course, we also impose the inequality $\alpha \leqslant \infty$ for all $\alpha \in \mathbb{Z}$. Extending v to K by putting $v(0) = \infty$ one obtains a map $v\colon K \to \mathbb{Z} \cup \{\infty\}$ still abiding by the two rules above, but without the limitation of only being defined for non-zero elements.

The two first statement of the following lemma are almost for free, and the third asserts that when $v(x)$ and $v(y)$ are different, the inequality in the second rule in fact is an equality:

LEMMA 15.2 *Let v be a valuation on the field K and x and y two non-zero elements from K. Then*

i) $v(x^n) = nv(x)$;
ii) $v(-x) = v(x)$;
iii) If $v(x) > v(y)$, *it holds that* $v(x+y) = v(y)$.

PROOF: All is clear when $x = 0$, so we may assume that $x \neq 0$, and then $y \neq 0$ as $v(x) > v(y)$. The first assertion follows just from v being group homomorphism and the second from the equality $(-1)^2 = 1$. As to the third, one has $v(x+y) = v(y(xy^{-1}+1)) = v(y) + v(xy^{-1}+1)$, so it suffices to show that $v(1-t) = 0$ whenever $v(t) > 0$ (indeed, set $t = -xy^{-1}$ and note that $v(t) = v(x) - v(y) > 0$), and since $v(1-t) \geqslant \min(0, v(t))$, showing that $v(1-t) \leqslant 0$ will be enough. To that end, consider the equality

$$(1-t)^{-1} = 1 + t + \ldots + t^{n-1} + t^n(1-t)^{-1}.$$

Since $v(t^i) = iv(t) > 0$ for $i > 1$, it yields

$$-v(1-t) = v(1-t)^{-1} \geqslant \min(v(1), v(t^n(1-t)^n)) = \min(0, nv(t) - v(1-t)) = 0,$$

when n is sufficiently big, and that's it. ❑

⁂ EXERCISE 15.1 Let $v_0\colon A\backslash\{0\} \to \mathbb{Z}$ be a function on a domain A satisfying the two axioms for a valuation. Prove that there is a unique discrete valuation v on the fraction field K extending v_0. HINT: Show that the function $v(ab^{-1}) = v(a) - v(b)$ is well defined and complies with the rules. ★

⁂ EXERCISE 15.2 Let p be an irreducible element in the UFD A. Show there is a unique discrete valuation v_p on K with $v_p(p) = 1$ taking non-negative values on A. ★

Valuation rings

The current section is about discrete valuations and discrete valuation rings, but a few of their properties have proofs valid for general valuation rings, so for a while, before coming back to the discrete ones, we shall work with general valuation rings.

Valuation rings (valuasjons ring)

(15.3) A domain A with field of fractions K is called a *valuation ring* if for each element from K either the element itself or its inverse belongs to A; that is, for each $x \in K$ either $x \in A$ or $x^{-1} \in A$. Note the field K it self will be a valuation ring. Equivalently, one may require that for any two elements x and y from A either $x|y$ or $y|x$; indeed, $x|y$ is equivalent to $xy^{-1} \in A$ and $y|x$ to $yx^{-1} \in A$. Yet another variant is to ask that the lattice of principal ideals in A be totally ordered (the condition $x|y$ translates into the inclusion $(y) \subseteq (x)$), and even more holds true:

PROPOSITION 15.4 *A domain A is a valuation ring if and only if the lattice $\mathcal{I}(A)$ of ideals in A is totally ordered.*

PROOF: Suppose first that A is a valuation ring and assume that two ideals satisfy $\mathfrak{a} \nsubseteq \mathfrak{b}$. We are to show that $\mathfrak{b} \subseteq \mathfrak{a}$. This is trivial when $\mathfrak{b} = (0)$, and at the outset $\mathfrak{a} \neq (0)$, so we may assume that both are non-zero. To proceed, pick an element $x \in \mathfrak{a}$ such that $x \notin \mathfrak{b}$, and let $y \in \mathfrak{b}$ be any non-zero element. Since $(x) \nsubseteq (y)$, it holds that $(y) \subseteq (x)$, and consequently $y \in \mathfrak{a}$.

For the converse implication, let $x \in K$ be an element in the fraction field of A and write $x = yz^{-1}$ with y and z from A. By hypothesis $\mathcal{I}(A)$ is totally ordered, and thus either $(z) \subseteq (y)$, in which case $x^{-1} = zy^{-1}$ belongs to A, or $(y) \subseteq (z)$, and $x = yz^{-1}$ lies in A. ❏

PROPOSITION 15.5 *Valuation rings are local rings integrally closed in their fraction fields.*

PROOF: That a valuation ring merely has one maximal ideal, ensues from the lattice of ideals being totally ordered; indeed, different maximal ideals are not comparable.

Let us then prove that a valuation ring A is integrally closed in its fraction field K. So assume that $x \in K$ is a non-zero element not lying in A, but satisfying an integral dependence relation

$$x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n = 0,$$

where the a_i's are elements from A. Since A is a valuation ring, it holds that $x^{-1} \in A$, and in fact, x^{-1} even lies in the maximal ideal $\mathfrak{m}$ of A since it is not a unit as $x \notin A$. A simple manipulation gives

$$1 = -\big(a_1 + \ldots + a_{n-1}x^{-(n-2)} + a_n x^{-(n-1)}\big)x^{-1},$$

from which ensues the absurdity that $1 \in \mathfrak{m}$. ❏

(15.6) Most valuation rings turn out not to be Noetherian—the exception being the discrete valuation rings—but they are what is commonly known as *Bézout rings*, which

we also met earlier (in Theorem 8.30 on page 215). These resemble the PID's in that every *finitely generated* ideal is principal.

Bézout rings (Bézout-ringer)

PROPOSITION 15.7 *In a valuation ring each finitely generated ideal is principal.*

PROOF: By induction on the number of generators it suffices to prove the proposition for ideals with two generators. So assume that $\mathfrak{a} = (x, y)$. Now, since A is a valuation ring, either $x|y$ and $\mathfrak{a} = (y)$, or $y|x$ in which case $\mathfrak{a} = (x)$. ❑

(15.8) Valuation rings are among the larger rings in their function fields. The class of valuation rings contained in a given field K is closed from above in the sense that any proper subring of K containing a valuation ring is itself a valuation ring. It ensues that localizations of a valuation ring is a valuation ring, and in fact, the only proper subrings of K larger than a valuation ring are the localizations at its prime ideals. In particular, valuation rings of Krull dimension one, *i. e.* those having just one non-zero prime ideal, will be maximal proper subrings of K.

PROPOSITION 15.9 *Let A be a valuation ring with fraction field K and let B be an overring of A different from K; that is, $A \subseteq B \subsetneq K$. Then the following holds true:*

i) The ring B is a valuation ring;
ii) The maximal ideal of B satisfies $\mathfrak{m} \subseteq A$, and $B = A_{\mathfrak{m}}$.

PROOF: That B is a valuation ring comes for free since if $x \notin B$, *a fortiori* $x \notin A$, and hence $x^{-1} \in A$ since A is a valuation ring. The maximal ideal $\mathfrak{m}$ is described as $\mathfrak{m} = \{ x \in B \mid x^{-1} \notin B \}$, so if $x \in \mathfrak{m}$, we may conclude that $x^{-1} \notin A$, and therefore $x \in A$ since A is valuation ring. To see that $B = A_{\mathfrak{m}}$, note that if $x \in A$, but $x \notin \mathfrak{m}$, it holds that $x^{-1} \in B$. ❑

Exercises

(15.3) Let A be a domain with fraction field K. Prove that A is a valuation ring if and only if the set of fractional ideals is totally ordered under inclusion.

⁂ (15.4) *DVR's are maximal subrings.* Let $A \subseteq B$ be two proper subrings of their common fraction field K. Prove that if A is a DVR, then $A = B$; that is, DVR's are maximal proper subrings of their fraction fields.

⁂ (15.5) Show that the intersection of any collection of valuation rings in a field K is a valuation ring. Show that the union of an ascending chain of valuation rings in K is a valuation ring.

⁂ (15.6) *Chevalley's lemma.* Prove the so-called Chevalley's lemma: Assume that A is a domain and let $\mathfrak{a}$ be a proper non-zero ideal in A. Let $x \in K$ be an element not in A. Then either $\mathfrak{a}A[x]$ or $\mathfrak{a}A[x^{-1}]$ is a proper ideal respectively in $A[x]$ and $A[x^{-1}]$.

HINT: Assume not, and express 1 both as a polynomial in x and as one in x^{-1}. Use polynomials of minimal degree, and deduce a contradiction.

✲ (15.7) *Existence of places.* Let A be a domain and $\mathfrak{p}$ a prime ideal in A. Prove that there is a valuation ring V in the fraction field K of A with maximal ideal $\mathfrak{m}_V$ such that $A \subseteq V$ and $\mathfrak{m}_V \cap A = \mathfrak{p}$ (such a valuation ring V is traditionally called *a place centred* at $\mathfrak{p}$).
HINT: Consider the set of local rings B containing A whose maximal ideal $\mathfrak{m}_B$ satisfies $\mathfrak{m}_B \cap A = \mathfrak{p}$. Use Zorn's lemma and the lemma of Chevalley from the previous exercise (Exercise 15.6).

★

Discrete valuation rings

With any discrete valuation v on a field K is associated a valuation ring. The underlying set consists of the elements where v assumes non-negative values; that is, the set

$$A = \{\, x \in K \mid v(x) \geqslant 0 \,\}.$$

That A really is a ring, ensues from the two axioms for valuations. Since $v(xy) = v(x) + v(y)$, it holds that $v(xy)$ is non-negative whenever $v(x)$ and $v(y)$ are, so A is closed under multiplication. And $v(x+y)$ is non-negative as well, being larger that both $v(x)$ and $v(y)$, which shows that A is additively closed. Furthermore, since $v(x^{-1}) = -v(x)$, either $x \in A$ or $x^{-1} \in A$; hence A is a valuation ring.

Using that $v(x^{-1}) = -v(x)$ once more, we see that the group A^* of units in A precisely is formed by the elements with $v(x) = 0$; indeed, both $v(x) \geqslant 0$ and $-v(x) \geqslant 0$ can merely be true for $v(x) = 0$. We infer that the maximal ideal in A consists of the elements in K where v is positive. Thus we have proven:

PROPOSITION 15.10 *Let v be a discrete valuation on K and let A be the corresponding valuation ring. Then the group of units in A is given as $A^* = \{\, x \in K \mid v(x) = 0 \,\}$, and the maximal ideal as $\mathfrak{m} = \{\, x \in K \mid v(x) > 0 \,\}$.*

Discrete valuation rings (diskrete valuasjonsringer)

Domains that arise as just described—*i. e.* as subsets of fields where a discrete valuation assumes non-negative values—are the famous *discrete valuation rings*. The initialism DVR is widely used.

(15.11) The descriptions of A, A^* and $\mathfrak{m}$ in the proposition are insensitive to changes of v by positive factors. In particular, they do not change when v is replaced by its normalization $e^{-1}v$ (the natural number e being the positive generator of $v(K^*)$), and one may as well restrict one's attention to the normalized valuations.

Uniformizing parameters (uniformiserende parametre)

(15.12) Ring elements t such that $v(t) = e$ (or $v(t) = 1$ in case v is normalized) are called *uniformizing parameters* for A. The terminology stems from the theory of Riemann surfaces where a holomorphic function that vanishes simply at a point, may serve as a local coordinate near the point; it defines an analytic isomorphism between an open

neighbourhood of the point and an open neighbourhood of the origin in the complex plane $\mathbb{C}$.

As we are about to explain, in general uniformizing parameters share a few properties with these local coordinates. In what follows, we assume for simplicity that the valuation is normalized.

PROPOSITION 15.13 *Let v be a normalized discrete valuation on the field K and A the corresponding valuation ring. Then the following hold true:*

i) The maximal ideal $\mathfrak{m}$ of A is principal, generated by any uniformizing parameter; that is, any element t with $v(t) = 1$;

ii) If t is a uniformizing parameter, each non-zero element in K may unambiguously be written as $x = a \cdot t^{v(x)}$ with a being a unit in A;

iii) The non-negative powers $\mathfrak{m}^i$ of the maximal ideal are the only non-zero ideals in A. In particular, the ring A is PID*, and it is Noetherian and of Krull dimension one.*

PROOF: We begin by attacking the second assertion, and to that end we choose an element such that $v(t) = 1$. That an element $x \in K$ may be written as $x = a \cdot t^{v(x)}$ with a a *unit* in A, amounts to $x \cdot t^{-v(x)}$ being a unit in A, or in terms of the valuation, that $v(x \cdot t^{-v(x)}) = 0$. But $v(x \cdot t^{-v(x)}) = v(x) - v(x)v(t) = 0$ as $v(t) = 1$. In particular, the element t generates the maximal ideal, and *i)* comes for free.

It remains to prove that all non-zero ideals are of shape $\mathfrak{m}^i$; so let $\mathfrak{a}$ be one. The image $v(\mathfrak{a})$ is a non-empty subset of the set of non-negative integers and thus has a least element, say i. Pick an $x \in \mathfrak{a}$ with $v(x) = i$. Thence $x = a \cdot t^i$ with $a \in A^*$, and $\mathfrak{m}^i = (t^i) \subseteq \mathfrak{a}$. By the minimality of i, it holds true that $v(yx^{-1}) = v(y) - v(x) \geqslant 0$ for any $y \in \mathfrak{a}$. Hence $yx^{-1} \in A$, and $y \in \mathfrak{m}^i$. ❑

A few comments are in place. If the valuation is not normalized, the first assertion still holds true, but the condition $v(t) = 1$ must be replaced by $v(t) = e$ with e being the positive generator for image $v(K^*)$.

Secondly, the proposition reveals that the lattice $\mathcal{I}(A)$ of ideals in A, ordered with reverse inclusion, is just the well ordered lattice of non-negative integers $\mathbb{N}_0$. Albeit all DVR's have this same ideal structure, they can be dramatically different; their innermost secrets are hidden in the group of units.

EXERCISE 15.8 Let A be a DVR with maximal ideal $\mathfrak{m}$ and residue class field $k = A/\mathfrak{m}$. Show that $\mathfrak{m}^n/\mathfrak{m}^{n+1}$ is a one dimensional vector space over k for each natural number $n \geqslant 0$. ★

DVR's *among local rings*

It is of course important to be able to recognise the DVR's among all the crowd of local rings. Having a uniformizing parameter is a strong requirement, and one may wonder if this caracterizes discrete valuation rings. There are, however, local domains

whose maximal ideal is principal which are not DVR's (we saw one in Example 11.10 on page 294), but they are not Noetherian—in fact one even finds such rings of any dimension superior to one—but among local Noetherian domains to have a principal maximal ideal will suffice for being a DVR. Even the *a priori* weaker condition that $\bigcap_i \mathfrak{m}^i = 0$ will do.

Above, in Proposition 15.13, the ideal structure of DVR was revealed; the only non-zero ideals were the powers $\mathfrak{m}^i$ of the maximal ideal, and so the ordered monoid of ideals is isomorphic to $\mathbb{N}_0$. It turns out that this property is charactertistic for a DVR, at least among reduced rings; but even the sole order-type of $\mathcal{I}(A)$ being $\mathbb{N}_0$ implies that A is a DVR (Exercise 15.11 below).

PROPOSITION 15.14 *Let A be a local ring without nilpotent elements that is not a field. Then the following four assertions are equivalent:*

i) A is a DVR*;*
ii) The powers $\mathfrak{m}^i$ of the maximal ideal are the only non-zero ideals in A;
iii) A is Noetherian and $\mathfrak{m}$ is principal;
iv) The maximal ideal $\mathfrak{m}$ is principal and $\bigcap_i \mathfrak{m}^i = 0$.

PROOF: Observe that the maximal ideal $\mathfrak{m}$ is not the zero ideal since A is assumed not to be a field.
i) $\Rightarrow$ ii): This is just statement *iii)* of Proposition 15.13 above.
ii) $\Rightarrow$ iii): Since the powers $\mathfrak{m}^i$ of the maxial ideal are the only non-zero ideals, every ascending chain must terminate (it is in fact finite) so A is Noetherian, and by Nakayama's lemma we may conclude that $\mathfrak{m}^2 \subsetneq \mathfrak{m}$. Pick an element $x \in \mathfrak{m}$, but not in $\mathfrak{m}^2$. Then the principal ideal (x) can not be equal to any power $\mathfrak{m}^i$ with $i \geqslant 2$, and neither can it be zero, hence $(x) = \mathfrak{m}$.
iii) $\Rightarrow$ iv): This is just Krull's Intersection Theorem.
iv) $\Rightarrow$ i): This is close to a repetition of Exercise 9.17 on page 244: Let x generate $\mathfrak{m}$ and let $y \in A$ be any non-zero element. Since $\bigcap_i \mathfrak{m}^i = 0$, there is a largest integer ν so that $y \in \mathfrak{m}^\nu$. We may thus write $y = ax^\nu$, and since $y \notin \mathfrak{m}^{\nu+1}$, the coefficient a is not a member of $\mathfrak{m}$ and is a unit. We conclude that A is a domain (indeed, if $ax^\nu \cdot bx^\mu = 0$ with a and b units, x will be nilpotent), and putting $v(y) = \nu$ gives a function on A^* that extends to a valuation on the fraction field in the usual ways (Exercise 15.1). The salient point is that v is unambiguously defined; indeed, an equality $ax^\nu = bx^\mu$ with $\mu > \nu$ and $a, b \in A^*$ entails that $x^\nu(a - bx^{\mu-\nu}) = 0$ leading to $x^\nu = 0$ since $a - bx^{\mu-\nu}$ is a unit. This in place, the axioms for a valuation ensue painlessly. ❏

Exercises

(15.9) Assume that A is a local ring whose maximal ideal $\mathfrak{m}$ is principal, say $\mathfrak{m} = (t)$. Prove that $\mathfrak{b} = \bigcap_r \mathfrak{m}^r$ is a prime ideal and $t \cdot \mathfrak{b} = \mathfrak{b}$. Conclude that if A is a local

one-dimensional ring without nilpotents whose maximal ideal is principal, then A is a DVR.

(15.10) Show that a local ring without nilpotents all whose ideals are principal is a DVR.

(15.11) Let A be a domain and assume that the lattice of ideals $\mathcal{I}(A)$ ordered by reverse inclusion is order-isomorphic to $\mathbb{N}_0$. Show that A is a DVR.

★

Characterization of DVR*'s among Noetherian domains*

(15.15) We proceeding to single out the DVR's in the class of local Noetherian rings. The two first assertions in the proposition below are equivalent by what we already have done, but the last statement is deeper and is a cardinal characterization of DVR's among Noetherian local rings.

THEOREM 15.16 (DVR'S AMONG NOETHERIAN DOMAINS) *Assume that A is a local Noetherian domain with maximal ideal $\mathfrak{m}$. The following statements are equivalent.*

i) A is a DVR*;*
ii) The maximal ideal is principal;
iii) A is normal and of Krull dimension one.

PROOF: We already have observed that *i)* and *ii)* are equivalent, and that *i)* implies *iii)*, is clear: All valuation rings are normal (Proposition 15.5) and the DVR's are of dimension one (statement *iii)* of Proposition 15.13). The only juicy part of the proof is that the two first assertions follow from the last; we shall show that *iii)* implies *ii)*. The proof leans heavily on the theory of primary decompositions, and we have formulated a separate lemma which also will be useful at a later occasion. From the lemma follows promtly that $\mathfrak{m}$ is principal: indeed, if A is of Krull dimension one, the maximal ideal $\mathfrak{m}$ is the only non-zero prime. ❑

LEMMA 15.17 *Let A be a Noetherian local normal domain and assume that the maximal ideal $\mathfrak{m}$ is associated to a principal ideal. Then $\mathfrak{m}$ is principal.*

PROOF: Let $x \in A$ be such that $\mathfrak{m}$ is associated to (x). Recall that by definition associate primes are transporters, so for some $y \in A$ with $y \notin (x)$ it holds that $y\mathfrak{m} \subseteq (x)$. Thence $\mathfrak{m}yx^{-1} \subseteq A$, but $yx^{-1} \notin A$. If $\mathfrak{m}yx^{-1} \subseteq \mathfrak{m}$ the element yx^{-1} would be integral over A by the third criterion of Proposition 12.4 on page 319 (because A is Noetherian, $\mathfrak{m}$ is finitely generated, and it is faithful as all ideals are). But this is impossible because A is normal and $yx^{-1} \notin A$. We deduce that $\mathfrak{m}yx^{-1} = A$, and consequently there is a relation $zyx^{-1} = 1$ with $z \in \mathfrak{m}$. Then $w = (wyx^{-1})z$ for all $w \in \mathfrak{m}$ (note that wyx^{-1} lies in A), and hence $\mathfrak{m} = (z)$. ❑

(15.18) The Noetherian hypothesis in Proposition 15.16 is essential. There are local rings

with principal maximal ideal of any Krull dimension, but the they will not be Noetherian when the dimension exceeds one (cfr Subsection 15.62), but curiously enough, those of dimension one are. There are also examples of normal domains of Krull dimension one that are not even valuation rings; Krull's example (in Subsection 14.2 on page 359) is one.

15.2 *Normal domains and discrete valuation rings*

A key property of normal Noetherian domains is that their local rings at height one prime ideals all are discrete valuation rings: they are all one dimensional—this is just the definition of being of height one—and being localizations of a normal Noetherian domain they are normal and Noetherian, so Theorem 15.16 above applies. The natural question then arises whether the converse holds true. The answer is no in general; normal rings have an additional and more delicate algebraic trait. A rich source of examples are the cones over space curves in projective three space which are projections from higher projective spaces; the simplest ones are the cones over quartic space curves, and we shall examine one in Example 15.1 below.

A recurring annoyance when practising commutative algebra or algebraic geometry is the occurrence of embedded components, so any result that describes a class of ideals guaranteed to be without embedded components will be welcome. Among such result are the so-called "unmixed theorems" (among which Macaulay's is the most famous); they predict that the height of the intervening associated prime ideals are uniform, and this guarantees that there are no embedded components. In this light, the main result in this section may be seen as a statement that among Noetherian domains the normal ones are precisely those being regular in codimension one and whose principal ideals all are unmixed.

A criterion for being normal

(15.19) The criterion we are about to establish is closely related to Serre's famous R_1–S_2 criterion, but it was first formulated as below by Abraham Seidenberg (See [**?**]). In the course we do not develop the notion of "the depth" of rings, needed for stating Serre's criterion, so we must be content with the minor perturbation of it due to Seidenberg:

THEOREM 15.20 (SEIDENBERG–SERRE) *Let A be a Noetherian domain. Then A is normal if and only if the following two conditions are fulfilled:*

i) The local rings $A_{\mathfrak{p}}$ at each height one prime ideal $\mathfrak{p}$ is a DVR*;*

ii) Each principal ideal is "unmixed"; i. e. it has no embedded components.

The first condition in the theorem is usually referred to as R_1, or expressed in words, that A is "regular in codimension one", and the second condition bears the label S_2;

indicating that every ideal in A of height at least two is of depth* at least two. The proof of the theorem is based on the following lemma:

*This means that there are elements x, y in $\mathfrak{p}$ so that x is not a zero divisor in A and y not one in $A/(x)A$, and it is equivalent to the second condition in the theorem.

LEMMA 15.21 *A domain A is the intersection $\bigcap_{\mathfrak{p}} A_{\mathfrak{p}}$ where $\mathfrak{p}$ runs through the prime ideals associated to principal ideals.*

PROOF: Seeking a contradiction, we assume there is an element ab^{-1} that lies in $\bigcap_{\mathfrak{p}} A_{\mathfrak{p}}$, but not in A. Consider the ideal $\mathfrak{a} = \{\, y \in A \mid ya \in (b) \,\}$, which is a proper ideal since $ab^{-1} \notin A$. Let $\mathfrak{p}$ be an associated prime to $(b)A$. Then $ab^{-1} \in A_{\mathfrak{p}}$ by assumption, and we may write $ab^{-1} = cd^{-1}$ with $c, d \in A$ but $d \notin \mathfrak{p}$. Hence $ad = bc$, and $d \in \mathfrak{a} \subseteq \mathfrak{p}$, which is absurd. ❑

PROOF OF THEOREM 15.20: We start by observing that A is normal when the two conditions are fulfilled: indeed, DVR's are normal and intersections of normal rings are normal, and moreover, the second condition combined with Krull's Principal Ideal Theorem ensures that all primes associated to a principal ideal are of height one.

Let us then show the other implication: that the two conditions are fulfilled when A is normal. We have already observed that in that case the local rings at height one primes are DVR's, so let $\mathfrak{p}$ be a prime in A associated to a principal ideal (x). Consider the local ring $A_{\mathfrak{p}}$. Its maximal ideal $\mathfrak{m} = \mathfrak{p}A_{\mathfrak{p}}$ persists being associated to $(x)A_{\mathfrak{p}}$, and citing Lemma 15.17 on page 379 we conclude that the maximal ideal $\mathfrak{m}$ is principal. Then $A_{\mathfrak{p}}$ is a discrete valuations ring; consequently $\mathfrak{p}$ is of height one, and therefore it can not be embedded. ❑

Corollaries

(15.22) The theorem has a corollary important in geometry, which in a geometric parlance loosely says that rational functions on a normal varieties can be extended over codimension two subsets; or equivalently, that the loci where they are not defined, are of codimension one. It is commonly referred to as *Hartogs' Extension Theorem*, even though it merely is an algebraic reflection of a much deeper result from complex function theory, proved by Friedrich Hartogs.

COROLLARY 15.23 (HARTOGS' EXTENSION THEOREM) *A normal Noetherian domain A satisfies $A = \bigcap_{\mathfrak{p}} A_{\mathfrak{p}}$ where the intersection extends over all prime ideals $\mathfrak{p}$ of height one.*

(15.24) Another by-product of Theorem 15.16 is that the only height one primary ideals in a normal Noetherian domain are the symbolic* powers of the height one primes. And because it also ensures that principal ideals are without embedded components, the principal ideals decompose as intersections of symbolic powers. Such decompositions are entirely described by the occurring "exponents". Each local ring $\mathfrak{p}A_{\mathfrak{p}}$ is a DVR and corresponds to a valuation, $v_{\mathfrak{p}}$ say, on the fraction field K of A. The exponent of a

*Recall that the symbolic power of a prime $\mathfrak{p}$ is $\mathfrak{p}^{(\nu)} = \mathfrak{p}^{\nu}A_{\mathfrak{p}} \cap A$; see Exercise 10.11 on page 275

"factor" $\mathfrak{p}^{(\nu)}$ in the decomposition of a principal ideal (f) is given as $\nu = v_{\mathfrak{p}}(f)$, and may be thought of as the generic order of vanishing of f along the subset $V(\mathfrak{p})$ of Spec A.

PROPOSITION 15.25 *Let A be a Noetherian domain.*

i) If A is normal and $\mathfrak{p}$ is a prime ideal of height one, then the $\mathfrak{p}$-primary ideals are precisely the symbolic powers $\mathfrak{p}^{(\nu)}$;

ii) The domain A is normal if and only if for every non-zero $f \in A$ the minimal primary decomposition of the principal ideal (f) is of shape

$$(f) = \mathfrak{p}_1^{(\nu_1)} \cap \ldots \cap \mathfrak{p}_r^{(\nu_r)}.$$

PROOF: Assume first that A is normal. To prove *i)*, observe that for each height one prime $\mathfrak{p}$, the local ring $A_{\mathfrak{p}}$ is a DVR with maximal ideal $\mathfrak{p}A_{\mathfrak{p}}$. Hence if $\mathfrak{q}$ is $\mathfrak{p}$-primary, it holds that $\mathfrak{q}A_{\mathfrak{q}}$ is power of $\mathfrak{p}A_{\mathfrak{p}}$, say $\mathfrak{q}A_{\mathfrak{p}} = \mathfrak{p}^{\nu}A_{\mathfrak{p}}$. Consequently, by definition of a symbolic power, we find $\mathfrak{q} = \mathfrak{q}A_{\mathfrak{p}} \cap A = \mathfrak{p}^{\nu}A_{\mathfrak{p}} \cap A = \mathfrak{p}^{(\nu)}$.

To prove the necessary part of *ii)*, we cite Krull's Principal Ideal Theorem: every minimal prime of (f) is of height one. So by statement *ii)* of Theorem 15.20 all associated primes are minimal and we conclude by *i)*.

Attacking the converse implication in *ii)*, we assume that all principal ideals have a primary decomposition as described. The second condition of 15.20 is then automatic, so we need only verify the first. To that end, let $\mathfrak{p}$ be a height one prime, say minimal over x. Then $(x) = \mathfrak{p}^{(\nu)} \cap \mathfrak{a}$, where $\mathfrak{a}$ is the intersection of primary ideals not belonging to $\mathfrak{p}$. We deduce that $(x)A_{\mathfrak{p}} = \mathfrak{p}^{(\nu)}A_{\mathfrak{p}} = \mathfrak{p}^{\nu}A_{\mathfrak{p}}$ which is $\mathfrak{p}A_{\mathfrak{p}}$-primary (because $\mathfrak{p}A_{\mathfrak{p}}$ is maximal). Lemma 15.17 then gives that $\mathfrak{p}A_{\mathfrak{p}}$ is principal, and $A_{\mathfrak{p}}$ is a DVR after Proposition 15.14 on page 378. ❑

Exercises

✲ (15.12) Let $\mathfrak{p}$ be a principal prime ideal in the ring A which is generated by a non-zero divisor x. Show that $\mathfrak{p}^{(v)} = \mathfrak{p}^{v}$. HINT: Induction on ν.

(15.13) Let A be a Noetherian local ring A with maximal ideal $\mathfrak{m}$ and residue class field k.

a) Prove that if some power $\mathfrak{m}^i$ is principal, then A is a DVR;

b) Prove that if $\dim_k \mathfrak{m}^i/\mathfrak{m}^{i+1} = 1$ for one i, then A is a DVR.

★

EXAMPLE 15.1 *The cone over a quartic space curve:* The subring $R = k[x^4, x^3y, x^2y^2, xy^3, y^4]$ of $k[x, y]$ is called the "*cone over a quartic rational normal curve*". In exercise 12.10 on page 327 you were in particular (with $d = 4$) asked to prove that R is normal ring.

The ring R is a quotient of the polynomial ring $k[t_{40}, t_{31}, t_{22}, t_{13}, t_{04}]$ with t_{ij} being variables, just send t_{ij} to x^iy^j. So Spec R is a closed subset of $\mathbb{A}_k^5$. It is a cone with apex at the origin, in the sense that any line joining a point on Spec R to the origin is entirely

contained in Spec R; and it is the cone over a curve; the *rational normal quartic curve*. The idea of the example is to project the curve into a hyperplane by forgetting one of the coordinates (*e.g.* the middle one), which on the level of cones corresponds to the inclusion of tings

$$A = k[x^4, x^3y, xy^3, y^4] \subseteq k[x^4, x^3y, x^2y^2, xy^3, y^4] = R. \tag{15.1}$$

Such projections of curves tend to be isomorphisms, and in our case it certainly will be, so on the level of cones the projections only affect the apices. I our case, these apices; that is, the points lying at the origin of $\mathbb{A}^5_k$ in the two cones, correspond to the ideals $\mathfrak{m} = A \cap (x, y)$ and $\mathfrak{n} = R \cap (x, y)$ respectively.

The extension (15.1) is an integral extension: indeed, R is generated over A by the element x^2y^2 which is integral since it satisfies the equation $t^2 - x^4 \cdot y^4 = 0$.

We contend that for all primes $\mathfrak{p}$ in A different from $\mathfrak{m}$ it holds that $A_\mathfrak{p} = R_\mathfrak{p}$. This hinges on the two facts that $\mathfrak{a} = (x^4, y^4)$ is an $\mathfrak{m}$-primary ideal (one easily checks that $\mathfrak{m}^4 \subseteq \mathfrak{a}$), from which follows that any prime different from $\mathfrak{m}$ does not contain both the elements x^4 and y^4, and secondly, that $x^2y^2 = (x^3y)^2x^{-4} = (xy^3)^2y^{-4}$. A consequence is that $A_\mathfrak{p}$ is a DVR for all height one primes $\mathfrak{p}$; indeed, as $\operatorname{ht}\mathfrak{m} = 2$, we know that $A_\mathfrak{p} = R_\mathfrak{p}$ when $\operatorname{ht}\mathfrak{p} = 1$, and by what you did (or should have done) in Exercise 12.10 the ring R is normal, and hence $R_\mathfrak{p}$ is a DVR.

However, A is not normal as we saw. What goes wrong is that the second condition in Theorem 15.20, the depth-condition, is not fulfilled at the apex: For instance, $\mathfrak{m}$ is associated to the principal ideal (x^4): the element x^6y^2 does not belong to (x^4) (precisely because x^2y^2 is not an element in A), and it holds true that $((x^4) : x^6y^2) = \mathfrak{m}$; indeed, one easily establishes the equalities

$$y^4 \cdot x^6y^2 = x^4 \cdot (x^3y)^2 \qquad xy^3 \cdot x^6y^2 = x^4 \cdot x^3y \cdot y^4 \qquad x^3y \cdot x^6y^2 = (x^4)^2 \cdot xy^3$$

☆

Exercises

In both the following exercises the notation is as in the example above.

(15.14) Show that the minimal primary decomposition of the ideal (x^4) in A is given as $(x^4) = (x^4, x^3y, y^3x) \cap (x^4, y^4)$.

(15.15) Find an element $e \in A$ such that $((x^3y) : e) = \mathfrak{m}$.

★

15.3 *Dedekind rings*

The original Dedekind rings, the rings that Dedekind studied and for which he proved his legendary factorisation theorem, were the ring of integers in algebraic function

fields, that is, the integral closures of $\mathbb{Z}$ in finite extensions of the rationals $\mathbb{Q}$. In the introduction to this chapter we referred to the discrete valuation rings as local variants of the Dedekind rings, and faithful to this point of view, we say that a Noetherian domain A is a *Dedekind ring* or a *Dedekind domain* if all the local rings $A_{\mathfrak{m}}$ at maximal ideals $\mathfrak{m}$ are discrete valuation rings.

Dedekind ring (Dedekind-ring)

Dedekind domain (Dedekind-område)

(15.26) There are serveral characetrisations of Dedekind rings, and we sum of the most important ones in the following proposition:

PROPOSITION 15.27 *Assume that A is a Noetherian domain. The following four statements are equivalent*

i) A is normal of Krull dimension one;
ii) Each maximal ideal in A is projective;
iii) Each maximal ideal is invertible;
iv) The ring A is a Dedekind ring; i. e. each localization $A_{\mathfrak{m}}$ in a maximal ideal is a DVR.

Moreover, in a Dedekind ring every non-zero ideal is projective and invertible.

PROOF: We begin with showing the equivalence of *i)* and *iv)*, and observe that being normal is a local property (Proposition 12.19 on page 323) so that A is normal if and only if all localizations $A_{\mathfrak{m}}$ are. Since $\dim A_{\mathfrak{m}} \leqslant \dim A$, it follows that if $\dim A = 1$, the same holds for all localizations; and if $\dim A_{\mathfrak{m}} = 1$ for all $\mathfrak{m}$, clearly no chain of prime ideals can have more than two members. Citing Proposition 15.16 we have shown the equivalence.

The equivalence of *ii)* and *iv)* ensues from finitely generated modules being projective is equivalent to they being locally free (Corollary 8.2 on page 204). A maximal ideal $\mathfrak{m}$ therefore is projective if and only if $\mathfrak{m}A_{\mathfrak{m}}$ is free (indeed, $\mathfrak{m}A_{\mathfrak{p}} = A_{\mathfrak{p}}$ for all $\mathfrak{p} \neq \mathfrak{m}$); that is, if and only if $\mathfrak{m}A_{\mathfrak{m}}$ is principal. And we are through citing 15.16 again.

Finally, the equivalence of *ii)* and *iii)* was already established in Proposition 8.24 on page 213 where invertible ideals were characterized; note that with the Noetherian hypothesis, all ideals are finitely generated.

Since all non-zero ideals in a DVR are invertible, the final assertion follows by the Localness of being equal, since the inclusion $\mathfrak{a}\mathfrak{a}^{-1} \subseteq A$ becomes an equality when localized at any maximal ideal where $A_{\mathfrak{m}}$ is a DVR (forming products and transportes of ideals commute with localization). ❑

Examples

(15.2) *Rings of integers:* Due to the first criterion, the original Dedekind rings; that is, the integral closures of $\mathbb{Z}$ in any finite field extension of $\mathbb{Q}$, will be Dedekind rings. By Going–Up (Corollary 12.35 on page 330) they will be of dimension one, and being finitely generated (Theorem 12.45 on page 335) modules over $\mathbb{Z}$ they are Noetherian, and of course, they will be integrally closed in K (Proposition 12.11 on page 320).

The same reasoning applies to every one dimensional domain provided one can prove that the integral closure is Noetherian. We established this for all algebras of finite type over a field (Theorem 12.47); but in fact, it holds true whenever the field extension is finite (the integral closure of a one dimensional domain in a finite extension of its fraction field will always be Noetherian, even though it not necessarily is finite over A).

(15.3) *Regular plane curves—The Jacobi criterion:* Another inexhaustible source of Dedekind rings are the coordinate rings of affine non-singular curves, and among those the plane curves are the more accessible. So let us consider an irreducible curve $C \subseteq \mathbb{A}^2_k$ where we for simplicity assume that k is an algebraically closed field. This means that $C = V((f))$ where f is an irreducible polynomial in the polynomial ring $k[x,y]$, and $C = \operatorname{Spec} A$ with $A = k[x,y]/(f)$.

Pick a closed point $P \in C$; it corresponds to a maximal ideal $\mathfrak{m} = (x-a, y-b)$ in A. We intend to examine the question whether the local ring $A_\mathfrak{m}$ is a DVR or not. It is certainly a Noetherian domain of dimension one, so the issue is whether $\mathfrak{m}$ is principal or not. Now, Taylor expansion gives $f(x,y) = \alpha(x-a) + \beta(x-b) + g(x,y)$ where $g(x,y) \in \mathfrak{m}^2$, where $\alpha = f_x(a,b)$ and $\beta = f_y(a,b)$. And since $f = 0$ in A this translates into the identity

$$\alpha(x-a) + \beta(y-b) + g = 0$$

in A. So if one of the partials does not vanishes at P, say that $\alpha \neq 0$, it follows that $(x-a) = -\alpha^{-1}\beta(x-b) - \alpha^{-1}g$, and since $\mathfrak{m}$ is generated by $x-a$ and $y-b$, it ensues that $\mathfrak{m}$ is principal.

Now, if the two partials f_x and f_y have no common zeros along C, which translates into the algebraic condition that the ideal (f, f_x, f_y) generated by the three is not proper, all the local rings $A_\mathfrak{m}$ will be DVR s, and we arrive at the following:

PROPOSITION 15.28 *Let k be an algebraically closed field. Let f be an irreducible polynomial in the polynomial ring $k[x,y]$ and assume that $(f, f_x, f_y) = k[x,y]$. Then $k[x,y]/(f)$ is a Dedekind ring.*

☆

Exercises

(15.16) *Hyperelliptic curves.* A particular class of plain regular curves are the so-called affine *hyperelliptic curves*[1]. They are given by an equation $y^2 - p(x) = 0$ where $p(x)$ is a polynomial in $k[x]$ without multiple roots. We assume that k is algebraically closed and not of characteristic two.

[1] There is a different and intrinsic definition of projective hyperelliptic curves in algebraic geometry, but many affine ones may be put on this form.

a) Show that the ring $k[x,y]/(y^2 - p(x))$ is a Dedekind ring.
b) Show that if a is not a root of p, then $x - a$ may serve as a uniformizing parameter at the point $(a, \sqrt{p(a)})$, and that y will be one when $p(a) = 0$.

✲ (15.17) Let d be a square free integer with $d \equiv 1 \mod 4$, and consider the quadratic extension $A = \mathbb{Z}[\sqrt{d}]$.

a) Show that the ideal $\mathfrak{m} = (\sqrt{d}+1, \sqrt{d}-1)$ is maximal in $\mathbb{Z}[\sqrt{d}]$ with residue class field equal to $\mathbb{F}_2$;
b) Show that $\mathfrak{m}A_\mathfrak{m}$ requires two generators and conclude that $A_\mathfrak{m}$ is not a DVR;
c) Show that $\mathfrak{m}$ is the only maximal ideal in A such that $A_\mathfrak{m}$ is not a DVR.

(15.18) Let $A = k[x,y]/(y^2 - x^3)$, and we persist with writing x and y for the images of x and y in A; so that $y^2 = x^3$. We are interested in ideals of A that are contained in the maximal ideal $\mathfrak{m} = (x,y)$.

We contend they are all of the form $(x^r, x^{r-2}(a+bx))$.

★

(15.29) One of the consequences of Proposition 15.27 is that on a Dedekind ring A, or rather on its fraction field K, there is for each maximal ideal $\mathfrak{m}$ a valuation $v_\mathfrak{m}$ whose valuation ring equals $A_\mathfrak{m}$. There may however be several more valuations on K, but among those being non-negative on A they are all.

PROPOSITION 15.30 *Assume that A is Dedekind ring with fraction field K and let v be a valuation on K which is non-negative on A. Then there is a maximal ideal $\mathfrak{m}$ in A such that $v_m = v$ up to normalization.*

PROOF: Denote by B the valuation ring associated to v and by $\mathfrak{n}$ its maximal ideal. By assumption $A \subseteq B$.

It suffices to see that $\mathfrak{n} \cap A \neq 0$. Indeed, since A is of dimension one, $\mathfrak{m} = \mathfrak{n} \cap A$ being prime will be maximal, and hence $A_\mathfrak{m} \subseteq B$. Now, $A_\mathfrak{m}$ is a DVR as A is Dedekind, and by Exercise 15.4 on page 375 DVR's are maximal subrings, and we infer that $A_\mathfrak{m} = B$. Hence v and $v_\mathfrak{m}$ are equal up to normalization.

Let $x \in \mathfrak{n}$ and write $x = ab^{-1}$ with $a, b \in A$. Then $0 < v(x) = v(a) - v(b) \leqslant v(a)$, hence $v(a) > 0$ and a belongs to $\mathfrak{n}$. ❑

EXAMPLE 15.4 The hypothesis in the Lemma above that v be non-negative on A is evidently necessary, a stupid example being the localization A_x in an element $x \in A$. Each maximal ideal $\mathfrak{m}$ containing x induces a valuation v_m on K that is not of the required form relative to A_x.

A more interesting example is derived from the degree of a rational function. Every polynomial in $k[t]$ has a degree $\deg f$, and the negative $-\deg f$ of the degree satisfies the two valuation axioms. Indeed, we have $\deg fg = \deg f + \deg g$ and

$\deg(f+g) \leqslant \max(\deg f, \deg g)$, and changing sign in the latter we obtain the inequality $-\deg(f+g) \geqslant -\max(\deg f, \deg g) = \min(-\deg f, -\deg g)$.

Consequently, it induces a valuation on the field $k(t)$ of rational functions. The function t^{-1} will be a uniformizing parameter and of course, it does not belong to any maximal ideal in $k[t]$ (not even to $k[t]$ itself). This valuation is just the order of vanishing at infinity of the rational functions. ☆

EXERCISE 15.19 The aim of this exercise is to construct another example of a valuation on a Dedekind ring A that is not induced by a maximal ideal. Students acquainted with projective geometry will at once see through the example and recognize the extra valuation as the order of vanishing of functions at the point of infinity of an elliptic curve.

a) Consider the algebra $A = k[x,y]$ with constituting relation $y^2 = x(x^2-1)$. The fraction field of A is a quadratic extension of $k(x)$ with $y^2 = x(x^2-1)$. Show that there is a valuation v on K with $v(x) = 2$ and $v(y) = 1$. HINT: If $\mathfrak{m} = (x,y)$, show that $A_{\mathfrak{m}}$ is a DVR.
b) Consider the two elements $z = y^{-1}$ and $t = xy^{-1}$ in K, and show that $K = k(z,t)$, Show that this a quadratic extension of $k(t)$ with $tz^2 - z + t^3 = 0$. Show that there is a valuation w on K with $w(t) = 1$ and $w(z) = 3$. HINT: let $B = k[z,t]$ with constituting relation $z - t(t^2 - z^2)$ and show that K is the fraction field of B; and then that $B_{\mathfrak{n}}$ with $\mathfrak{n} = (z,t)$ is a DVR.
c) Conclude that $w(x) = -2$ and $w(y) = -3$.

★

Main Theorem on ideals in Dedekind rings

We have now come Richard Dedekind's honoured generalization of The Fundamental Theorem of Arithmetic, his theorem about factorization of ideals in the rings named after him. After the groundbreaking contributions of Noether and Lasker, this is not very difficult, but is merely an analysis of how primary decompositions are shaped in Dedekind rings.

(15.31) So let $\mathfrak{a}$ be an ideal in the Dedekind ring A. That the local ring $A_{\mathfrak{m}}$ at a maximal ideal $\mathfrak{m}$ is a DVR, has the effect that $\mathfrak{a}A_{\mathfrak{m}} = \mathfrak{m}^{\nu}A_{\mathfrak{m}}$ for an unambiguously determined non-negative integer ν, simply because the powers of $\mathfrak{m}A_{\mathfrak{m}}$ are all ideals in $A_{\mathfrak{m}}$. And in that way we may associate a number $v_{\mathfrak{m}}(\mathfrak{a})$ with $\mathfrak{a}$ for each maximal ideal $\mathfrak{m}$, it is called the order of *order of* $\mathfrak{a}$ at $\mathfrak{m}$, and these numbers will be the exponents appearing in Dedekind's theorem. Observe that an ideal $\mathfrak{a}$ is contained in a maximal ideal $\mathfrak{m}$ precisely when $\mathfrak{a}A_{\mathfrak{m}}$ is proper; that is, when $v_{\mathfrak{m}}(\mathfrak{a}) > 0$.

The order of ideals at a maximal ideals (ordenen til idealer i et maksimalt ideal)

THEOREM 15.32 (MAIN THEOREM OF DEDEKIND RINGS) *Let A be a Dedekind ring. Then*

every non-zero ideal $\mathfrak{a}$ equals a product of maximal ideals. That is

$$\mathfrak{a} = \mathfrak{m}_1^{\nu_1} \cdot \ldots \cdot \mathfrak{m}_r^{\nu_r}, \tag{15.2}$$

where the $\mathfrak{m}_i$'s are different maximal ideals, and the ν_i's are natural numbers. The ideals $\mathfrak{m}_i$ and the exponents ν_i are unambiguously determined by $\mathfrak{a}$.

Note that since the $\mathfrak{m}_i$'s are maximal, their powers are primary, and since they are comaximal, their product equals their intersection. Thus the $\mathfrak{m}_i$'s appearing in (15.2) are the minimal primes of $\mathfrak{a}$, and the exponents ν_i are the orders $\nu_i = v_{\mathfrak{m}_i}(\mathfrak{a})$ of $\mathfrak{a}$ at $\mathfrak{m}_i$; or if one prefers, the ν_i's are the unique numbers such that $\mathfrak{a}A_{\mathfrak{m}_i} = \mathfrak{m}_i^{\nu_i} A_{\mathfrak{m}}$. This takes care of the unicity. In a compact notation the equality (15.2) takes the form

$$\mathfrak{a} = \prod \mathfrak{m}^{v_{\mathfrak{m}}(\mathfrak{a})},$$

(where it is understood that the infinitely many factors equal to A do not contribute to the product).

PROOF: Only existence needs an argument. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ be the minimal primes of $\mathfrak{a}$. As always in Noetherian rings they are finite in number, and because A is one-dimensional, they will all be maximal. Letting $\nu_i = v_{\mathfrak{m}}(\mathfrak{a})$ we have the inclusion

$$\mathfrak{a} \subseteq \mathfrak{m}_1^{\nu_1} \cap \ldots \cap \mathfrak{m}_r^{v_r} = \mathfrak{m}_1^{\nu_1} \cdot \ldots \cdot \mathfrak{m}_r^{v_r} \tag{15.3}$$

where the intersection equals the product as the different powers ${\mathfrak{m}_i}^{\nu_i}$ are comaximal. Finally, by the very definition of the integers ν_i, the inclusion in (15.3) becomes an equality when localised at any maximal ideal and therefore is an equality. ❑

(15.33) The orders of results of the usual operations on ideals are easily expressibel in terms of the orders of the invoved ideals:

PROPOSITION 15.34 *Let A be a Dedekind ring and $\mathfrak{a}$ and $\mathfrak{b}$ two ideals in A. Then the following statements hold true for all prime ideals $\mathfrak{m}$:*

i) $v_{\mathfrak{m}}(\mathfrak{a} \cap \mathfrak{b}) = \max(v_{\mathfrak{m}}(\mathfrak{a}), v_{\mathfrak{m}}(\mathfrak{b}))$;
ii) $v_{\mathfrak{m}}(\mathfrak{a} + \mathfrak{b}) = \min(v_{\mathfrak{m}}(\mathfrak{a}), v_{\mathfrak{m}}(\mathfrak{b}))$;
iii) $v_{\mathfrak{m}}(\mathfrak{a}\mathfrak{b}) = v_{\mathfrak{m}}(\mathfrak{a}) + v_{\mathfrak{m}}(\mathfrak{b})$;

Moreover, it holds true that

iv) $\mathfrak{a} \subseteq \mathfrak{b}$ *if and only if* $v_{\mathfrak{m}}(\mathfrak{a}) \leqslant v_{\mathfrak{m}}(\mathfrak{b})$ *for all* $\mathfrak{m}$.

PROOF: Let $\mathfrak{m}$ be a maximal ideal in A. When localized at $\mathfrak{m}$, the ideals $\mathfrak{a}$ and $\mathfrak{b}$ are transformed into $\mathfrak{a}A_{\mathfrak{m}} = \mathfrak{m}^{\nu} A_{\mathfrak{m}}$ and $\mathfrak{b}A_{\mathfrak{m}} = \mathfrak{m}^{\mu} A_{\mathfrak{m}}$, and consequently we find:

$$(\mathfrak{a} \cap \mathfrak{b})A_{\mathfrak{m}} = \mathfrak{a}A_{\mathfrak{m}} \cap \mathfrak{b}A_{\mathfrak{m}} = \mathfrak{m}^{\nu} A_{\mathfrak{m}} \cap \mathfrak{m}^{\mu} A_{\mathfrak{m}} = \mathfrak{m}^{\max(\nu,\mu)} A_{\mathfrak{m}}$$
$$(\mathfrak{a} + \mathfrak{b})A_{\mathfrak{m}} = \mathfrak{a}A_{\mathfrak{m}} + \mathfrak{b}A_{\mathfrak{m}} = \mathfrak{m}^{\nu} A_{\mathfrak{m}} + \mathfrak{m}^{\mu} A_{\mathfrak{m}} = \mathfrak{m}^{\min(\nu,\mu)} A_{\mathfrak{m}}.$$

This takes care of *i)* and *ii)*, and statement *iii)* is trivial. One of the implications in *iv)* is obvious, so assume that $v_{\mathfrak{m}}(\mathfrak{a}) \leqslant v_{\mathfrak{m}}(\mathfrak{b})$ for all maximal ideals $\mathfrak{m}$. This means that $\mathfrak{a}A_{\mathfrak{m}} \subseteq \mathfrak{b}A_{\mathfrak{m}}$ for all $\mathfrak{m}$; and the inclusion $\mathfrak{b} \subseteq \mathfrak{a} + \mathfrak{b}$ will when localized in each $\mathfrak{m}$, be an equality, hence it is an equality by the localness of being equal, and we can conclude that $\mathfrak{a} \subseteq \mathfrak{b}$. ❑

EXAMPLE 15.5 There is a clear qualitative explanation of the ambiguous factorization $2 \cdot 3 = (1 + i\sqrt{5}) \cdot (1 - i\sqrt{5})$ in the ring $\mathbb{Z}[i\sqrt{5}]$, which involves the following three ideals $\mathfrak{q} = (2, 1 + i\sqrt{5})$, $\mathfrak{p}_1 = (3, 1 + i\sqrt{5})$ and $\mathfrak{p}_2 = (3, 1 - i\sqrt{5})$. They are all prime ideals, and an easy computation shows that $\mathfrak{q}^2 = (2)$ and $\mathfrak{p}_1\mathfrak{p}_2 = (3)$. It follows that the factorization of (6) into a product of prime ideals is $(6) = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}^2$. One checks that $\mathfrak{p}_1\mathfrak{q} = (1 + i\sqrt{5})$ and that $\mathfrak{p}_2\mathfrak{q} = (1 - i\sqrt{5})$. The factorization $2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5})$ thus arise from grouping the four factors of $\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}^2$ into pairs in two different ways. ☆

Exercises

(15.20) Let k be a field and $A = k[x, y]$ with constituting relation $(y^2 - x(x^2 - 1))$. Determine the factorization of $(y - ax)$ in A.

(15.21) Let A be a Dedekind ring and let $a \in A$ be an element. Assume that $(a) = \prod_{1 \leqslant i \leqslant r} \mathfrak{p}_i$ with the $\mathfrak{p}_i$'s being prime ideals. Show that a is irreducible if and only if for no proper subset $J \subsetneq \{1, \ldots, r\}$ the ideal $\prod_{i \in J} \mathfrak{p}_i$ is principal.

(15.22) Let A be a Noetherian ring and S a multiplicative set in A. Assume that S is generated by prime elements; that is, every element $s \in S$ is a product $s = p_1 \cdot \ldots \cdot p_r$ with p_i a prime element and $p_i \in S$.

a) Show that if $\mathfrak{p} \subseteq A$ is a prime ideal and $\mathfrak{p}A_S$ is principal, then $\mathfrak{p}$ is principal.

b) Assume further that A be a Dedekind ring that is not a PID, and let S be the multiplicative set generated by all prime elements in A. Prove that $S^{-1}A$ is a Dedekind ring where no prime ideal is principal.

★

Corollaries

Combining the Main Theorem with the Chinese Remainder Theorem we deduce three corollaries.

(15.35) Given r points $x_1, \ldots, x_r$ in the complex line $\mathbb{C}$ with r multiplicities $\nu_1, \ldots, \nu_r$ associated to the points, one easily finds a polynomial having a zero of order exactly ν_i at each x_i and having no other zeros. Just take the product of the $(x - x_i)^{\nu_i}$ as i varies from 1 to r.

A similar, but weaker, statement holds true in any Dedekind ring, but with two important modifications. The obvious first twist is to replace the r points by r maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ and ask for an element a with $v_{\mathfrak{m}_i}(a) = \nu_i$. Secondly, and much more substantially, one can no longer guarantee to find elements vanishing *only* at the

specified prime ideals $\mathfrak{m}_i$. They might, and often will, have other zeros away from the given primes.

PROPOSITION 15.36 *Let A be a Dedekind ring. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ be different maximal ideals in A and denote by $v_1 \ldots, v_r$ the corresponding valuations. Given non-negative integers $n_1, \ldots, n_r$. Then there exist elements $a \in A$ so that $v_i(a) = n_i$ for $1 \leqslant i \leqslant r$; or phrased differently, so that $a \in \mathfrak{m}_i^{n_i} \backslash \mathfrak{m}_i{}^{n_i+1}$ for $1 \leqslant i \leqslant r$.*

PROOF: In view of the powers $m_i^{n_i}$ being pairwise comaximal this is just the Chinese Remainder Theorem: the map

$$A \to A/\mathfrak{m}_1^{n_1+1} \times \ldots \times A/\mathfrak{m}_r^{n_r+1}$$

that sends a to $([a]_1, \ldots, [a]_r)$, where $[a]_i$ denotes the class of a modulo $\mathfrak{m}_i{}^{n_i+1}$, is surjective. If for each index i one lets t_i be a uniformizing parameter at $\mathfrak{m}_i$, any element a that maps to the sequence $([t_1]_1^{n_1}, \ldots, [t_r]_r^{n_r})$, is as wanted. ❑

(15.37) Ideals in a Dedekind ring require at most two generators, and this is a particular property even among one-dimensional domains. Heuristically, the coordinate ring of a curve which near the origin approximates the union of the coordinate axes in $\mathbb{C}^n$, will have a maximal ideal that requires n generators.

COROLLARY 15.38 *Any ideal $\mathfrak{a}$ in a Dedekind ring A is generated by at most two elements.*

PROOF: Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ be the minimal primes of $\mathfrak{a}$ and chose an element $a \in A$ such that $v_{\mathfrak{m}_i}(a) = v_{\mathfrak{m}_i}(\mathfrak{a})$. The ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ are certainly among the minimal primes of (a), but there might be others, say $\mathfrak{n}_1, \ldots, \mathfrak{n}_s$. Chose an element $b \in A$ such that $v_{\mathfrak{m}_i}(b) = v_{\mathfrak{m}_i}(\mathfrak{a})$ for $1 \leqslant i \leqslant r$ and $v_{\mathfrak{n}_j}(b) = 0$ for $1 \leqslant j \leqslant s$; *i. e.* such that $b \in \mathfrak{a}$, but $b \notin \mathfrak{n}_j$ for $1 \leqslant j \leqslant s$. Then $(a, b) \subseteq \mathfrak{a}$, and this inclusion becomes an equality when localized at each maximal ideal; hence it is an equality. ❑

EXERCISE 15.23 Consider the polynomial ring $k[x, y, z]$ over the field k. Let $\mathfrak{a} = (x, y) \cap (x, z) \cap (y, z)$. Show that $A = k[x, y, z]/\mathfrak{a}$ is of dimension one, but that $(x, y, z)A$ cannot be generated by two elements. Can you find a domain with the same properties? ★

(15.39) The third corollary comes here:

COROLLARY 15.40 *A Dedekind ring with finitely many maximal ideals is a* PID,

PROOF: Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$ be the maximal ideals in A. Each ideal $\mathfrak{a}$ can according to the Main Theorem be expressed as $\mathfrak{a} = \mathfrak{m}_1^{\nu_1} \cdot \ldots \cdot \mathfrak{m}_r^{\nu_r}$. In view of Proposition 15.36 there is an element $a \in A$ with $v_{\mathfrak{m}_i}(a) = \nu_i$ for each i. Then $(a) \subseteq \mathfrak{a}$ and the inclusion becomes an equality when localized at each $\mathfrak{m}_i$; that is, at every maximal ideal of A. Hence it is an equality. ❑

The ideal class group and the Picard group

For Dedekind rings the ideal class group is, together with the group of units A^*, the most important invariant, and it is fair to say that in traditional algebraic number theory they even were the main objects of study, and the reason is as follows (remember that the theory arose out of the failure of the factorization theorem). As all ideals in a Dedekind ring A are invertible, A is a PID precisely when the Ideal class group vanishes, and since Dedekind domains are factorial if and only if they are principal ideal domains, one has the important:

PROPOSITION 15.41 *A Dedekind domain A is factorial if and only if* $\mathrm{Cl}(A)$ *is trivial.*

(15.42) In the "Kummer-set-up", when element (*i. e.* numbers) are replaced by ideals, one is tempted to state that the group of units A^* is the measure of contraction while $\mathrm{Cl}(A)$ measures the expansion, as illustrated by the exact sequence

$$\{1\} \longrightarrow A^* \longrightarrow K^* \longrightarrow I(A) \longrightarrow \mathrm{Cl}(A) \longrightarrow \{1\}.$$

The Krull–Akizuki Theorem

Back in chapter 12 we experienced the "finiteness issue"; the integral closure of a domain A in a finite extension of the fraction field is not always a finitely generated A-module. However, in the special case of closing up a Dedekind rings, which is important in number theory, the closure will at least be Dedekind. The pertinent issue is whether the integral closure is Noetherian, all other requests for being Dedekind come easily. There are several more general versions of this result treating integral closures of one-dimensional rings, but the "soul" of all other proofs appears in this simplest version (which also is the most important one in number theory).

THEOREM 15.43 (KRULL–AKIZUKI) *Let A be a Dedekind ring with fraction field K and let L be a finite extension of K. Then the integral closure of A in L is a Dedekind ring*

PROOF: By Going–Up, B is one-dimensional, and being the integral closure it is integrally closed, and remains to show it is Noetherian, and that is Corollary 15.46 below. ❑

(15.44) The proof of the integral closure being Noetherian hinges on the following lemma:

LEMMA 15.45 *Let A be a Dedekind ring and let M be a torsion free A-module of finite rank. Then for each maximal ideal $\mathfrak{m}$ of A the dimension of $M/\mathfrak{m}M$ as a vector space over $A/\mathfrak{m}$ is bounded by* $\operatorname{rk} M$.

PROOF: Let $n = \operatorname{rk} M$ and let K denote the fraction field of A. If $m_1, \ldots, m_{n+1}$ be arbitrary elements in M our task is to show that the classes $[m_i]$ in $M/\mathfrak{m}M$ are linearly dependent. To that end, observe that they are dependent in $M \otimes_A K$, so there is a relation

$\sum_{1\leqslant i\leqslant n+1} a_i m_i = 0$ with coefficients $a_i \in A$. The ideal $\mathfrak{a} = (a_1, \ldots, a_{n+1})$ is invertible, and because $\mathfrak{a}^{-1}\mathfrak{a} = A$, there is an $x \in \mathfrak{a}^{-1}$ so that $x\mathfrak{a} \nsubseteq \mathfrak{m}$. Then each xa_i lie in A, and at least one does not lie in $\mathfrak{m}$. Reducing the relation above mod $\mathfrak{m}M$ then yields a linear dependence relation among the classes $[m_i]$. ❏

COROLLARY 15.46 *If B is a the closure of A in L, then B is Noetherian.*

PROOF: By Lying–Over, each non-zero prime $\mathfrak{p}$ in B intersects A in maximal ideal $\mathfrak{m}$; hence $\mathfrak{m}B \subseteq \mathfrak{p}$. By the lemma $B/\mathfrak{m}B$ is of finite dimension over $A/\mathfrak{m}$, and thus it is Noetherian. It follows that each $\mathfrak{p}/\mathfrak{m}B$ is finitely generated, which certainly implies that $\mathfrak{p}$ is finitely generated, and we are through citing Cohen's criterion. ❏
Note that the fibres over $\mathfrak{m}$ are finite; the cardinality is bounded by $[L : K]$.

15.4 *Finitely generated modules over Dedekind rings*

(15.47) Just as is the case for principal ideal domains, it turns out that finitely generated torsion free modules over Dedekind rings are projective, and as a the corollary submodules of finitely generated projective modules will be projective. Over Dedekind rings there is also a quite satisfactory classification of projective modules up to isomorphism: they are all direct sums of invertible ideals (*i. e.* projective modules of rank one)—in a geometric language this says that vector bundles over regular affine curves decompose as direct sums of line bundles—and there is a workable criterion for such sums to be isomorphic.

THEOREM 15.48 (TORSION FREE MODULES OVER DEDEKIND RINGS) *Every finitely generated torsion free module over a Dedekind ring A is isomorphic to a direct sum of (invertible) ideals; or in other words, to a direct sum of projective rank one modules.*

We remind you of Theorem 8.30 about torsion free modules over PID's, and the proof of the precent theorem is *mutatis mutandis* the same as of that result, except that ideals will be projective and not free.
PROOF: The proof goes by induction on the rank of the module, which we call E. Citing Lemma 8.31 on page 216 we may find a non-zero A-linear map $\pi\colon E \to A$. Its image is an ideal $\mathfrak{a}$, which is projective because A is Dedekind, and consequently π is a split surjection onto it image. Hence $E \simeq F \oplus \mathfrak{a}$ where F is torsion free and $\operatorname{rk} F = \operatorname{rk} E - 1$.

If E is of rank one, the rank of F will be zero, and being torsion free it must reduce to zero. Hence E is isomorphic to an ideal. If $\operatorname{rk} E > 1$, we invoke the induction hypothesis to infer that F is a direct sum of ideals. Obviously the same is then true for E. ❏

We immediately obtain the following corollary:

COROLLARY 15.49 *Every finitely generated torsion free module over a Dedekind ring is projective. A submodule of a finitely generated projective module over a Dedekind ring is projective.*

(15.50) Like the Bézout rings being non-Noetherian analogues of the PID's, the Dedekind rings have non-Noetherian cousins called Prüfer rings, in which all finitely generated ideals are projective. The theorem holds true for these, and the proof goes through without modifications, merely supplemented with the observation that direct summands of finitely generated modules are finitely generated. The second assertion in the corollary is not true as stated, but persists being true with the additional assumption that the submodule be finitely generated.

(15.51) From Theorem 15.48 above arises the natural question when two direct sums of ideals are isomorphic, and it seems that Steinitz (at least Kaplansky attributes the result to him) was the first to give an adequate answer. The easier part is that the number of summands—*i. e.* the rank of the two modules—must be equal, the subtler part that the products of the ideals involved must be isomorphic. This gives a complete classification up to isomorphism of finitely generated torsion free modules over Dedekind rings (provided the Picard group is known). To a finitely generated projective module over any domain, there is associated an invertible module called the *determinant* (for those who know those creatures, it is the highest non-vanishing exterior power). In the present case the notion boils down to the product of the involved invertible summands, and in order to be in sync with general accepted usage, we shall write $\det M = \mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r$ when $M = \bigoplus_i \mathfrak{a}_i$ and name it the *determinant* of M (note that this determinant is not *a priori* an invariant of M, but depends on the specific decomposition as a direct sum).

The determinant of a projective module (determinanten til en projective modul)

THEOREM 15.52 (PROJECTIVE MODULES OVER DEDEKIND RINGS) *Any finitely generated torsion free module M over a Dedekind ring is isomorphic to a direct sum of ideals. Two such direct sums are isomorphic if and only they have the same rank, and their determinants are isomorphic.*

The first step in the proof is to treat the special case that all save one of the summands in one of the sums are trivial:

LEMMA 15.53 *There are isomorphisms $\bigoplus_{1\leqslant i\leqslant r} \mathfrak{a}_i \simeq (r-1)A \oplus \mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r$.*

PROOF: We begin with reducing the general case to the case of two summands by induction on r. Let M=$\bigoplus_{1\leqslant i\leqslant r} \mathfrak{a}_i$ be a given direct sum. By induction there is an isomorphism $\bigoplus_{2\leqslant i\leqslant r} \mathfrak{a}_i \simeq (r-2)A \oplus \mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r$, but by the case of two summands there is an isomorphism $\mathfrak{a}_1 \oplus \mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r \simeq A \oplus \mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r$, and consequently we obtain an isomorphism $\mathfrak{a}_1 \oplus (r-2)A \oplus \mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r \simeq (r-1)A \oplus \mathfrak{a}_1 \cdot \ldots \cdot \mathfrak{a}_r$.

When $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are two comaximal ideals, it holds true that $\mathfrak{a}_1 \cap \mathfrak{a}_2 = \mathfrak{a}_1\mathfrak{a}_2$ and $\mathfrak{a}_1 + \mathfrak{a}_2 = A$, so that the Chinese exact sequence takes the form

$$0 \longrightarrow \mathfrak{a}_1\mathfrak{a}_2 \longrightarrow \mathfrak{a}_1 \oplus \mathfrak{a}_2 \longrightarrow A \longrightarrow 0.$$

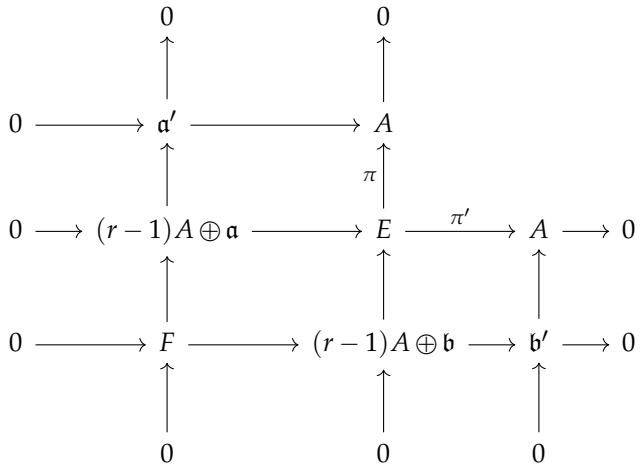
It splits and gives the desired isomorphism. The final step is to see that for every two non-zero ideals $\mathfrak{a}_1$ and $\mathfrak{a}_2$ one may find elements x and y in A so that $x\mathfrak{a}_1$ and $y\mathfrak{a}_2$ are

comaximal: two ideals are comaximal precisely when they are not contained in the same maximal ideal, so let $\{\mathfrak{m}_i\}_i$ and $\{\mathfrak{n}_j\}_j$ be the maximal ideals containing $\mathfrak{a}_1$ and $\mathfrak{a}_2$. The submodules $\mathrm{Hom}_A(\mathfrak{m}_i, \mathfrak{n}_j) = (\mathfrak{n}_j : \mathfrak{m}_i)$ of K are finitely generated and finite in number, and hence their sum S does not equal K (which is not a finitely generated module over A unless A is semi-local in which case A is a PID and the theorem is trivial). So pick an element xy^{-1} from K but not in S. Then $x\mathfrak{m}_i \nsubseteq y\mathfrak{n}_j$ for each i and j, and hence $x\mathfrak{a}_1$ and $y\mathfrak{a}_2$ are comaximal. ❏

The next lemma finishes the proof of Theorem 15.52:

LEMMA 15.54 *Let A be Dedekind ring, $\mathfrak{a}$ and $\mathfrak{b}$ two ideals and r a natural number. If there is an isomorphism $rA \oplus \mathfrak{a} \simeq rA \oplus \mathfrak{b}$, then $\mathfrak{a} \simeq \mathfrak{b}$.*

PROOF: The political correct—and admittedly the best—proof uses exterior powers and determinants of vector bundles. However, we do not have tools of that sophistication at our disposal and shall present a completely elementary *ad hoc* proof. Once the appropriate diagram is displayed, it is an entirely self-propelled induction on r. We may certainly assume that $rA \oplus \mathfrak{a} = rA \oplus \mathfrak{b}$ and shall denote this module by E. Furthermore, we pick two surjections π and π' from E onto A having kernels $(r-1)A \oplus \mathfrak{a}$ and $(r-1)A \oplus \mathfrak{b}$ respectively (we may surely assume that $r \geqslant 1$ and that the submodules $\mathfrak{a}$ and $\mathfrak{b}$ of E are different). They enter in the diagram



with exact columns and rows, where F is the intersection of $(r-1)A \oplus \mathfrak{a}$ and $(r-1)A \oplus \mathfrak{a}$ inside E, and $\mathfrak{a}'$ and $\mathfrak{b}'$ denote their images in A. With attention on the missing upper right corner, and activating the snake lemma, we infer that $A/\mathfrak{a}' \simeq A/\mathfrak{b}'$. This implies that $\mathfrak{a}' = \mathfrak{b}'$, and because all ideals are projective modules, we may conclude that there are isomorphisms $(r-1)A \oplus \mathfrak{a} \simeq F \oplus \mathfrak{a}' \simeq F \oplus \mathfrak{b}' \simeq (r-1)A \oplus \mathfrak{b}$. By induction it follows that $\mathfrak{a} \simeq \mathfrak{b}$. ❏

EXERCISE 15.24 *The approximation theorem.* Assumethat A is a Dedekind ring with fraction fiels K. Let $\nu_1, \ldots, \nu_r$ be distinct discrete valuations on K all being positive on A, and let $n_1, \ldots, n_r$ be integers. Show that there are elements $x \in K$ such that $\nu_i(x) = n_i$.
HINT: Appeal to the Chinese. ★

EXERCISE 15.25 Let $\mathfrak{a}$ and $\mathfrak{b}$ be any two ideals in A. Show that $\mathfrak{a} \oplus \mathfrak{a}^{-1} \simeq \mathfrak{b} \oplus \mathfrak{b}^{-1}$ ★

EXERCISE 15.26 Let k be an algebraically closed field whose carachteristic is not two, and let $A = k[x, y]$ with constituting relation $y^2 = x(x-a)(x-b)$; that is, the coordinate ring of an affine elliptic curve on Weierstrass form. Show that any finitely generated projective module over A is isomorphic to $nA \oplus \mathfrak{m}$ for a uniquely defined maximal ideal $\mathfrak{m}$. ★

15.5 *Appendix: General valuations*

The origin of valuation theory, and the notions of valuations and a valuation rings, is found far back in the history of mathematics, but its systematic study and axiomatic development started in 1912 by the work of the Hungarian mathematician József Kürschák whose aim was to bring Kurt Hensel's theory of p-adic numbers on a solid footing. The theory was soon substantially extended and improved by several people, until it reached the form in which we know it to day, which mainly is due to Wolfgang Krull. General valuation theory was an important part in the formulation of algebraic geometry a la Zariski. In the book "Commutative Algebra" by Oscar Zariski and Pierre Samuel, which was one of the bibles of the field in the 1960's, general valuation theory takes up about 125 pages. In the post-Grothendieck area general valuation rings went out of fashion, even though they are indispensable a few places and seem to have gain popularity the latest years. The Noetherian ones, however, the discrete valuation ring, persisted and are still flourishing. They are indispensable in both contemporary algebraic geometry and number theory.

József Kürschák (1864–1933) Hungarian mathematician

Like most text on commutative algebra we concentrate on the DVR's, but shall use some space on the general ones here. If only for they being good examples of non-noetherian rings that well illustrates the importance of the Noetherian hypothesis. To put it like this: You best appreciate your own way of life by casting a glance on the struggles of your neighbour.

(15.55) The axioms for general valuations are the same as those for discrete valuations, the only difference being the group where the valuation assumes values, the so-called *value group*. It will be any *totally ordered group*; that is, an additively written abelian group G equipped with a total order of the elements. And, of course, the two structures must be compatible so that $\alpha + \gamma < \beta + \gamma$ whenever $\alpha < \beta$.

The set of positive elements of G; *i. e.* those with $\alpha > 0$, form a submonoid G^+ of G, called The *positive cone*, and which is such that* $G^+ \cap -G^+ = \{0\}$ and

*For any subset $S \subseteq G$, the negative of S is the subset $-S = \{ -x \mid x \in S \}$

$G^+ \cup -G^+ = G$. In fact, giving such a submonoid is the same thing as equipping G with a total order; indeed, declaring $\alpha \geqslant \beta$ to mean that $\alpha - \beta \in G^+$, we obtain a total order compatible with the group law: symmetry holds as $G^+ \cap -G^+ = \{0\}$, transitivity because $(\alpha - \beta) + (\beta - \gamma) \in G^+$ and G^+ is closed under addition, and compatibility ensues from $(\alpha + \gamma) - (\beta + \gamma) = \alpha - \beta$. Finally, either $\alpha - \beta$ or $\beta - \alpha$ belongs to G^+ since $G^+ \cup -G^+ = G$, so the order is total.

Totally ordered groups are torsion free; indeed, from $g > 0$ ensues $ng > 0$ for all natural numbers n by a straightforward induction argument.

Example 15.6 Every subgroup of a totally ordered group inherits a total order from the surrounding group; so for instance, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ and all their subgroups are totally ordered. ☆

The lexicographical order (den leksikografiske ordningen)

Example 15.7 Recall that if T is any totally ordered set the cartesian product T^n is endowed with the total order called the *lexicographical order*: an element $x = (x_1, \ldots, x_n)$ precedes an element $y = (y_1, \ldots, y_n)$ (that is, $x > y$) if at the first place i where x and y differ, x_i is greater than y_i. When T is the ordinary alphabet, this order is the one we know well from dictionaries, hence the name. ☆

General valuations (generelle valuasjoner)

(15.56) This paves the way for the definition of a *general valuation* v on a field K with value-group G. It is a surjective map $v\colon K^* \to G$ that abide by the two rules

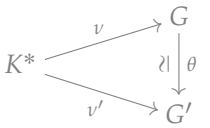
- $v(xy) = v(x) + v(y)$;
- $v(x + y) \geqslant \min(v(x), v(y))$,

whenever the sum $x + y$ is non-zero. The first requirement can of course be phrased as v being a group homomorphism. Notice that the minimum in the second requirement is meaningful since G is totally ordered, and just as for discrete valuation, one finds that $v(1) = 0$ because $v(1) = v(1^2) = 2 \cdot v(1)$, and consequently it holds true that $v(x^{-1}) = -v(x)$ for all $x \in K^*$.

Not all texts require the map v to be surjective, and in fact v's that are not surjective arise naturally in several contexts; however, replacing G by the image of v brings such a situation in accordance with our convention.

Equivalent valuations (ekvivalente valusjoner)

Two valuations v and v' on a field K whose respective value groups are G and G', are said to be *equivalent* if there is an isomorphism $\theta\colon G \to G'$ of ordered groups such that $v' = \theta \circ v$.



Valuations and valuation rings

Recall that a domain A with fraction field K is called a *valuation ring* if it has the property that for any member x of K either x or its inverse x^{-1} belongs to A. Mimicking what we did for discrete valuations, we associate a valuation ring to any valuation v on K

and through the definition

$$A = \{\, x \in K \mid v(x) \geqslant x \,\}.$$

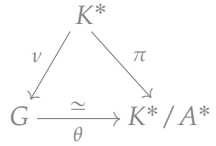
The two axioms yields that A is a ring, it will be closed under multiplication by the first and under addition by the second, and since G is totally ordered, it will be a valuation ring (either $v(x)$ or 0 is the smaller, and in the former case, $-v(x) \geqslant 0$).

Notice that the group A^* of units in A is given as $A^* = \{\, x \in A \mid v(x) = 0 \,\}$; indeed, an element $x \in A$ is precisely a unit when $x^{-1} \in A$; in other words, precisely when $v(x) \geqslant 0$ and $v(x^{-1}) = -v(x) \geqslant 0$. The group of units is thus the kernel of v, and being surjective by assumption, v induced a group isomorphism $K^*/A^* \simeq G$. The elements of K^* are ordered by divisibility; that is x dominates y when $x|y$, or in other words when $xy^{-1} \in A$. The valuation respects this order in that $v(x) \geqslant v(y)$ precisely when $xy^{-1} \in A$, and the isomorphism v induces between K^*/A^* and G will be an isomorphism of ordered groups if K^*/A^* is equipped with the induced order. Note as well that the the quotient K^*/A^* is an abelian group whose elements are the principal fractional ideals—the group operation corresponds to multiplication of ideals

(15.57) Every valuation gives rise to a valuation ring, and the converse is true as well, Indeed, the the quotient K^*/A^* is an abelian group whose elements are the principal fractional ideals—the group operation corresponds to multiplication of ideals—and the set of fractional ideals are totally ordered precisely when A is a valuation ring: the set of principal *ideals* is totally order under inclusion according to Proposition 15.4, and this implies that the set of fractional ideals is totally ordered as well (see Exercise 15.3). So if v is a given valuation, the isomorphism $K^*/A^* \simeq G$ respects the orders and is an isomorphism of ordered groups. The canonical projection $\pi\colon K^* \to K^*/A^*$ will be a valuation (the group K^*/A^* is written multiplicatively and one has to switch to an additive notation). So, we have proven most of the following:

PROPOSITION 15.58 (VALUATIONS VS VALUATION RINGS) *Let K denote a field. Every valuation ring in K arises as associated to a valuation. The valuation is, up to equivalence, unambiguously determined by the subring.*

PROOF: What remains to do, is to verify that the valuation is determined up to equivalence by the valuation ring it determines. But given a valuation ν with ring A, the kernel $\ker \nu$ equals A^*, and by the Isomorphism Theorem for abelian groups it factors through the canonical map $\pi\colon K^* \to K^*/A^*$ showing that π and ν are equivalent. ❑



Final segments (sluttlige segmenter)

Ideals in valuation rings

(15.59) One may wonder what kind of subsets of G the images $v(\mathfrak{a})$ of the ideals are. For instance, the image $v(A)$ is G^+, and the image $v(\mathfrak{m})$ of the maximal ideal is $G^+\backslash\{0\} = \{\, \alpha \in G \mid \alpha > 0 \,\}$. In general the answer is that they are the so-called *final*

segments contained in G^+: subsets $\Gamma \subseteq G^+$ that are closed from above; *i. e.* if $\alpha \in \Gamma$ and $\beta \geqslant \alpha$, then $\beta \in \Gamma$. There is no hocus pocus about this: if $x \in A$, it holds that $v(x) \geqslant 0$, so that $v(xy) = v(x) + v(y) \geqslant v(y)$ for all y; hence if S is a final segment and $v(y) \in S$, it ensues that $v(xy) \in S$. Moreover, one verifies painlessly that complements in G^+ of final segments are characterized by the property that $\alpha \in T$ implies $\gamma \in T$ for all $\gamma \leqslant \alpha$.

Some of the final segments have a least element and are shaped like $\Delta_\alpha = \{\, \beta \mid \beta \geqslant \alpha \,\}$, but many do not—just think of open intervals in the rationals $\mathbb{Q}$. We shall call segments with a least element *principal*. Obviously principal segments correspond to principal ideals: if $x \in A$ is an element with $v(x) = \alpha$, then $v((x)) = \Delta_\alpha$.

Principal segments (hovedsegmenter)

(15.60) The next pertinent question is "what are the prime ideals?" The complement of a prime ideal $\mathfrak{p}$ is closed under multiplication, which means that the complement T of $v(\mathfrak{p})$ in G^+ is closed under addition; hence the complement is a monoid, and in fact, it is the positive part of a subgroup in that $N = T \cup \{0\} \cup -T$ is a subgroup (the point is that G is totally ordered so that if $\alpha \in T$ and $-\beta \in -T$ either $\alpha \geqslant \beta$ and $\alpha + (-\beta) \in T$ or $\beta \geqslant \alpha$ and $\alpha + (-\beta) = -(\beta - \alpha) \in -T$) Additionally T has the property that $\gamma \leqslant \alpha$ lies in T when α does. This inspires the following definition: a proper subgroup $N \subseteq G$ is called *isolated* if elements that are squeezed between two elements from N, belong to N; *i. e.* if $\alpha \leqslant \gamma \leqslant \beta$ and $\alpha, \beta \in N$, then $\gamma \in N$.

Isolated subgroups (isolerte undergrupper)

PROPOSITION 15.61 *Let v be a valuation on K with value group G and let A be the corresponding valuation ring. For the correspondence $S \mapsto v^{-1}(S)$ between subsets of G^+ and A the following hold:*

i) Ideals in A correspond to final segments in G^+;
ii) Principal ideals correspond to principal final segments in G^+;
iii) Prime ideals correspond to complements in G^+ of isolated subgroups.

PROOF: Let $\alpha = v(x)$ and $\beta = v(y)$. The crucial observation is that $v(yx^{-1}) = \beta - \alpha$ so that yx^{-1} lies in A if and only if $\beta \geqslant \alpha$. We infer that, for a subset $S \subseteq G^+$, it holds that S is a final segment if and only if $v^{-1}(S)$ is closed under multiplication with elements from A; that is, $v^{-1}(S)$ is an ideal. ❑

EXERCISE 15.27 Show that the prime ideals correspond to final segments whose complement in G^+ are submonoids; that is, the complements are closed under addition. ★

EXERCISE 15.28 Let G be a totally ordered group. Recall that a subgroup $N \subseteq G$ is called *isolated* if elements that are squeezed between two elements from N, belong to N; *i. e.* if $\alpha \leqslant \gamma \leqslant \beta$ and $\alpha, \beta \in N$, then $\gamma \in N$.

Isolated subgroups (isolerte undergrupper)

a) Prove that kernels of homomorphisms of ordered groups are isolated.
b) Prove that If N is an isolated subgroup, then $N \cap G^+$ is a monoid whose complement in G^+ is a final segment.

c) The ordered group G is called Archimedean if for any two elements α and β there is an integer so that $n\alpha \geqslant \beta$. Prove that an Archimedean group has no non-trivial isolated subgroups.

d) Assume that A is the valuation ring associated with a valuation whose value group is a subgroup of the reals $\mathbb{R}$. Show that $\dim A = 1$.

★

Every totally ordered group appears as a value group

(15.62) With the construction of the monoidal algebras associated to a monoid in Exercise 1.19 on page 24 is rather straight forward to exhibit examples of valuation rings with any given totally ordered group as value group. This shows the extreme diversity of valuations rings.

Let G be a totally ordered group and let k be a field, let $k(G)$ denote the fraction field of the monoidal algebra $k[G]$. The monoidal algebra $k[G]$, consists of elements which are finite k-linear combination $f = \sum_{\gamma\in G} a_\gamma x^\gamma$. For such f define $\nu(f) = \min\{\,\gamma \mid a_\gamma \neq 0\,\}$, which is meaningful as G is totally ordered and every finite subset has a least element. We may expressed f as

$$f = a_\nu x^\nu + \sum_{\gamma>\nu} a_\gamma x^\gamma$$

with $\nu)\nu f$ and with $a_\nu \neq 0$. It is fairly easy to verify that this is a valuation: let g be another element and write g as

$$g = a_\mu x^\mu + \sum_{\gamma>\mu} b_\gamma x^\gamma$$

Then

$$fg = a_\nu b_\mu x^{\nu+\mu} + \sum a_\gamma b_{\gamma'} x^{\gamma+\gamma'}$$

where both $\gamma > \nu$ and $\gamma' > \mu$ so that $\gamma + \gamma' > \nu + \mu$. and hence $\nu(fg) = \nu(f) + \nu(g)$.

As to the second axiom, expand the sum $f + g$ i terms of the x^γ to get

$$f + g = a_\nu x^\nu + b_\mu x^\mu + \sum c_\gamma x^\gamma$$

where the sum extends over γ with $\gamma > \min(\nu, \mu)$. If ν and μ are different, say $\nu < \mu$, if follows tat $v(f + g) = \nu 0 \min(v(f), v(g))$, and in case $\nu = \mu$ the two first terms might cancel, but in any case $v(f + g) \geqslant \nu = \min(v(f), v(g))$. The final step is to extend v to the fraction field of $K[G]$ by the usual procedure from Exercise 15.1 setting $v(fg^{-1}) = v(f) - v(g)$.

Valuation rings with value group $\mathbb{Z}^n$

For each integer i with $1 \leqslant i \leqslant n$ we let N_i be the subgroup of $\mathbb{Z}^n$ of elements whose i first coordinate vanish; that is, the set of elements of the form $(0, \ldots, 0, x_{i+1}, \ldots, x_n)$ and for consistency, we let $N_0 = \mathbb{Z}^n$.

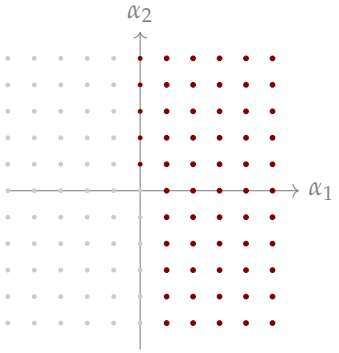
PROPOSITION 15.63 *The N_i's are the only isolated subgroups of $\mathbb{Z}^n$. A valuation ring having $\mathbb{Z}^n$ as value group is of Krull dimension n.*

PROOF: The second statement follows because the subgroups N_i form a chain of length n and because the prime ideals in a valuation ring are in a one-to-one order reversing correspondence with the isolated subgroups of the value group.

To prove that the N_i's are the only isolated subgroups, it suffices, by induction on n, to show that the proper ones are all contained in N_1. But this is quite clear: if N is an isolated subgroup, the smallest non-negative first coordinate x_1 of any element in N must be zero since any element in $\mathbb{Z}^n$ having a positive first coordinated less than x_1, would lie in N as N is isolated. ❑

EXAMPLE 15.8 Concrete examples of valuations with value group $\mathbb{Z}^n$ can be the following valuations on the rational function field $k(x_1, \ldots, x_n)$. It is is just a specific instance of the monoidal construction in Paragraph 15.62 above: As above, the construction is most smoothly performed on the monoidal algebra $A_n = k[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]$ for subsequently to be extended to the fraction field $k(x_1, \ldots, x_n)$. Elements f in A_n are presented as finite sums $f = \sum_\alpha a_\alpha x^\alpha$, with $a_\alpha \in k$ and $\alpha \in \mathbb{Z}^n$, and the valuation is given as $v(f) = \min\{\, \alpha \mid a_\alpha \neq 0 \,\}$. ☆

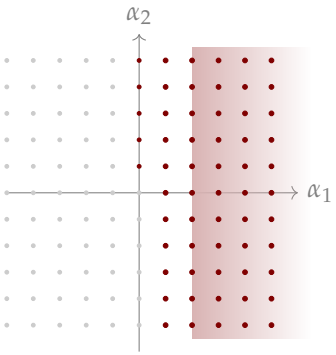
✲ EXERCISE 15.29 Let A be a valuation ring with value group $\mathbb{Z}^n$. Show that the maximal ideal is principal. ★



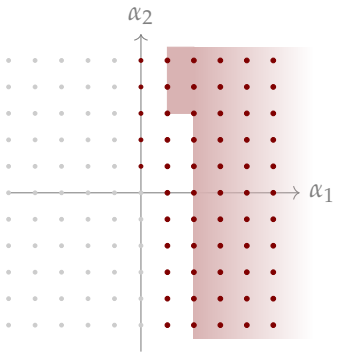
EXAMPLE 15.9 *Ideals in a two-dimensional valuation ring:* It is illustrative to study the ring A_2 and its ideal structure in more more detail. Exploring the ideal structure of A_2 amounts to exploring the segment-structure of the positive cone of $\mathbb{Z}^2$. In the figure we

have depicted part of $\mathbb{Z}^2$, with the positive cone drawn with red dots; remember that $\mathbb{Z}^2$ is ordered such that elements increase upwards and to the right.

There are two types of proper final segments in $\mathbb{Z}^2$. Apart from the principal ones, there is one segment $\Delta_n = \{\, \alpha \mid \alpha_1 \geqslant n \,\}$ for each integers n, having as elements the grid points in the region of the plane to the right of the vertical line $\alpha_1 = n$. If Δ is a proper non-principal segment the first coordinate of elements in Δ will be bounded below, and hence there is smallest first coordinate n. Either all integers appear as a second coordinates of members of Δ having first coordinate n, and $\Delta = \Delta_n$, or there is a least second coordinate, say m. But then Δ will be the principal segment defined by (n, m)



A non-principal segment



A principal segment

The subset of $\mathbb{Z}^2$ with $\alpha_1 = 0$; that is, the α_2-axis, is an isolated subgroup as the inequality

$$(0, \alpha_2) \leqslant (\gamma_1, \gamma_2) \leqslant (0, \beta_2)$$

trivially implies that $\gamma_1 = 0$. And it is the sole proper and non-trivial isolated subgroup. Indeed, assume the subgroup $N \subseteq \mathbb{Z}^2$ is isolated and has an element $\gamma \in N$ whose first coordinates is non-zero. Since the first coordinate dominates the ordering, there is for any $\alpha \in \mathbb{Z}^2$ an integer n so that $n\gamma \geqslant \alpha$, and it follows that $\alpha \in N$ as N is isolated.

In addition to the maximal and the zero-ideal there is one other prime ideal in A, namely the one corresponding to the segment Δ_1. Hence $\dim A = 2$. The other ideals are the principal ones $(x_1^{\alpha_1} x_2^{\alpha_2})$, with $\alpha = (\alpha_1, \alpha_2)$ in the positive cone, and the ones corresponding to the segments Δ_n with $n \geqslant 2$. These, and Δ_1 as well, are generated by any infinite subset of $\{\, x_1^n x_2^{-m} \mid m \in \mathbb{N} \,\}$. Notice that the maximal ideal is generated by x_2. ☆

Lecture 16

Hilbert functions

When investigating a graded algebra R, or a graded module M for that matter, which is finitely generated over a field k—a frequent activity among algebraic geometers—an extremely powerful tool is the so-called *Hilbert function* of R. This is simply the function defined as $h_R(n) = \dim R_n$; that is the dimension of the grade piece of R consisting of homogeneous elements of degree n. Or in case of a mosul $h_M(n) = \dim Mn$ The homogeneous piece M_n is a vector space over k and turns out to be of finite dimension whenever R is Noetherian and M is finitely generated over R, which makes the definition is legitimate. There is also the notion of of the Hilbert series of R, the generating series of the Hilbert function; that is, $P_M(t) = \sum_i \dim_k M_i t^i$ called the Hilbert–Poincré series of M.

The prototypical example would the the polynomial algebra $R = k[x_1, \ldots, x_r]$ in n variables. The space R_r of hogomeneous forms of degree n has the monomials of degree n as a basis, and it is is an exercise in elementary combinatorics to show that $\dim R_n$ is the binomial coefficient $\binom{n+r-1}{r-1}$.

The Hilbert functions have the virtue of behaving like a polynomial (with rational coefficients) for large values of the variable, and the coefficients of this polynomial furnish invariants of the algebra. Of course, by their very definition, these polynomials takes integral values on integers (at leat for large arguments); such polynomials are called numerical polynomials.

(which will be integral after a simple change). The dimension of a varietriy, the degrees of a projective varities, and the famous genus of a curve, are all of this type. One may also define the nultiplicity of a singular point (+++ other invariants) in this way.

The Riemann–Roch theorem iis a topological interpretation of some of this invariants.

One comes across graded rings in several corners of algebraic geometry. The most frequently met are the coordinate rings of cones. These can be of several types, the most prominent ones being cones over projective varieties, and graded rings and their Hilbert

polynomials are of paramount importance to the projective geometers

An other place is the local study of a variety and the local multiplicities. Contrary to manifolds varieties may acquire singularities, and they do not have tangent space but so-called tangent cones. Heuristically such a cone at a point P is union of all lines emanating from the point and being tangent to the variety at P. There is an algebraic intrinsic definition of the tangent cone which is a graded ring associated with any local ring containing much local information.

EXAMPLE 16.1 *Th conormal cone:* To a given ideal $\mathfrak{a}$ in a ring A one may associate a graded ring whose underlying abelain groups is

$$\mathrm{Gr}_{\mathfrak{a}} = A/\mathfrak{a} \oplus \mathfrak{a}/\mathfrak{a}^2 \oplus \ldots = \bigoplus_{i \geqslant 0} \mathfrak{a}^i/\mathfrak{a}^{i+1}$$

That there is a multiplication needs to be explained, so let $a \in \mathfrak{a}^s$ and $b \in \mathfrak{a}^t$ be elements and consider the classes $[a]$ and $[b]$ in respectively $\mathfrak{a}^s/\mathfrak{a}^{s+1}$ and $\mathfrak{a}^t/\mathfrak{a}^{t+1}$. The product ab lies in $\mathfrak{a}^{s+t}$ and the salient point is that $[ab]$ does not depend on the choice of representatives a and b: if $\mathfrak{a} \in \mathfrak{a}^{s+1}$ and $\beta \in \mathfrak{a}^{t+1}$, one finds $(a+\alpha)(b+\beta) = ab + a\beta + b\alpha + \alpha\beta$, where the three terms involving α or β belong to $\mathfrak{a}^{s+t+1}$. This gives the product of two homogeneous elements, which is extended by linearity to a product of arbitrary elements, and the ring axioms follow straightforwardly.

In the particular case of a maximal ideal $\mathfrak{m}$, the cone $\mathrm{Gr}_{\mathfrak{m}} A$ is called the *cotangent cone* of A at $\mathfrak{m}$. ☆

16.1 Numerical polynomials

numerical function (nummeriske funksjoiner)

The Hilbert functions are *a priori* so-called *numerical function*, they accept integral arguments and return integral values; *i. e.* so they are function $h\colon \mathbb{Z} \to \mathbb{Z}$. However, they are not any such functions, but turn out have the very special property of behaving as a polynomial for sufficiently large arguments: there is a polynomial $P \in \mathbb{Q}[t]$ so that $h(n) = P(n)$ for $n >> 0$. The polynomial P will necessarily* be a so-called *numerical polynomial* in that it takes integral values on the integers.

numerical polynomial (nummeriske polynomer)

(16.1) A numerical function h has a "discrete derivative" which is defined as

$$\Delta h(t) = h(t) - h(t-1).$$

*Convince yourself that this is so

It shares the property with the usual calculus-derivative that $\Delta h = 0$ is equivalent with h being constant. There is also an analogue to the integral, in that it holds true that

$$h(r) - h(s-1) = \sum_{s}^{r} \Delta h(t)$$

for all integers r and s with $r \geqslant s$. Primeordial prototypes of numerical polynomials are the *binomial coefficients* which we remind you are given as

$$\binom{t+n}{n} = \frac{(t+n)(t+n-1)\dots(t+1)}{n!}, \tag{16.1}$$

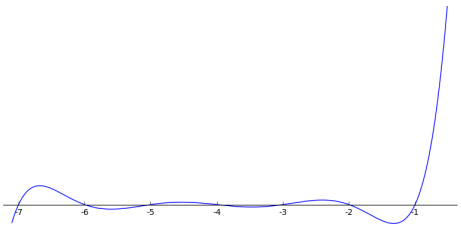
where n is any non-negative integer. It is classical that they assume integral values on the integers, and moreover, the binomial coefficient in (16.1) is of degree n, and its leading coefficient equals $1/n!$.

We already encountered the numerical polynomials in Exercise 9.8 which asked you to show that the subring $\mathrm{Int}(\mathbb{Z})$ of $\mathbb{Q}[t]$ they form is not Noetherian.

As the usual calculus derivative does, the discrete derivative lowers the degree of a polynomial by one; indeed, the Binomial Theorem yields

$$t^n - (t-1)^n = t^n - (t^n - nt^{n-1} + \dots) = nt^{n-1} + \dots.$$

We also observe that the leading coefficient picks up a factor n, so writing the leading coefficient of a degree n polynomial P as $a_0/n!$, the leading coefficient of $\Delta P(t)$ will be $a_0/(n-1)!$.



A part of the graph of $\binom{t+7}{7}$, slightly scaled to better show the behavior.

(16.2) The binomial coefficients form a basis for the ring $\mathrm{Int}(\mathbb{Z})$ of numerical polynomials which is well adapted to the discrete derivative operator Δ as the well known identity from Pascal's triangle holds true:

$$\Delta\binom{t+n}{n} = \binom{t+n-1}{n-1}. \tag{16.2}$$

Moreover, a polynomial assuming integral values at all sufficiently large integers, is a numerical polynomial; we have:

PROPOSITION 16.3 *The following two assertions hold true:*

i) A polynomial $P(t)$ is a numerical polynomial if and only if it assumes integral values for sufficiently large integers; i. e. $P(n) \in \mathbb{Z}$ when $n >> 0$;

ii) A numerical polynomial P of degree n has a development

$$P(t) = a_0\binom{t+n}{n} + a_1\binom{t+n-1}{n-1} + \dots + a_{n-1}\binom{t+1}{1} + a_n$$

where the coefficients a_i are uniquely defined integers and $a_0 \neq 0$.

PROOF: One of the implications in *i)* is tautological; attacking the other one we assume that $P(n) \in \mathbb{Z}$ for $n \geqslant s$. It follows that $\Delta P(t) = P(t) - P(t-1) \in \mathbb{Z}$ for $t \geqslant s+1$, and by induction on the degree we know that ΔP is a numerical polynomial (the claim is obviously true for polynomials of degree zero, so that the induction may start). Reintegrating, we find for any pair of integers with $r \geqslant s$ the equality

$$P(s) = P(r) + \sum_{s+1 \leqslant t \leqslant r} \Delta P(t),$$

in which the terms in the right hand sum all belong to $\mathbb{Z}$. Moreover $P(r)$ is integral when r is chosen sufficiently large, and thus we may conclude that $P(s) \in \mathbb{Z}$.

The proof of *ii)* also relies on induction on the degree, and this time the crucial observation is Pascal's identity (16.2). Obviously a polynomial of the form as in *ii)* is numerical and of the degree n. To prove the converse, assume h is a numerical polynomial of degree n; then the derivative Δh will be one of degree $n-1$, and by induction we may express Δh as

$$\Delta h(t) = \sum_{0 \leqslant i \leqslant n-1} a_i \binom{t+i}{i}.$$

Consequently the difference

$$h(t) - \sum_{0 \leqslant i \leqslant n-1} a_i \binom{t+i+1}{i+1}$$

has a vanishing derivative and is therefore constant; with a_n being this constant the claim follows. ❑

PROPOSITION 16.4 *A numerical function $h(t)$ equals a numerical polynomial of degree n for for $t >> 0$ if and only if the discrete derivative $\Delta h(t)$ equals a numerical polynomial of degree $n-1$ for $t >> 0$.*

PROOF: If $\Delta h(t)$ equals a numerical polynomial for $t >> 0$, this polynomial is of the form as in *ii)* in Proposition 16.3, and the argument just given shows that $h(t)$ is of the same form for $t >> 0$. ❑

Exercises

(16.1) Show that

$$\binom{t+n}{n} = t^n/n! + \binom{n+1}{2}/n! \cdot t^{n-1} + \dots$$

(16.2) Show that $\binom{t+n}{n}$ is the unique polynomial that vanishes at the negative integers between -1 and $-n$ and assumes the value one at zero.

★

Positively graded rings

In the introduction to this chapter we mentioned graded rings with R_0 a field, say k, and graded R-modules with each graded piece M_n of finite dimension over k; the point was that these modules possess the Hilbert functions $h_M(n) = \dim_k M_n$. The conditions may certainly be relaxed: when each homogenous piece M_n is of finite length over R_0, the module has a Hilbert function defined as $h_M(n) = \ell_{R_0}(M_n)$. The current subsection is devoted to describing large workable classes of graded rings R and graded modules M fulfilling this.

(16.5) Save for a few very special cases, graded rings with non-zero elements both of positive and of negative degree do not have very informative Hilbert Functions, if having one at all, and they are not interesting in the present context. So we shall confine our study to the so-called positively graded rings: a graded ring R is said to be *positively graded* if it has no non-zero homogeneous elements of negative degree; in other words, we require that $R_i = 0$ for $i < 0$. In that case the additive subgroup

Positively graded rings (positivt graderte ringer)

$$R_+ = \bigoplus_{i>0} R_i$$

generated by the homogeneous elements of positive degree will be an ideal in R, which is called the *irrelevant ideal**; indeed, homogeneous elements being of positive degree multiply R_+ into R_+ since $\deg x \cdot y = \deg x + \deg y \geqslant \deg y$ for all x. The graded rings $\mathrm{Gr}_{\mathfrak{a}} A$ constructed in the introduction (Example 16.1) are all positively graded; they are even generated by elements of degree one; namely by the classes of elements generating $\mathfrak{a}$.

The irrelevant ideal (det irrelevante idealet)

*This is another instance of highly delusive naming in mathematics, albeit having irrelevant in the name, these ideals are seriously relevant. The reason for this apparent malpractice is found in projective geometry; apices of cones over projective varieties are invisible to the projective geometers eye.

✲ EXERCISE 16.3 The exercise is an illustration of the restrictions one imposes on graded domains with non-zero elements of both positive degree and of negative degree when requiring the degree zero part to be a field. Let R be one of the kind and assume $R_0 = k$ is a field. Show that $R = k[w, w^{-1}]$ for some homogenous element $w \in R$. HINT: Show first that each homogeneous element is invertible. ★

(16.6) Most finiteness results about positively graded rings are rooted in the following observation which shows the relevance of the irrelevant ideal:

LEMMA 16.7 *Let R be a positively graded ring and assume that $x_1, \ldots, x_r$ are homogeneous elements that generate the irrelevant ideal R_+. Then the x_i's generate R as an algebra over R_0; that is, $R = R_0[x_1, \ldots, x_r]$.*

PROOF: We shall show, by induction on the degree of x, that each homogeneous element x of R of positive degree belongs to $R_0[x_1, \ldots, x_r]$; this will suffice since every element is the finite sum of its homogeneous constituents. So we assume that all elements of degree lower than $\deg x$ lie in $R_0[x_1, \ldots, x_r]$. Now $x \in R_+$, and hence $x = \sum a_i x_i$, for elements $a_i \in R$, and replacing the a_i's by their homogeneous components, we may assume that

a_i's are homogeneous and comply with the constraint $\deg a_i + \deg x_i = \deg x$. Thus $\deg a_i < \deg x$; by induction we infer that a_i belongs to $R_0[x_1, \ldots, x_r]$, and we are done. ❏

PROPOSITION 16.8 *Let R be a positively graded ring.*

i) Then R is Noetherian if and only if R_0 is Noetherian and the irrelevant ideal R_+ is finitely generated.

ii) If R is Noetherian, the homogenous pieces R_n are all finitely generated over R_0.

Note that the first statement appears remarkably strong; just *one* ideal, although special, being finitely generated implies that all are. The reason behind, as the proof will show, is that these rings will turn out to be quotients of polynomial rings over R_0 and so Hilbert's Basis Theorem takes effect.

PROOF: We begin with proving *i)*. If R_0 is Noetherian and R_+ is finitely generated, in fact it is generated by finitely many homogeneous elements, as just mentioned, Hilbert's Basis Theorem together with Lemma 16.7 closes the case.

So assume that R is Noetherian, the of course R_+ is finitely generated as any other ideal is, and we merely have to show that R_0 is Noetherian. The crucial observation is that for any ideal $\mathfrak{a}$ in R_0 one has $\mathfrak{a}R \cap R_0 = \mathfrak{a}$ (from which the claim follows by extending and recontracting chains). Indeed, let the sum $\sum a_i x_i$ with $x_i \in \mathfrak{a}$ and $a_i \in R$ belong to R_0. Replacing the a_i's by their homogeneous components, we may assume that the a_i's are homogeneous; furthermore, the terms whose degree is non-zero, add up to zero, and can be discarded. Hence for each i it holds that $0 = \deg a_i + \deg x_i = \deg a_i$.

The second statement follows readily: Let $x_1, \ldots, x_r$ be homogeneous generators of R_+. A monomial $x_1^{\alpha_1} \cdot \ldots \cdot x_r^{\alpha_r}$ belongs to R_n precisely when $\alpha_1 \deg x_1 + \ldots + \alpha_r \deg x_r = n$ and this equation has only finitely many solutions because $\deg x_i > 0$ (each α_i is bounded by $n/\deg x_i$). Hence there only finitely many monic monomials in R_n, and these generate R_n over R_0. ❏

The Hilbert function

The next proposition is an easy consequence of Proposition 16.8 above and describes a large class of module having a Hilbert function.

PROPOSITION 16.9 *Let R be a positively graded Noetherian ring and M a finitely generated graded R-module.*

i) All the graded pieces M_n are finitely generated over R_0.

ii) If additonally R_0 is Artinian, each M_n will be of finite length over R_0.

PROOF: The lemma is true for R itself by Proposition 16.8, hence for all shifts $R(d)$ (shifting does not alter any algebraic property merely changes degrees), and consequently for every finite direct sum $\bigoplus_i R(-d_i)$. And if M is finitely generated, it is a quotient of

such a finite sum $\bigoplus_i R(-d_i)$, and the graded pieces M_d of M of degree d are quotients of the graded pieces $\bigoplus_i R(-d_i)_d$. ❑

(16.10) So when R is Noetherian and R_0 Artinian, we my define the important invariant $h_M(n)$, the *Hilbert function*, of such finitely generated modules by putting

Hilbert function (Hilbert function)

$$h_M(n) = \ell_{R_0}(M_n).$$

It is of course a numerical function of n and it has the all important property of being an *additive* function on the category Grmod_A, which allows one to calculate it in many instances. This means that if

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0 \tag{16.3}$$

is an exact sequence in the category Grmod_R of finitely generated graded modules, then the equality

$$h_M(n) = h_{M'}(n) + h_{M''}(n)$$

holds true. It ensues from the additive character of the length. Indeed, the maps in the sequence 16.3 preserves homogeneous elements and degrees, and for each degree n the sequence therefore induces the sequence

$$0 \longrightarrow M'_n \longrightarrow M_n \longrightarrow M''_n \longrightarrow 0.$$

which is exact.

(16.11) As with any additive function, if $C_\bullet$ is a bounded complex—so that $C_n = 0$ for $|n|$ sufficiently large—of finitely generated R-modules it holds that

$$\sum_i (-1)^i h_{C_i}(t) = \sum_i (-1)^i h_{H^i(C_\bullet)}(t).$$

Examples

(16.2) Recall that the Hilbert function of the polynomial ring $R_r = k[x_0, \dots, x_r]$ over a field k is given as

$$h_{R_r}(t) = \begin{cases} 0 & \text{when } t < 0; \\ \binom{r+t}{r} & \text{when } t \geqslant 0. \end{cases}$$

The first claim is trivial. In case you don't know the second formula already: it follows immediately by induction on r and Pascal's identity (16.2). Indeed, the short exact sequence

$$0 \longrightarrow R_r \xrightarrow{x_r} R_r \longrightarrow R_{r-1} \longrightarrow 0$$

yields that $\Delta h_{R_r}(t) = h_{R_{r-1}}(t)$, which shows that the claim is true up to a constant, and evaluation at $t = 0$ ensures that the constant is zero.

(16.3) The shifts $R_r(-m)$ of the polynomial ring $R_r = k[x_0, \ldots, x_r]$ has the Hilbert function

$$h_{R_r(-m)}(t) = \begin{cases} 0 & t < m; \\ \binom{r+t-m}{r} & t \geqslant m, \end{cases}$$

since the graded pieces of $R_r(-m)$ are given as $R_r(-m)_t = (R_r)_{t-m}$. Note that $h_{R_r(-m)}(t)$ is not a polynomial, but equals one for $t \geqslant m$.

(16.4) Consider the principle ideal $(F)R_r$ in the polynomial ring $R_r = k[x_0, \ldots, x_r]$ generated by a homogeneous form F of degree m. Multiplication by F induces a homogeneous isomorphism between $R_r(-m)$ and (F) since for an element $a \in R_r(-m)_d$ it holds that $\deg a = d - m$, so consequently $\deg aF = d$.

The classical short exact sequence is therefore an exact sequence of graded modules:

$$0 \longrightarrow R_r(-m) \xrightarrow{\mu} R_r \longrightarrow R_r/(F)R_r \longrightarrow 0,$$

where the map μ is multiplication by F. Hence we find that

$$h_{R_r/(F)R_r}(t) = \begin{cases} 0 & \text{if t<0;} \\ \binom{r+t}{r} & \text{if } 0 \leqslant t < m; \\ \binom{r+t}{r} - \binom{r+t-m}{r} & \text{if } t \geqslant m. \end{cases}$$

One observes that the function $h_{R_r/(F)R_r}$is a piecewise polynomial, and that for $t \geqslant m$ it is equal to the polynomial

$$\chi_{R_r/(F)R_r}(t) = \binom{r+t}{r} - \binom{r+t-m}{r} = mt^{r-1}/r! + \ldots$$

whose degree is one less than the Krull dimension of $R_r/(F)R_r$. This illustrates the general feature of many* Hilbert functions: they are polynomials for large values of the variable of degree one less than then Krull dimension of the module. We also observe that the leading coefficient, up to the factor $1/r!$, equals the degree of F. In general this coefficient will be of the form $a/r!$ with a a natural number which, being a consequence of the standard from Proposition 16.3, holds for any numerical polynomial.

As we are soon to see, this holds for modules over graded rings which are Noetherian and generated in degree one.

Graded algebras generated in degree one over fields all have a geometric incarnation, which is a *projective varieties*, and projective varieties have a degree, which turns out to be equal to the number a.

(16.5) The next examples is slightly more elaborated. It plays an important role in the theory of plane curves where the result goes under the name of Bezout's theorem.

The givens are two homogenous polynomials f and g in $R = k[x,y,z]$ without common factors; their degrees are respectively n and m. The subsets $V(f)$ and $V(g)$ of $\mathbb{A}^3$ are cones over projective plane curves, and we intend to investigate their intersection. It is represented by the cone $V(f,g)$ over the common zeros of f and g, and whose algebraic incarnation is the homogeneous ring $R/(f,g)$. The two curves should intersect in finitely many points, so one suspects $V(f,g)$ to be finite; that is, the cone $R/(f,g)$ should be one dimensional. Moreover, one believes that "the number of common zeros" should be the product nm of the degrees; or translated into properties of the Hilbert function, it should holds that $h_{R/(f,g)}(t)$ is constant for $t > 0$ (*i. e.* of degree one less than $\dim R/(f,g)$) and the constant value should equal nm. And indeed, we shall prove that the Hilbert function $h_{R/(f,g)}(t) = nm$ for $t \geqslant n+m$.

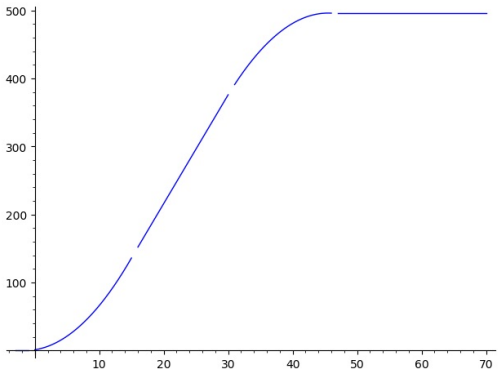
Then crucial point is the Koszul complex which we introduced in Example 5.9 on page 145, and which under current cicumstances furnishes a free resolution of $R/(f,g)$ shaped like:

$$0 \longrightarrow R(-n-m) \xrightarrow{d_2} R(-n) \oplus R(-m) \xrightarrow{d_1} R \longrightarrow R/(f,g) \longrightarrow 0 \qquad (16.4)$$

where $d_2(a) = (ag, -af)$ and $d_1(a,b) = af + bg$. Applying the additivity of the Hilbert function, this exact sequenecs followed by a small trivial computation one finds that for $t \geqslant n+m$ it holds that

$$h_{R/(f,g)}(t) = \binom{t+2}{2} - \binom{t+2-n}{2} - \binom{t+2-m}{2} + \binom{t+2-n-m}{2} = nm.$$

For values of t smaller than $n+m$ not all four binomial coefficients do appear in the expression for $h_{R/(f,g)}$; remember that $h_{R(-n)}(t) = 0$ for $t < n$. Which ones do, depends on the size of t compared to 0, n, m and $n+m$. In the figure below the graph of $h_{R/(f,g)}(t)$ is sketched for $n = 16$ and $m = 31$; It is a piecewise polynomial with two infinite constant sectors and two quadratic parts separated a linear sector.



☆

The Hilbert polynomial

The main result in this paragraph is that the Hilbert function associated with a graded module M finite over a Noetherian ring A generated in degree one and with A_0 Artinian, is equal to a polynomial for large values of the variables. This polynomial is called the *Hilbert polynomial* of M. It is a numerical polynomial whose coefficients are important invariants of the graded module M. We shall not say much about these invariants, but contend ourself to show that the degree equals the dimension of the support of M.

The most common application of the Hilbert polynomial is in cases when A is a polynomial ring over a field with all the variables being of degree one. The hypothesis that A be generated in degree one, is an essential one; a stupid example is obtained by assigning the degree two to the variable x in $A = k[x]$. Then the Hilbert function $h_A(n)$ vanishes for all odd numbers and therefore cannot be equal to a polynomial for large n. (16.12) Recall that the dimension $\dim M$ of a module M finite over a Noetherian ring A by definition equals the dimension of its support; that is, it is equal to $\dim A/\operatorname{Ann} M$. Moreover, $\dim M = 0$ if and only if M is a module of finite length. Here comes the result:

THEOREM 16.13 (HILBERT–SERRE) *Let R be a Noetherian graded ring with R_0 being Artinian and M a finitely generated graded A-module Assume that R is generated in degree one. Then the following hold true:*

i) The Hilbert function $h_M(t)$ equals a polynomial $\chi_M(t)$ for $t >> 0$;
ii) The degree r of χ_M is equal to $\dim M - 1$, and its leading coefficient is of shape $d/r!$ with $d \in \mathbb{N}$.

It is crucial that R be generated in degree one; a stupid example with $R = k[t]$ where t is given the degree two. Then $R_n = 0$ when n is odd, so the Hilbert function has infinitely many zeros and can not equal a polynomial for large integers.

PROOF: The proof goes by induction on $\dim M$. The induction can begin because $\dim M = 0$ implies that the module M is of finite length and merely has finitely many non-vanishing homogeneous parts, so that $h_M(t) = 0$ for $t >> 0$.

Each finitely generated module M has, as the Structure theorem for graded modules tells us (Proposition 10.41 on page 279), a finite ascending chain M_i of graded submodules whose subquotients are of the form $A/\mathfrak{p}_i$ with the $\mathfrak{p}_i$'s being homogeneous prime ideals. Taking the grading into account we arrive at a series of exact sequences in GrMod_A, one for each $0 \leqslant i \leqslant r$, which all are shaped like:

$$0 \longrightarrow M_{i-1} \longrightarrow M_i \longrightarrow A/\mathfrak{p}_i(m_i) \longrightarrow 0.$$

Moreover, it holds that $M_r = M$ and $M_0 = 0$.

A successive application of the additivity of the Hilbert function yields the equality

$$h_M(t) = \sum_{1 \leqslant i \leqslant r} h_{A/\mathfrak{p}_i}(t + m_i).$$

Since the the dimension of the support Supp M equals the maximum of the dimensions $\dim A/\mathfrak{p}_i$ of the subquotients, and since shifting a module merely affects the Hilbert functions by translating the variable, it will suffice to show the proposition for quotients $A/\mathfrak{p}$ with $\mathfrak{p}$ being a homogeneous prime ideal. To that end, pick any $x \in A_1$ not lying in $\mathfrak{p}$ (we may safely assume that $\dim A/\mathfrak{p} > 0$ so that $\mathfrak{p} \subsetneq A_+$) and form the exact sequence

$$0 \longrightarrow A/\mathfrak{p}(-1) \longrightarrow A/\mathfrak{p} \longrightarrow A/\mathfrak{p} + (x) \longrightarrow 0.$$

According to Krull's Principal Ideal Theorem it holds that $\dim A/\mathfrak{p} + (x) = \dim M - 1$, induction applies to $A/\mathfrak{p} + (x)$ and $\Delta h_{A/\mathfrak{p}}(t) = h_{A/\mathfrak{p}+(x)}(t)$ equals a polynomial of degree $\dim M - 1$ for $t >> 0$. Evoking Proposition 16.3 on page 405 we infer that $h_{A/\mathfrak{p}}(t)$ is a polynomial of degree $\dim M$ for large values of t. ❑

LEMMA 16.14 *Let R be a Noetherian graded ring with R_0 Artinian which is generated by r elements of degree one. Then the Hilbert polynomial $\chi_A(t)$ is of degree less than r unless R is isomorphic to the polynomial ring in r variables over R_0, in which case* $\deg \chi_A(t) = r$.

PROOF: Assume that the elements $x_1, \ldots, x_r$ are elements od R_1 that generates R over R_0. Let $X_1, \ldots, X_r$ be variables and define a map $\phi\colon R_0[X_1, \ldots, X_s] \to R$ by sending X_i to x_i. This is a map of graded rings which at the outset is surjective. It follows that $h_R(t) \leqslant \binom{t+r}{r}$ for all t.

If the map ϕ is not injective, we may chose a non-zero homogeneous polynomial F from its kernel. Then ϕ factors through $B = R_0[X_1, \ldots, X_r]/(F)$, and consequently $h_R(t) \leqslant h_B(t)$ for all t. According to Proposition 11.9 on page 293 it holds that $\dim B \leqslant r - 1$, and by Proposition 16.13 above $\chi_B(t)$ is of degree at most $r - 1$, so that $\deg \chi_R(t) \leqslant r - 1$. ❑

This is the so-called Roman surface, a realisation with self intersections of the real avatar of the Veronese surface in $\mathbb{R}^3$. Modern technology not only produce beautiful pictures of surfaces, but with 3D-printers also intriguing models are made. This one find at Mathematical Art Galleries

Examples

(16.6) Given a natural number d. In this example we consider the subalgebra $A = k[u^d, u^d v, \dots, uv^{d-1}, v^d]$ of the polynomial ring $R = k[u, v]$ generated by all monomials of degree d. It is a graded subalgebra, and one easily verifies that A decomposes as $A = \bigoplus_i R_{id}$; that is, as the sum of the homogeneous pieces of R of degree a multiple of d. Now, changing the degrees of the elements in A by factoring out d, the algebra A will be generated in degree one, and by the decomposition of A above we find $\chi_A(t) = \chi_R(dt) = dt + 1$.

More generally, let R be the polynomial ring $R = k[u_0, \dots, u_n]$ and consider the subalgebra* A of R generated by all the monomials of degree d; *i. e.* those shaped like $u^\alpha = u^{\alpha_0} \cdot \dots \cdot u_r^{\alpha_r}$ with $\sum_i \alpha_i = d$. Just as in the previous case, the decomposition of A into homogeneous pieces appears as $A = \bigoplus_i R_{id}$, form which ensues the identity

*These algebras are called Veronese algebras, since they are the homogeneous coordinate rings of the so-call Veronese varieties.

$$\chi_A(t) = \chi_R(dt) = \binom{dt+n}{n} = d^n t^n / n! + \dots .$$

We conclude that $\dim A = n + 1$ and that the degree of the corresponding Veronese variety is d^n.

(16.7) Let us examine the algebra $R = k[x, y, x]$ with constituting relations $x^2 - y^2 = x^2 - z^2 = 0$ with standard grating. The plolynomial ring $k[, Y, Z]$ is factorial and the polynomials $X^2 - Y^2$ and $X^2 - Z^2$ are without common factors, so by Example 16.5 the Koszul complex is exact and the Hilbert polynomial is constant and equal to the product of the two degrees; that is, it is equal to four.

It is worthwhile examining the situation more closely. The relations start to be visible for the Hilbert functtion In degree two: since all the squares of the variable are equal, R_2 is generated by x^2, xy, xz, yz, and more generally, the same reasoning gives that $x^n, x^{n-1}y, x^{n-1}z, x^{n-2}yz$ generate R_n.

This immediately gives that $h_R(t) \leqslant 4$ for $t \geqslant 2$, but to prove equality with this approach, one would also need to show they are linearly independent, which amounts to seeing that $\mathfrak{m} = (x, y, z)$ is not an associated prime. Indeed, if $0 = \alpha x^n + \beta x^{n-1}y + \gamma x^{n-1}z + \delta x^{n-2}yz = x^{n-2}(\alpha x^2 + \beta xy + \gamma xz + \delta yz)$, there will be an element killed by a high power of x and hence by high powers of y and z as well. It is not to hard to give an *ad hoc* argument for this not being the case, but with the Koszul complex it comes for free. In fact, one may turn the argument around, and knowing the Hilbert function $h_R(t)$, conclude that $\mathfrak{m}$ is not associated. The primary decomposition of $(X^2 - Y^2, X^2 - Z^2)$ is thence

$$\begin{aligned}&(X^2 - Y^2, X^2 - Z^2) = \\ &= (X - Y, X - Z) \cap (X - Y, X + Z) \cap (X + Y, Y - Z) \cap (X + Y, X + Z).\end{aligned} \tag{16.5}$$

Localized at a prime ideal $\mathfrak{p}$ not containing two of the linear factors, the inclusion (16.5) becomes an equality, and if three of them belongs to $\mathfrak{p}$, one easily checks that $\mathfrak{p} = (x, y, z)$.

(16.8) *A determinantal variety and the Hilbert–Burch complex:* Our next example is about the ideal generated by the maximal minors of the generic 3×2-matrix

$$\begin{pmatrix} x_{00} & x_{10} \\ x_{01} & x_{11} \\ x_{02} & x_{12} \end{pmatrix}$$

where the x_{ij}'s are variables, so that M has entries from the polynomial ring $R = k[x_{ij} | 0 \leqslant i, j \leqslant 2]$. The maximal minors of M are the 2×2-minors, and the ideal we shall explore will be

$$\mathfrak{a} = \big(x_{01}x_{12} - x_{02}x_{11}, x_{02}x_{10} - x_{00}x_{12}, x_{00}x_{11} - x_{01}x_{10}\big). \tag{16.6}$$

When k is algebraically closed, the closed points of the space $\mathbb{A}^6_k = \operatorname{Spec} k[x_{ij}]$ parametrize the 3×2-matrices with entries from k, and those lying in $V(\mathfrak{a})$ constitue the locus where the matrix drops rank. For general k the same applies to the set of k-points $\mathbb{A}_6(k)$, and quite generally any 3×2-matrix $N = (b_{ij})$ with coefficients in a k-algebra B is obtained from M by changing base by the map $R \to B$, that sends x_{ij} to b_{ij}.

We intend to compute the Hilbert function of $A = R/\mathfrak{a}$, which gives us the opportunity to introduce the *Hilbert–Burch complex*. This is the complex

$$C_\bullet : \ 0 \longrightarrow 2R(-3) \xrightarrow{M} 3R(-2) \xrightarrow{D} R$$

where D is the 3×1-matrix

$$D = \big(x_{01}x_{12} - x_{02}x_{11}, -(x_{00}x_{12} - x_{02}x_{11}), x_{00}x_{11} - x_{01}x_{10}\big).$$

Obviously the images of D is equal to our ideal $\mathfrak{a}$, and that $D \cdot M = 0$ is seen by successively enlarging M by the each row and then expanding the determinant. For instance, expansion along the first row gives

$$0 = \det \begin{pmatrix} x_{00} & x_{01} & x_{02} \\ x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \end{pmatrix} = x_{00}(x_{01}x_{12} - x_{02}x_{11}) - x_{01}(x_{00}x_{12} - x_{02}x_{11}) + x_{02}(x_{00}x_{11} - x_{01}x_{10}),$$

and in a similar way, expanding M by adding t the other row and expanding along it, gives the second equation required for $D \cdot M$=0 to hold true. We have by that established

that 16.8 is a complex. Most often one the complex is displayed with the cokernel of D, wich equals A, included:

$$0 \longrightarrow 2R(-3) \xrightarrow{M} 3R(-2) \xrightarrow{D} R \longrightarrow A \longrightarrow 0. \tag{16.7}$$

The salient point is that the sequence (16.7) is exact. Given this, we easily find the Hilbert polynomial:

$$\chi_A(t) = \binom{t+5}{5} - 3\binom{t+3}{5} + 2\binom{t+2}{5} = \frac{1}{2}t^3 + 2t^2 + \frac{5}{2}t + 1$$

or expanded in the basis of the binomial polynomials

$$\chi_A(t) = 3\binom{t+3}{3} - 2\binom{t+2}{2}.$$

Note that this one of rare cases that Hilbert polynomial and the Hilbert function coincide for all $t \geqslant 0$.

LEMMA 16.15 *The Hilbert–Burch complex resolves A; that is, the sequence 16.6 is exact.*

PROOF: That M is injective follows by passing to the quotient field K of R; Then the minors of M will be invertible, and $\ker M \otimes_R K = 0$. But the kernel of M is torsion free being contained in a free R-module, hence it vanishes.

So, the only hot spot is the middle homology $H^1(C_\bullet) = \ker D / \operatorname{im} M$. The proof it vanishes has two steps. Firstly, we contend that the support of $H^1 C_\bullet$ is contained* in $V(x_{00})$; that is, $H^1(C_\bullet) \otimes_R R[x_{00}^{-1}] = 0$. So assume that $(a, b, c) \in 3R$ satisfies the relation $D \cdot (a, b, c)^t = 0$; a totally elementary manipulation gives

$$(x_{00}c - x_{02}a)(x_{11}x_{00} - x_{10}x_{01}) + (x_{00}b + x_{01}a)(x_{12}x_{00} - x_{10}x_{02}) = 0$$

and hence since the two involved minors are irreducible and the polynomial ring is a UFD, there is a polynomial $d \in R$ with

$$\begin{aligned} x_{00}b &= -d(x_{11}x_{00} - x_{10}x_{01}) + x_{01}a \\ x_{00}c &= d(x_{12}x_{00} - x_{10}x_{02}) + x_{02}a \end{aligned}$$

Then one finds that

$$\begin{pmatrix} x_{00} & x_{10} \\ x_{01} & x_{11} \\ x_{02} & x_1 \end{pmatrix} \begin{pmatrix} -dx_{10} + a \\ dx_{00} \end{pmatrix} = \begin{pmatrix} x_{00}a \\ x_{00}b \\ x_{00}c \end{pmatrix}$$

and since x_{00} is inveritble on $R_{x_{00}}$, we are done.

The use of x_{00} is dictated by pure convenience. By elementary row and column operations each of the variables, in fact any linear combination of the variables, may be brought to the upper left corner of M, so they are all on equal footing.

This shows that $H^1(C_\bullet)\otimes_R R[x_{00}^{-1}] = 0$, and consequently $H^1(C_\bullet)$ is killed by a high power of x_{00}. The sequence

$$0 \longrightarrow C_\bullet \xrightarrow{x_{00}} C_\bullet \longrightarrow C_\bullet/x_{00}C_\bullet \longrightarrow 0$$

induces a long exact sequence of homology modules, the part of which that interests us is

$$H^2(C_\bullet/x_{00}C_\bullet) \longrightarrow H^1(C_\bullet) \xrightarrow{x_{00}} H^1(C_\bullet)$$

The point is that $H^2(C_\bullet/x_{00}C_\bullet) = 0$, so that multiplication by x_{00} is injective But we already observed that a power of x_{00} kills $H^1(C_\bullet)$, and consequently $H^1(C_\bullet) = 0$. To see that $H^2(C_\bullet/x_{00}C_\bullet) = 0$ observe that killing x_{00} gives us a the matrix

$$M' = \begin{pmatrix} 0 & x_{01} \\ x_{01} & x_{11} \\ x_{02} & x_{12} \end{pmatrix}$$

with coefficients in the polynomial ring $R' = R/x_{00}R = k[x_{01}, x_{02}, x_{10}, x_{11}, x_{12}]$, and M' is is generically of rank 2, hence the map $M'\colon 2R'(-3) \to 3R'(-2)$ is injective. ❑

☆

✲ EXERCISE 16.4 The Hilbert–Burch complex $C_\bullet$ is a meaningful construct for any 3×2-matrix $M = (a_{ij})$ with coefficient from any ring R. As in the example, let $\mathfrak{a}$ be the ideal generated by the 2×2-minors of M and let $A = R/\mathfrak{a}$. Mimicking relevant parts of the approach in the example, show the following claims:

a) If $\mathfrak{a} = R$, then the Hilbert–Burch complex is exact.
b) If $\mathfrak{a}$ contains a non-zero divisor x, then $H^2(C_\bullet) = 0$.
c) If $\mathfrak{a}$ contains a regular sequence of length two, then the Hilbert–Burch complex is a resolution of $A = R/\mathfrak{a}$.

★

EXERCISE 16.5 Let $\mathfrak{a}$ be the ideal $\mathfrak{a} = (x_1x_3 - x_2^2, x_0x_3 - x_1x_2, x_0x_2 - x_1^2)$ in the polynomial ring $k[x_0, x_1, x_2, x_3]$ and let $A = R/\mathfrak{a}$ equipped with the standard grading. Determine the Hilbert polynomial of A. The ring A is the coordinate ring of the cone over the so-called *twisted cubic curve*. ★

The Hilbert-Poincaré series

There is another way of encoding the sizes of the graded pieces of graded modules by forming their *generating series*, the so called *Hilbert–Poincaré series* Contrary to the approach relying on numerical polynomials, which requires the graded ring to be generated in degree one, the Hilbert–Poincaré series is usefull for (Noetherian) postively graded rings without further limitations on the degree of teh genators.

(16.16) So let R a positively graded Noetherian ring with R_0 artinian and let M be a finitely generated R-module. The graded pieces M_n are the of finite length over R_0, and

we may form the formal Laurent series

$$P(M,t) = \sum_{n\in\mathbb{Z}} \ell_{R_0}(M_n)t^d.$$

Hilbert–Poincaré series (Hilbert–Poincaré rekken)

As M finiteenerted over R, which is positively graded, M will be bounded below, and the series has only finitely many non-zero terms with negative exponents, in other words, it is a Laurent series; the *Hilbert–Poincaré series* of M.

(16.17) The Hilbert–Poincaré series are clearly additive invariants; indeed, an exact sequence

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

of graded M-modules is exact degree by degree, and hence $P(M,t) = P(M',t) + P(M'',t)$. Moreover, they are well behaved with respect to the shift operators:

LEMMA 16.18 $P(M(m),t) = t^{-m}P(M,t)$

PROOF: It holds true that $M(m)_n = M_{n+m}$ so that $P(M(m),t) = \sum \ell_{R_0}(M(m)_n)t^n = \sum_n \ell_{R_0}(M_{n+m})t^n$, and changing the summation variable by putting $n' = n+m$ one obtains $\sum \ell_{R_0}(M_{n'})t^{n'-m} = P(M,t)t^{-m}$. ❑

(16.19) The next theorem describes the over all structure of the Hilbert–Poincaré series of a finitely generated graded module over Noetherian and positively graded ring. They turn out to be rational function with poles at certain roots of unity determined by the generators of the graded ring.

THEOREM 16.20 *Let R be a graded ring with R_0 being Artinian. Assume that R is generaed over R_0 by elements $x_1,\dots,x_r$ whose degrees are $d_1,\dots,d_r$, all being positive. Let M be a finite R-module. Then the Hilbert–Poincare series $P(M,t)$ of M is a rational function of the type*

$$P(M,t) = f(M,t)/t^m \prod_i (1-t^{d_i}) \tag{16.8}$$

where $f(M,t)$ is a polynomial with integral coefficients. If M is positively generated as well, then $m = 0$.

PROOF: The proof goes by induction on the number of generaors of R_+ and relies on Structure of Graded Modules (Theorem 10.41 on page 279), which allows us to reduce to the case that $M = R/\mathfrak{p}$ for a homogenous prime ideal $\mathfrak{p}$. Indeed, the class of rational functions appearing on the right hand side of in (16.8) is closed under addition and invariant under multiplication by powers of t (both positive and negative) and the Hilbert–Poincaré series $P(M,t)$ is an additive invariant.

If $R_+ \subseteq \mathfrak{p}$, we are done, since then $R/\mathfrak{p}$ is Artinian being contained in R_0 and the Hilbert–Poincaré series is in fact a polynomial. If $\mathfrak{p}$ is not contained i R_+, one of the

generators, say x_i, does not belong to $\mathfrak{p}$, and one may form the following sequence which is exact in GrMod_R:

$$0 \longrightarrow R/\mathfrak{p}(-d_i) \xrightarrow{x_i} R/\mathfrak{p} \longrightarrow R/\mathfrak{p} + (x_i) \longrightarrow 0. \tag{16.9}$$

Now, by Lemma 16.18 above $P(R/\mathfrak{p}(-d_i)) = t^{d_i}P(R/\mathfrak{p})$, and hence it ensues form (16.9) that

$$P(R/\mathfrak{p} + (x_i), t) = P(R/\mathfrak{q}, t) - P(R/\mathfrak{p}(-s_i), t) = P(R/\mathfrak{q}, t) - t^{d_i}P(R/\mathfrak{p}, t).$$

Consequently we find that

$$P(R/\mathfrak{p}, t) = P(R/\mathfrak{p} + (x_i), t)/(1 - t^{d_i}),$$

and we are done since the right hand side by induction on the number of generators is of the desired shape. ❏

As a corollaries of the proof we have

COROLLARY 16.21 *Let $k[x_1, \ldots, x_r]$ be a polynomial ring given a grading by letting x_i be of degree d_i. Then $P(R, t) = \prod(1 - t^{d_i})^{-1}$.*

PROOF: Induction on r, successively killing each x_i. ❏

THEOREM 16.22 *Let R be positively graded Noetherian ring with R_0 Artinian and M a graded module finite over R_0. Then the pole order of $P(M, t)$ at $t = 1$ equals the dimension* $\dim M$*, and the residue of $P(M, t)$ at $t = 1$ is positive rational number unless M is the zero module.*

If $P(M, t)$ happens to be a polynomial, in other words if M is Artinian, then the residue should be interpreted as the value $P(M, 1)$, that is the total length of M. It is never zero unless M is the zero module.

PROOF: The proof follows the same pattern as the proof of Theorem 16.20. By the Structure Theorem 10.41 on page 279 it suffices to verify the claims for the quotients $R/\mathfrak{p}$ where $\mathfrak{p}$ is a homogeneous prime; indeed, if $R/\mathfrak{p}_i(m_i)$ with $1 \leqslant i \leqslant r$ are the arising subquotients, it holds that

$$P(M, t) = \sum_{1 \leqslant i \leqslant r} P(R/\mathfrak{p}_i, t)t^{-m_i}.$$

The residues at 1 of all the $P(R/\mathfrak{p}_i, t)$ being positive, no cancellation takes place, and the pole order of the sum equals the maximum of the pole orders of the summands. By the quotient case this equals $\max \dim R/\mathfrak{p}_i$; but this maximum is also equal to the dimension $\dim M$.

As to the case of the quotients $R/\mathfrak{p}$, one infers from (16.9) that the pole order of the Hilbert–Poincare series goes up by one when one passes from $R/\mathfrak{p} + (x_i)$ to

$R/\mathfrak{p}$, as does the dimension by Krull's Principal Ideal Theorem. Hence by induction the dimension and the pole order agree. Keeping the classical identity $(1-t^d) = (1-t)(1+t+\ldots+t^{d-1})$ in mind one also infers from (16.9) that

$$\mathrm{Res}_{t=1}P(R/\mathfrak{p},t) = \mathrm{Res}_{t=1}P(R/\mathfrak{p}+(x_i))d_i^{-1},$$

and the claim about the residues being positive rational numbers ensues. ❏

The residue of $P(M,t)$ at 1 is not always an integer, but it follows from the proof that it is a rational number belonging to $\mathbb{Z}[d_1^{-1},\ldots,d_r^{-1}]$; that is, only primes diving one of the degrees appear in its denominator. When R is generated in degree one so that all d_i's equal one, the residue will be an integer, and coincides with the degree of $V(\mathfrak{p})$ in case $M = R/\mathfrak{p}$. In the general case some would call the residue the *orbifold degree* of $V(\mathfrak{p})$.

Examples

EXAMPLE 16.9 Let $R = k[u,v]$ and $A = k[uv,u^n,v^n]$. Then $A = k[x,y,z]/(z^n - xy)$. ☆

EXAMPLE 16.10 Consider the surface with equation $x^2+y^3+z^5 = 0$ in $\mathbb{A}^3$. It is another of the so-called du Val singularities which goes under the name of the E_8-singularity (there is an E and an E_7 singularities as well, but no E_9 or E_ν for $\nu \geqslant 9$). Giving the degrees to the variables in t $\deg x = 15$, $\deg y = 10$ and $\deg z = 6$ the polynomial $x^2+y^3+z^5$ becomes homogeneous of degree 30; and from the exact sequence

$$0 \longrightarrow R[-30] \longrightarrow R \longrightarrow A \longrightarrow 0$$

we find $P(A,t) = (1-t^{15})^{-1}(1-t^{10})^{-1}(1-t^6)^{-1}(1-t^{30})$

The Laurent expantion of $P(A,t)$ to the third round $t=1$ looks like

$$\frac{1}{30}(x-1)^{-2} + \frac{1}{60}(x-1)^{-1} + \frac{269}{2^3\cdot 3^2\cdot 5} - \frac{269}{2^4\cdot 3^2\cdot 5}(x-1) + \frac{611}{2^5\cdot 3^3\cdot 5^2}(x-1)^2 + \frac{231}{2^6\cdot 5^2}(x-1)^3 + \ldots$$

and we notice that the pole order is two corresponding to the dimension of A being two and the orbifold degree is $1/30$. With modern computational software such series might feed the numerological inclinations you might have; for instance, the 50^{th} terms of the series is

$$\frac{\cdot 383\cdot 6173996327\cdot 927594632681286257477}{2^{53}\cdot 3^{27}\cdot 5^{14}}(x-1)^{50}.$$

There are nice primes appearing, and immediately a conjecture about the power of two in the denominator surfaces. It is of course also interesting to expand P in powers of t: ☆

EXERCISE 16.6 Let R be a Noetherian graded ring. Show that $N \subseteq M$ is an R_0-submodule then $RM \cap M = N$. Conclude that if R is Noetherian, then the all R_0-submodules

$N_{r,s} = \bigoplus_{r\leqslant n\leqslant s} R_n$ are Noetherian R_0-modules, i particular they are finitely generated R_0-modules. ★

EXERCISE 16.7 Show that $R^+ = \bigoplus_{n\geqslant 0} R_n$ and $R^- = \bigoplus_{n\leqslant 0} R_n$ are graded subrings of R. Show that R is Noetherian if ad only if R_0 is Noetherian and R is a finitely generted R_0-algebra. ★

16.2 *Multiplicities and Hilbert–Samuel functions*

Filtrations

A *filtration* of a ring A is a descending chain $\{A_i\}$ of ideals

filtration of rings (filtrasjon av ringer)

$$\ldots \subseteq A_{i+1} \subseteq A_i \subseteq \ldots \subseteq A_0 = A$$

with $A_0 = A$ such that $A_iA_j \subseteq A_{i+j}$ for all i and j. The only filtrations on rings we shall meet in this course are the so-called $\mathfrak{a}$-*adic* ones; they are shaped like $\{\mathfrak{a}^i\}$ for an ideal $\mathfrak{a}$ in A. These adic filtrations include the two trivial filtrations, one with $A_i = A$ for all i and one with $A_0 = A$ and $A_i = 0$ for $i > 0$.

A *filtration* of an A-module M compatible with a given filtration of A is just a descending chain $\mathcal{M} = \{M_i\}$ in M *Filtration (filtrasjoner)*

$$\ldots \subseteq M_{i+1} \subseteq M_i \subseteq \ldots \subseteq M_0 = M$$

satisfying the compatibility requirement $A_iM_j \subseteq M_{i+j}$. If $\mathcal{M}$ is compatible with the $\mathfrak{a}$-adic filtration of A; that is, if $\mathfrak{a}^iM_j \subseteq M_{i+j}$ for all i and j, it is said to be an $\mathfrak{a}$-*filtration*. A self-propelled induction yields that for $\mathcal{M}$ to be $\mathfrak{a}$-*filtration* it suffices that $\mathfrak{a}M_i \subseteq M_{i+1}$ for all i. In case equality eventually reigns; that is, when it holds that $\mathfrak{a}^iM_j = M_{i+j}$ for all i and $j >> 0$, the filtration is said to be $\mathfrak{a}$-*stable*. Again by a straightforward induction, this is equivalent to $\mathfrak{a}M_i = M_{i+1}$ for $i >> 0$.

$\mathfrak{a}$-filtrations ($\mathfrak{a}$-filtrasjoner)

$\mathfrak{a}$-stable filtrations ($\mathfrak{a}$-stabile filtrasjoner)

(16.23) A filtration $\{M_i\}$ on an A-module M induces one on every submodule N which simply is given as $N_i = M_i \cap N$. There is also induced a filtration on the quotient module M/N whose terms are the images of the M_i's, in other words, $M_i + N/N$. In both cases the compatibility with the filtration on A is obvious. If $\{M_i\}$ is $\mathfrak{a}$-stable, the induced filtration $M_i \cap N$ is $\mathfrak{a}$-stable as well. Indeed, $\mathfrak{a}^{i+r}M_i \cap N = \mathfrak{a}^i(\mathfrak{a}^rM_r \cap N)$.

(16.24) To any filtration $\mathcal{M}$ is associated the module $\mathrm{Gr}_{\mathcal{M}} \bigoplus_{i\geqslant 0} M_i/M_{i+1}$.

If $\mathfrak{a}$ is finitely generated, say by $a_1, \ldots, a_s$, the ring $\mathrm{Gr}_{\mathfrak{a}}(A)$ will generated by the classes $[a_i]_1$ of the a_i's in $\mathfrak{a}/\mathfrak{a}^2$, so $\mathrm{Gr}_{\mathfrak{a}}$ is finitely generated in degree one over $A/\mathfrak{a}$.

The arche-type of an $\mathfrak{a}$-adic filtration is the one whose terms are $\mathfrak{a}^nM$.

16.3 *Graded rings and modules*

Recall from Section 2.8 on page 59 that a graded ring A is a ring whose underlying abelian group decomposes as a direct sum $A = \bigoplus_i A_i$ of subgroups where the sum

extends over all integers. Every element $x \in A$ can thus be decomposed as a sum $x = \sum_i x_i$ with each x_i being homogeneous of degree i; that is, belonging to A_i, and only finitely many of them being non-zero. The non-zero x_i's are called the homogenous components of x of degree i, and the degree of x is the degree of the homogeneous component of x of highest degree.

The decomposition $A = \bigoplus_i A_i$ must be compatible with the multiplication in A in sense that

$$A_i A_j \subseteq A_{i+j}$$

for all i and j; in other words, if x and y are homogeneous of degree i and j respectively, their product is homogeneous of degree $i + j$. In particular, the graded piece A_0 of degree zero will be a subring of A, and each A_i is a module over A_0.

EXAMPLE 16.11 The archetype of a graded ring is of course the polynomial ring $R = R_0[x_0, \ldots, x_n]$ over some ring R_0 with the standard grading, the one giving all the variable x_i the degree one. The homogenous part R_i degree i has an R_0-basis which consists of the monomials of degree i, and hence is a free R_0-module of rank $\binom{n+i}{i}$. In most examples occurring in algebraic geometry the ring R_0 will be a field.

Other examples, also omnipresent in algebraic geometry, are the quotients $R/\mathfrak{a}$ where $\mathfrak{a}$ is a homogeneous ideal. It holds true that $\mathfrak{a}_i = \mathfrak{a} \cap R_i$ and hence the induced decomposition of $R/\mathfrak{a}$ into homogeneous pieces is $R/\mathfrak{a} = \bigoplus_i (R_i/\mathfrak{a}_i)$. ☆

Associated graded rings

With any ideal in a ring A one associates a graded ring $\operatorname{Gr}_{\mathfrak{a}}(A)$ by the construction

$$\operatorname{Gr}_{\mathfrak{a}} A = \bigoplus_{i \geqslant 0} \mathfrak{a}^i/\mathfrak{a}^{i+1} = A/\mathfrak{a} \oplus \mathfrak{a}/\mathfrak{a}^2 \oplus \ldots.$$

The ring structure is defined in the straightforward manner: If $[a]_i$ and $[b]_j$ are classes in $\mathfrak{a}^i/\mathfrak{a}^{i+1}$ and $\mathfrak{a}^j/\mathfrak{a}^{j+1}$ of elements $a \in \mathfrak{a}^i$ and $b \in \mathfrak{a}^j$ respectively, their product equals the class $[ab]_{i+j}$ in $\mathfrak{a}^{i+j}/\mathfrak{a}^{i+j+1}$, which is a legitimate definition since obviously $ab \in \mathfrak{a}^{i+j}$, and altering a or b by an element from respectively $\mathfrak{a}^{i+1}$ or $\mathfrak{a}^{j+1}$ changes ab by an element in $\mathfrak{a}^{i+j+1}$ so that the class $[ab]_{i+j}$ is well defined. The axioms for a graded ring are gotten almost for free, and any serious student should check them. The ring $\operatorname{Gr}_{\mathfrak{a}} A$ is sometimes called the *normal cone* of Spec A along $V(\mathfrak{a})$, or if $\mathfrak{m}$ is maximal, the *tangent cone* of Spec A at $\mathfrak{m}$.

Normal cones (normalkjegler)

Tangent cones (tangentkjegler)

EXAMPLE 16.12 To justify the name tangent cone, let us consider the example of a simple double point in the plane located at the origin. It has two tangents, and the only reasonable interpretation of the tangent cone is the union of the two lines. To see this is the case, let $A = k[x, y]$ with constituting relation $y^2 = x^2(x + 1)$, and let $\mathfrak{m}$ be the maximal ideal (x, y). We contend that the associated graded algebra $\operatorname{Gr}_{\mathfrak{m}} A$ is

isomorphic to $k[u,v]$ with constituting relation $u^2 - v^2 = 0$, where u denotes the class of x in $\mathfrak{m}/\mathfrak{m}^2$ and v that of y. Geometrically, the spectrum of $\mathrm{Gr}_\mathfrak{m} A$ equals the union of the two lines $y = \pm x$ in the plane which is what we wanted; at least this holds true when k is not of characteristic two. In case it is, $u + v$ is a nilpotent element of $\mathrm{Gr}_\mathfrak{m} A$, the square equals zero, and the spectrum represents a double line.

Let X and Y be variables. The assignments $X \mapsto u$ and $U \mapsto v$ induce a surjection $k[X,Y]$ onto $\mathrm{Gr}_\mathfrak{m} A$. To figure out what the kernel is, observe that a homogeneous form $F_d(X,Y)$ of degree d maps to elements $F_d(x,y)$ in A and $F_d(u,v)$ in $\mathrm{Gr}_\mathfrak{m} A$. The latter vanishes precisely when the former; that is $F_d(x,y)$, belongs to $\mathfrak{m}^{d+1}$, and then there is a polynomial $G(X,Y)$ of degree exceeding d so that $F_d(x,y) - G(x,y) = 0$. Hence

$$F_d(X,Y) - G(X,Y) = H(X,Y)(Y^2 - X^2 + X^3)$$

for some polynomial $H(X,Y)$. Separating out the initial homogeneous parts from both sides, which both are of degree d, yields the equality

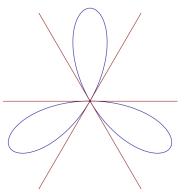
$$F_d(X,Y) = H_{d-2}(X,Y)(Y^2 - X^2)$$

where $H_{d-2}(X,Y)$ is the homogeneous component of H of degree $d-2$, and this is precisely to say that $\mathrm{Gr}_\mathfrak{m} A \simeq k[X,Y]/(Y^2 - X^2)$. ☆

(16.25) This example is at its base generic; the argument persists holding water when the double point is replaced by any hypersurface in $\mathbb{A}^n$; that is, by any polynomial $f \in k[X_1, \ldots, X_n]$ vanishing at the origin (which at least when k is algebraically closed, imposes no restriction as every point then is the origin in an appropriate coordinate system). We let $A = k[X_1, \ldots, X_n]/(f)$ and $\mathfrak{m}$ the maximal ideal at the origin. Moreover, we let f_μ denote the initial form of f; that is, the homogeneous component of f of lowest degree. Then one verifies *mutatis mutandis* as for the simple double point, that $\mathrm{Gr}_\mathfrak{m} A = k[X_1, \ldots, X_n]/(f_\mu)$.

For brevity we write $R = k[X_1, \ldots, X_n]$. Then $A = R/(f)$ is the coordinate ring of the hypersurface $C = V(f) \subseteq \mathbb{A}^N$, and as usual the lower case letters x_i will denote the the classes of the upper case letters X_i in A. In the polynomial ring R we may write $f = f_\mu + h$ where h is a polynomial of higher degree than μ. Denote by u_i the classes of x_i in $\mathrm{Gr}_\mathfrak{m} A$.

We contend that the assignments $X_i \mapsto u_i$ define an isomorphism between the tangent cone $\mathrm{Gr}_\mathfrak{m} A$ and $R/(f_\mu)$. Indeed, any form $F_d(X_1, \ldots, X_n)$ yields elements $F_d(x_1, \ldots, x_n) \in A$ and $F_d(u_1, \ldots, u_n) \in \mathrm{Gr}_\mathfrak{m} A$, and the latter being zero means that $F_d(x_1, \ldots, x_n) = G(x_1, \ldots, x_n)$ with $G(x_1, \ldots, x_n) \in \mathfrak{m}^{i+1}$; hence an equality $F_d - G = A \cdot (f_\mu + h)$, holds in $k[X_1, \ldots, X_n]$, and when we isolate the homogeneous parts of degree d, we find $F_d = A_{d-\mu} f_\mu$ where $A_{d-\mu}$ is the homogeneous component of A of degree $d - \mu$; but this means that $F_d \in (f_\mu)R$, and we are happy.



Exercises

(16.8) Describe the the tangent cone at the origin of the super-cusp $y^p = x^q$, the trefoil in $\mathbb{A}^2$ with equation $(x^2+y^2)^2+3x^2y-y^3=0$ and the du Val singularity $xy+z^{n+1}=0$ in $\mathbb{A}^3$.

(16.9) Consider the algebra $A = k[x,y]$ with constituting relation $y^2+x^2-x^3=0$. Name three different fields k so that $\mathrm{Gr}_{\mathfrak{m}} A$ respectively is a domain, is reduced but not a domain, and is not reduced.

(16.10) Let $\mathfrak{a} \subseteq k[x_1,\dots,x_n]$ be a homogeneous ideal and let $A = k[x_1,\dots,x_n]/\mathfrak{a}$. Show that $A \simeq \mathrm{Gr}_{\mathfrak{m}} A$ where $\mathfrak{m} = (x_1,\dots,x_n)$. This is as should be, a cone is its proper tangent cone

(16.11) Recall that any $f \in k[x_1,\dots,x_n]$ has at initial term $Init(f)$ which is the homogeneous component of lowest degree. Let $\mathfrak{a} \subseteq k[x_1,\dots,x_n]$ be an ideal and let $Init(\mathfrak{a})$ be the ideal generated by all initail terms of members of $\mathfrak{a}$; that is by the set $\{\, Init(f) \mid f \in \mathfrak{a} \,\}$. Show that $\mathrm{Gr}_{\mathfrak{m}} A \simeq k[x_1,\dots,x_n]/Init(\mathfrak{a})$.

David Rees (1918–2013)

British mathematician

16.4 *Filtrations, the Artin–Rees lemme and Samuel functions*

The Artin–Rees lemma is one more important result having kept the status of a lemma, at least in the name. It was more or less simultaneously found by Emil Artin and David Rees in the mid 1950's. The proof we shall give is the one from Nagata's book; albeit simplistic, it is transparent, and one can see what is going on, at least in the case of ideals. The general lemma is inferred from the ideal case by using an unabashed trick Nagata calls "The principal of idealization".

PROPOSITION 16.26 (ARTIN–REES LEMMA) *Let $\mathfrak{a}$ and I be two ideals in the Noetherian ring A. Then there is an integer r so that*

$$\mathfrak{a}^n(\mathfrak{a}^r \cap I) = \mathfrak{a}^{n+r} \cap I$$

for all non-negative integers n.

PROOF: It is clear that $\mathfrak{a}^n(\mathfrak{a}^r \cap I) \subseteq \mathfrak{a}^{n+r} \cap I$, so we merely have to verify the converse inclusion. The ideal $\mathfrak{a}$ is finitely generated, and we may choose generators $a_1,\dots,a_s$ for it. Let Σ be the subset of $A[x_1,\dots,x_s]$ form by the homogeneous polynomials f such that $f(a_1,\dots,a_s) \in \mathfrak{a}^r \cap I$. The polynomial ring $A[x_1,\dots,x_s]$ being Noetherian by Hilbert's basis theorem, the ideal that Σ generates has a finite generator set $f_1,\dots,f_\nu$ consisting of homogeneous forms. The integer r in the statement is the maximum of their degrees.

Let $a \in \mathfrak{a}^{n+r} \cap I$. Then $a \in \mathfrak{a}^{n+r}$, and $a = f(a_1,\dots,a_s)$ for some homogeneous polynomial f of degree $n+r$. As $a \in \mathfrak{a}^{n+r} \cap I \subseteq \mathfrak{a}^r \cap I$, the polynomial belongs to Σ, and

we may write $f = \sum g_i f_i$ with the g_i's are homogeneous and $\deg g_i + \deg f_i = n + r$. Hence

$$a = f(a_1, \ldots, a_s) = \sum g_i(a_1, \ldots, a_s) f_i(a_1, \ldots, a_s).$$

Since $\deg g_i = n + r - \deg f_i \geqslant n$, it holds that $g_i(a_1, \ldots, a_s) \in \mathfrak{a}^n$, and we are through. ❑

PROPOSITION 16.27 (GENERAL ARTIN–REES LEMMA) *Let A be a Noetherian ring and $\mathfrak{a}$ an ideal in A. Let M a finitely generated A-module and $N \subseteq M$ a submodule. Then there is a natural number r so that*

$$\mathfrak{a}^n(\mathfrak{a}^r M \cap N) = \mathfrak{a}^{n+r} M \cap N$$

for all non-negative integers n.

PROOF: The simple trick is to replace the ring A by the ring $A' = A \oplus M$ in which multiplication is defined as $(a + m)(a' + m') = aa' + am' + a'm$ so that $M^2 = 0$. The module M and all its submodules will be ideals in A'. Moreover, one replaces the ideal $\mathfrak{a}$ in A by the ideal $\mathfrak{a}' = \mathfrak{a} \oplus M$ in A'; then $(\mathfrak{a}')^m L = \mathfrak{a}^m L$ for all natural numbers m and all submodules L of M, and the proposition follows immediately from the Artin–Rees lemma for ideals. ❑

(16.28) Assume that M/M_i is of finite length over A, then the Samuel-function associated with $\mathcal{M}$ is the numerical function

$$S_{\mathcal{M}}(n) = \ell_A(M/M_n).$$

LEMMA 16.29 *If M is finite A-module and the filtration $\mathcal{M}$ of M is an $\mathfrak{a}$-filtration, the Samuel function is a numerical polynomial.*

PROOF: To the filtration $\mathcal{M}$ one associates the graded module $\mathrm{Gr}_{\mathcal{M}} M = \bigoplus_i M_i/M_{i+1}$. It holds true that $\Delta S_{\mathcal{M}} = h_{\mathrm{Gr}_{\mathcal{M}}} \mathrm{Gr}_{\mathfrak{a}} A$ so that we may awake prop. ❑

LEMMA 16.30 *Assume that A is a local ring and $\mathfrak{q}$ is an $\mathfrak{m}$-primary ideal. Assume that $\mathcal{M} = \{M_i\}$ and $\mathcal{N} = \{N_i\}$ are two $\mathfrak{q}$-stable filtrations of M. The the Samuel functions $S_{\mathcal{M}}$ and $S_{\mathcal{N}}$ have the same degree and the same leading coefficient.*

PROOF: We may certainly assume that one of the filtrations, say $\mathcal{N}$, is the $\mathfrak{q}$-adic one. Now we have the inclusions $\mathfrak{q}^i M = \mathfrak{q}^i M_0 \subseteq M_i$ and $M_{i+r} = \mathfrak{q}^i M_r \subseteq \mathfrak{q}^i M_0 = \mathfrak{q}^i M$ hence it holds true that

$$S_{\mathcal{M}}(i) \leqslant S_{\mathfrak{q}}(i) \leqslant S_{\mathcal{M}}(i + r)$$

for $i >> 0$ from which ensues that the two leading terms coincide. ❑

LEMMA 16.31 *Assume given an exact sequence*

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

of graded modules over the local ring A, then $d(M) = \max d(M'), d(M'')$.

PROOF: The filtration $\mathcal{M} = \mathfrak{q}^v M \cap M'$ induced in M' from the $\mathfrak{q}$-adic filtration on M is according to the Artin–Rees lemma a $\mathfrak{q}$-stable filtration. By additivity

$$S_{M,\mathfrak{q}} = S_{M',\mathcal{M}} + S_{M'',\mathfrak{q}}$$

and since the leading coefficient of all Samuel polynomials are positive no cancellation can occur, and the degree of $S_{M,\mathfrak{q}}$ equals the larger of the degrees of $S_{M',\mathcal{M}}$ and $S_{M'',\mathfrak{q}}$. ❏

THEOREM 16.32 *Let A be a local ring with maximal ideal. Then it holds true that $d(M) =$* $\dim M$

PROOF: So show that $d(M) \leqslant \dim M$ choose a system of parameters $x_1, \ldots, x_d$ for A; that is $\mathfrak{q}$ they generate is $\mathfrak{m}$-primary and $h_{\mathfrak{m}}(v)$ and $h_{\mathfrak{q}}(v)$ are of the same degree by xxx. But the associated graded ring $\mathrm{Gr}_{\mathfrak{q}} A$ is a quotient of the polynomial ring $A/\mathfrak{q}[t_1, \ldots, t_d]$ where t_i is the class of x_i in $\mathrm{Gr}_{\mathfrak{q}} A$ from which ensues that the degree of $h_{\mathfrak{q}}(v)$ is at most d.

We attack the other inequality, that $\dim M \leqslant d(M)$ by induction on $\dim M$, and to that end there is a chain $\{M_i\}$ whose subquotients are $A/\mathfrak{p}_i$ with the $\mathfrak{p}_i$'s being prime ideals:

$$0 \longrightarrow M_{i+1} \longrightarrow M_i \longrightarrow A/\mathfrak{p}_i(m_i) \longrightarrow 0$$

Now ❏

Lecture 17

Regular sequences

17.1 Depth, regular sequences and unmixedness

An important ingredient in the full proof of Bézout's theorem is the concept of so-called unmixed rings. These are Noetherian rings all whose associated prime ideals are of the same height, or what amounts to the same in our context of algeras of finite type over a field, that dim $A/\mathfrak{p}$ is the same for all associated primes $\mathfrak{p}$. In particual A has no embedded components, the height of an embedded prime would of course be larger than the height of at least on of the others. In geometric terms, if $A = k[x_1, \ldots, x_n]/\mathfrak{a}$, all the components of the closed algtebraic subset $X = Z(\mathfrak{a})$ are of the same dimension and A has no embedded component.

Macaulay showed that if $(F_1, \ldots, F_r)$ is of height r, then $k[x_1, \ldots, x]/(f_1, \ldots, f_r)$ is unmixed. That the irreducible components of the closed algebraic set $Z(f_1, \ldots, f_r)$ all are of codimension r is clear—the height being the smallest codimension of a components, and Krull's Hauptidealsatz tells us that every component is of codimension most r— so the subtle content is that there are no embedded components. This has consequence that if F_{r+1} is a new polynomial not vanishing along any of the components, then F_{r+1} is a non-zero divisor in $k[x_1, \ldots, x_n]/(f_1, \ldots, f_r)$. So we see that $(F_1, \ldots, F_r)$ being of height r is equivalent to $F_1, \ldots, F_r$ being a regular sequence.

Regular sequences

The theory of Cohen–Macaulay rings and more generally of the Cohen–Macaualy modules, is based on the concept of *regular sequences* which was introduced by Jean Pierre Serre in 1955. Their basic properties are described in this paragraph.

(17.1) The stage is set as follows. We are given a ring A together with a proper ideal $\mathfrak{a}$ in A and an A-module M. Most of the time An will be local and Noetherian and M will be finitely generated over A.

Regular sequences (regulære følger)

A sequence $x_1, \ldots, x_r$ of elements belonging to the ideal $\mathfrak{a}$ is said to be *regular for* M, or M*-regular* for short, if the following condition is fulfilled where for notational

convenience we let $x_0 = 0$.

- For any i with $1 \leqslant i \leqslant r$ the multiplication-by-x_i map

$$M/(x_1, \ldots, x_{i-1}) \longrightarrow M/(x_1, \ldots, x_{i-1})$$

  is injective.

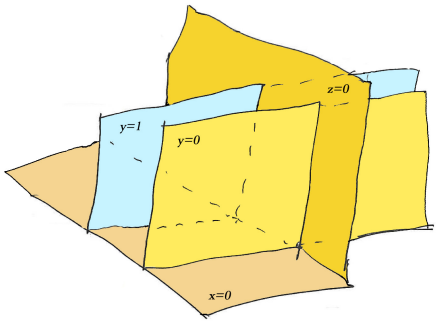
In other words, x_i is a not a zero-divisor in $M/(x_1, \ldots, x_{i-1})$. In particular, x_1 is not a zero-divisor in M, and this has lead to the usage that x_1 being regular in M is synonymous with x_1 being a non-zero divisor in M.

Maximal regular sequences (maksimale regulære følger)

(17.2) A regular sequence $x_1, \ldots, x_r$ is said to be *maximal* if it is no longer regular when an element is added to it. When M is a Noetherian module, this is equivalent to $\mathfrak{a}$ being contained in one of the associated primes of $M/(x_1, \ldots, x_r)$; indeed, the union of the associated primes of $M/(x_1, \ldots, x_r)$ is precisely the set of zero-divisors in $M/(x_1, \ldots, x_r)$.

EXERCISE 17.1 Show that maximal regular sequences for Noetherian modules are finite. Exihibit a counterexample when M is not Noetherian. HINT: Consider the ascending chain $(x_1, \ldots, x_i)$ of ideals. ★

(17.3) Be aware that in general the order of the x_i's is important, permute them and the sequence may no more be regular. However, regular sequences for modules finitely generated over local Noetherian rings remain regular after an arbitrary permutation, and the same holds true for graded rings with appropriate finiteness conditions. Henceforth, we shall work with rings that are Noetherian and local and with modules finitely generated over A (but with a sideways glimpse into the graded case, so important for projective geometry).



EXAMPLE 17.1 The simplest example of a sequence that ceases being regular when permuted is as follows. Start with the three coordinate planes in $\mathbb{A}^3$; they are given

as the zero loci of x, y, z. Add a plane disjoint from one of them to the two others; *e.g.* consider the zero loci of the three polynomials $x(y-1)$, y and $z(y-1)$.

Clearly $x(y-1), z(y-1), y$ is *not* a regular sequence in $k[x,y,x]$. The point is that $z(y-1)$ kills any function on $Z(x(y-1))$ that vanishes on the component $Z(x)$ (for example x) and is thus not a zero-divisor in $k[x,y,z]/(x(y-1))$.

On the other hand, the sequence $x(y-1), y, z(y-1)$ is regular. Indeed, it holds that $k[x,y,z]/(x(y-1),y) = k[z]$, and in that ring $z(y-1)$ is congruent to z and thus not a zero-divisor. Geometrically, capping $Z(x(y-1))$ with $Z(y)$ makes the villain component $Z(y-1)$ go away.

This example is in fact arche-typical. The troubles occur when two of the involved closed algebraic sets have a common component disjoint from one of the components of a third. If all components of all the closed algebraic subsets involved have a point in common, one is basically in a local situation, and permutations are permitted. ☆

Permutation permitted

(17.4) As mention in the previous example, in local Noetherian rings a sequence being regular is a property insensitive to order. The same holds true in a graded setting, and in both cases Nakayama's lemma is the tool that makes it work.

LEMMA 17.5 *Assume that A is a local Noetherian ring with maximal ideal $\mathfrak{m}$ and M a finitely generated A-module. If x_1, x_2 is a regular sequence in $\mathfrak{m}$ for M, then x_2, x_1 is one as well.*

PROOF: There are two things to be checked. Firstly, that x_2 is a non-zero divisor in M. The annihilator $(0 : x_2)_M = \{\, a \in M \mid x_2\, a = 0 \,\}$ must map to zero in M/x_1M because multiplication by x_2 in M/x_1M is injective. Hence $(0 : x_2)_M + x_1M = x_1M$, and since $x_1 \in \mathfrak{m}$ and M is finitely generated, Nakayama's lemma applies and $(0 : x_2)_M = 0$.

Secondly, we are to see that multiplication by x_1 is injective on M/x_2M, so assume that $x_1a = x_2b$. But multiplication by x_2 is injective on M/x_1M, and it follows that $b = cx_1$ for some c; that is, $x_1a = x_1x_2c$. Cancelling x_1, which is legal since x_1 is a non-zero divisor in M, we obtain $a = cx_2$. ❑

PROPOSITION 17.6 *Let A be a local Noetherian ring with maximal ideal $\mathfrak{m}$ and M a finitely generated A-module. Assume that $x_1, \dots, x_r$ is a regular sequence in $\mathfrak{m}$ for M. Then for any permutation σ the sequence $x_{\sigma(1)}, \dots, x_{\sigma(n)}$ is regular.*

PROOF: It suffices to say that any permutation can be achieved by successively swapping neighbours . ❑

(17.7) The graded version reads as follows:

PROPOSITION 17.8 *Let A be a graded ring satisfying $A_i = 0$ when $i < 0$, and let M be a finitely generated graded A-module. If $x_1, \dots, x_r$ is a sequence of elements from A, homogeneous*

of positive degree, that form a regular sequence in M, then for any permutation σ the sequence $x_{\sigma(1)}, \dots, x_{\sigma(r)}$ is also a regular in M.

PROOF: As above, one may assume that $r = 2$. The proof of Lemma 17.5 goes through *mutatis mutandis*; the sub module $(0 : x_2)_M$ will be a graded submodule because x_2 is homogeneous, and a version of Nakayma's lemma for graded things is available. ❑

EXERCISE 17.2 With assumptions as in 17.6 or 17.8, prove that if $x_1, \dots, x_r$ is a regular sequence for M and $\nu_1, \dots, \nu_r$ is a sequence of natural numbers, then $x_1^{\nu_1}, \dots, x_r^{\nu_r}$ will be a regular sequence as well. HINT: Reduce to the case of $x_1, \dots, x_{r-1}, x_r^{\nu}$. ★

EXERCISE 17.3 Jean Dieudonné gave the following example of a regular sequence x_1, x_2 in local non-Noetherian ring such that x_2, x_1 is not regular. Consider the ring B of germs of C^∞-functions near 0 in $\mathbb{R}$. It is a local ring whose maximal ideal $\mathfrak{m}$ constists of the functions vanishing at zero. Let $\mathfrak{a}$ be the ideal $\mathfrak{a} = \bigcap_i \mathfrak{m}^i$ of functions all whose derivatives vanish at the origin. Let $A = B[T]/\mathfrak{a}TB[T]$. Let I be the function $I(x) = x$. Show that the sequence I, T is a regular sequence in A whereas T, I is not. ★

Enters homolocigal algebra—the depth

(17.9) One of the first appearances of homological methods in commutative algebra was in the circle of ideas round of regular sequences and Cohen-Macaulay modules. These methods give a caracterization of the maximal length of M-regular sequences in terms of certain homologically defined modules. The criterion has the virtue of not explicitly refering to any sequence, and has the consequence that all maximal sequence are of the same length.

The homological modules in question are modules $\operatorname{Ext}^i_A(M, N)$ associated with a pair of A-modules M and N. In the lingo of homological algebra they appear as derived functors of the functor $\operatorname{Hom}_A(-,-)$. Students not already acquainted with these useful creatures should consult a textbook about homological algebra for the few of their very basic properties we shall need.

The depth of a module (dybden til en modul)

(17.10) It is natural to introduce the number $\operatorname{depth}_{\mathfrak{a}} M$ as the length of the longest (maximal) regular M-sequence in $\mathfrak{a}$. It is called the *depth* of M in $\mathfrak{a}$. In the end, it turns out that all maximal M-sequences in $\mathfrak{a}$ have the same length, but for the moment we do not know that, and *a priori* the number is not even bounded. However, we have:

LEMMA 17.11 *If A is a local Noetherian ring, $\mathfrak{a}$ a proper ideal and M a finitely generated A-module, then* $\operatorname{depth}_{\mathfrak{a}} M \leqslant \dim M$. *In particular,* $\operatorname{depth}_{\mathfrak{a}} M$ *is finite.*

PROOF: Induction on $\dim M$ (which is finite!). If $\dim M = 0$, the maximal ideal $\mathfrak{m}$ is the only associated prime of M. Therefore every element in $\mathfrak{m}$ is a zero divisor and $\operatorname{depth}_{\mathfrak{a}} M = 0$.

Next, observe that if x is a non-zero divisor in M, it holds true that $\dim M/xM < \dim M$, and by induction one may infer that

$$\operatorname{depth}_{\mathfrak{a}} M/xM \leqslant \dim M/xM < \dim M. \tag{17.1}$$

So if $x_1, \dots, x_r$ is a maximal regular sequence in M (they are all finite after Problem 17.1), the sequence $x_2, \dots, x_r$ will be one for M/x_1M, and by (17.1) $r - 1 < \dim M$; that is $r \leqslant \dim M$. ❑

(17.12) We have comes to the homological characterization. It is notable since it determines the depth of a module without referring to any regular sequence. We introduce a number $p(M)$ which is the smallest integer i such that $\operatorname{Ext}^i_A(A/\mathfrak{a}, M) \neq 0$.

PROPOSITION 17.13 *Let A be a local Noetherian ring, $\mathfrak{a}$ a proper ideal and M a finitely generated A-module. Then* $\operatorname{depth}_{\mathfrak{a}} M = p(M)$.

PROOF: The proof goes by induction on the depth of M (which is finite by Lemma 17.11 above). That $\operatorname{depth}_{\mathfrak{a}} M = 0$, means that there no element in $\mathfrak{a}$ is regular in M. In other words, $\mathfrak{a}$ is contained in one of the associated primes of M, say $\mathfrak{p}$. There is then an inclusion $A/\mathfrak{p} \to M$ and a surjection $A/\mathfrak{m} \to A/\mathfrak{p}$. Consequently $\operatorname{Hom}_A(A/\mathfrak{a}, M) \neq 0$, and $p(M) = 0$.

Assume next that $\operatorname{depth}_{\mathfrak{a}} M > 0$. If x is the first member of an M-regular sequence of maximal length, the quotient M/xM satisfies $\operatorname{depth}_{\mathfrak{a}} M/xM = \operatorname{depth}_{\mathfrak{a}} M - 1$. Moreover, since x is regular on M, one has the short exact sequences

$$0 \longrightarrow M \xrightarrow{x} M \longrightarrow M/xM \longrightarrow 0,$$

from which one derives a long exact sequence the relevant part for us being

$$\operatorname{Ext}^i_A(A/\mathfrak{a}, M) \longrightarrow \operatorname{Ext}^i_A(A/\mathfrak{a}, M/xM) \longrightarrow \operatorname{Ext}^{i+1}_A(A/\mathfrak{a}, M) \xrightarrow{x=0} \operatorname{Ext}^{i+1}_A(A/\mathfrak{a}, M).$$

Since $x \in \mathfrak{a}$ the multiplication by x on the ext-modules is the zero map. Now, if $i + 1 < p(M)$ it ensues that $\operatorname{Ext}^i_A(A/\mathfrak{a}, M/xM) = 0$, and we may conclude that $p(M/xM) + 1 \leqslant p(M)$. And if $i + 1 = p(M)$ it follows that $\operatorname{Ext}^i_A(A/\mathfrak{a}, M/xM) \simeq \operatorname{Ext}^{i+1}_A(A/\mathfrak{a}, M) \neq 0$ so equality holds.

So both the quantities $\operatorname{depth}_{\mathfrak{a}} M$ and $p(M)$ drops by one when we mod out by x, and thence they are equal by induction. ❑

The proposition has an important corollary, which in fact is the main target of this paragraph:

THEOREM 17.14 *Let A be a local Noetherian ring, $\mathfrak{a}$ an ideal in A and M a finitely generated A-module. Then all maximal regular M-sequences in $\mathfrak{a}$ have the same length.*

EXAMPLE 17.2 A Noetherian zero-dimensional local ring has of course depth zero. A Noetherian one-dimensional local ring A has depth one if and only if the maximal ideal is not associated; that is, A has no embedded component. ☆

As usual, we also give a graded version:

THEOREM 17.15 *Let A be a graded ring satisfying $A_i = 0$ when $i < 0$, and let M be a finitely generated graded A-module. Then all homogeneous maximal regular M-sequences have the same length.*

The bound

In geometry the dimension of a close algebraic set is the maximum dimension of the irreducible components, and the algebraic counterpart is that the dimension of a ring is the maximum of the dimensions $\dim A/\mathfrak{p}$ for $\mathfrak{p}$ running through the associated prime ideals of A. This maximum is never assumed at an embedded prime since these by definition strictly contain another associated prime. For a module M, the same holds true as $\dim M = \dim / \operatorname{Ann} M$.

The word *depth* has the flavour of something down, and indeed, $\operatorname{depth}_{\mathfrak{m}} M$ is smaller then all the dimensions $\dim A/\mathfrak{p}$ where this time $\mathfrak{p}$ runs through *all* of the associated primes, including the embedded ones. And this is the crucial point.

PROPOSITION 17.16 *As usual, let A be local Noetherian ring with maximal ideal $\mathfrak{m}$ and let M be a finitely generated A-module. It then holds true that*

$$\operatorname{depth}_{\mathfrak{m}} M \leqslant \dim A/\mathfrak{p}$$

for all prime ideals associated to M.

PROOF: The proof goes by induction on the depth of M. If $\operatorname{depth}_{\mathfrak{m}} M = 0$ there is nothing to prove. So assume that $\operatorname{depth}_{\mathfrak{m}} M \geqslant 1$. Then there is a short exact sequence

$$0 \longrightarrow M \xrightarrow{x} M \longrightarrow M/xM \longrightarrow 0. \tag{17.2}$$

Let $\mathfrak{p}$ be prime ideal associated to M. The sequence (17.2) above induces an exact sequence

$$0 \longrightarrow \operatorname{Hom}_A(A/\mathfrak{p}, M) \xrightarrow{x} \operatorname{Hom}_A(A/\mathfrak{p}, M) \longrightarrow \operatorname{Hom}_A(A/\mathfrak{p}, M/xM),$$

and by Nakayama's lemma the cokernel of the multiplication-by-x map is a non-zero submodule of $\operatorname{Hom}_A(A/\mathfrak{p}, M/xM)$. Hence $\operatorname{Hom}_A(A/\mathfrak{p}, M/xM) \neq 0$, and the ideal $\mathfrak{p} + (x)$ is contained in an associated ideal prime ideal $\mathfrak{q}$ of M/xM. Since x is a non-zero divisor in M and $\mathfrak{p}$ is associated to M, we may infer that $x \notin \mathfrak{p}$, and therefore $\mathfrak{p}$ is

strictly contained in $\mathfrak{q}$. It follows that $\dim A/\mathfrak{p} > \dim A/\mathfrak{q}$. Now, $\operatorname{depth}_m M/xM = \operatorname{depth}_{\mathfrak{m}} M - 1$, and by induction

$$\operatorname{depth}_{\mathfrak{m}} M - 1 \leqslant \dim A/\mathfrak{q} < \dim A/\mathfrak{p}.$$

❑

(17.17) The A-module M is said to be a *Cohen-Macaulay module* if $\operatorname{depth}_{\mathfrak{m}} M = \dim M$, in particular, the ring A itself is *Cohen-Macaulay* if $\operatorname{depth}_{\mathfrak{m}} A = \dim A$. If x is a non-zero divisor in A, both the depth and the dimension of A/xA are one less than of A, and hence A is Cohen–Macaulay if and only if A/xA is.

Cohen-Macaulay modules
(Cohen-Macaulay moduler)

THEOREM 17.18 *Assume that A is a local Noetherian Cohen-Macaulay ring. Then A is unmixed. That is $\dim a/\mathfrak{p} = \dim A$ for all associated primes $\mathfrak{p}$ of A; in particular, A has no embedded components.*

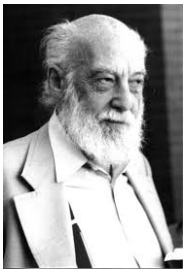
PROOF: In view of Proposition 17.16 this is almost a tautology. The lower and the upper bound of the dimensions $\dim A/\mathfrak{p}$ for $\mathfrak{p}$ associated with A coincide, hence these dimensions all coincide. ❑

(17.19) To check that a ring is Cohen–Macaulay, it suffices to exhibit one regular sequence of length the dimension of the ring. For instance, the local rings $A_n = k[x_1, \dots, x_n]_{\mathfrak{m}_n}$ where $\mathfrak{m}_n = (x_1, \dots, x_n)$ are Cohen–Macaulay since the sequence $x_1, \dots, x_n$ is regular. This follows easily by induction because there are natural isomorphisms $A_n/x_nA_n \simeq A_{n-1}$ induced by the maps $k[x_1, \dots, x_n] \to k[x_1, \dots, x_{n-1}]$ that send x_n to zero.

Lecture 18

Categories

The concepts of categories, functors and natural transformations were introduced during World War II, or more precisely at end of it in 1945, in paper by Saunders Mac Lane and Samuel Eilenberg and was a giant step in the development of homological algebra. Functors and natural transformations existed in mathematics long time before, but without being neither explicitly defined nor named. After the birth of algebraic topology and the subsequent development of homological algebra they surfaced as natural foundations for the theory, and the concepts in a fairly explosive manner permeated a wide variety of subjects, to lend Peter Freyd's words.

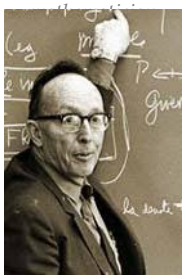


Samuel Eilenberg (1919–1998)

Hungarian–American

Some years later, in then 1950's the theory category were substantially refined and developed by the french mathematical milieu with Alexander Grothendieck in the front row. His monumental contribution to mathematics was based on categories; you meet them in every thing he did.

This short chapter does not have an ambitious approach to categories and functors, and in very few words, merely aims at giving a very first introduction to the language. In the course the language is used as no more than a notational devise which one hardly can do without in a modern introduction to commutative algebra. Students aiming at a pursuit of studies in algebraic geometry or algebraic topology will certainly need a deeper understanding and a much broader mastering of the subject, but this text will hopefully be a good beginning.



Saunders Mac Lane (1909–2005)

American mathematician

(18.1) Anyone studying mathematics at a certain level has experienced that the introduction of a new class of mathematical objects is accompanied by the introduction of a fresh class of maps; namely the maps that respect the structures one wants to investigate. Cases to have in mind can be vector space and linear maps or topological spaces together with continuous maps.

A category is an axiomatic construct mimicking this situation. A category C has two components; the class* of objects and the class of "maps" usually called *morphisms*.

Morphisms (morfier)

** Well... set-theoretical issue here with paradoxes like Russel's lurking , hence the word* class *and not* set. *In our shallow introduction for beginners we just ignore this issue and think about classes being sets (as in fact, most researchers do as well)*

From the concrete model cases one merely carries along two things, the possibility to compose two composable maps—that is, two maps with the target of one equals the source of the other—and secondly, that every object has an identity map.

Even though the concept of categories is modelled on concrete situations, as in any axiomatized theory, objects and morphisms can be any class and the composition any collection of maps as longs the axioms are obeyed. One may be tempted to call the morphisms *arrows*, when wanting to emphasise the axiomatic nature of the matter. So category theory is a game of arrows!

Categories (kategorier)
Source of morphisms (kilden til morfier)
target of morphisms (målet til morfier)

(18.2) Here comes the formal definition: A *category* consists of a class $\mathrm{Ob}\,\mathsf{C}$ of objects, and secondly, for each pair X and Y of objects in $\mathrm{Ob}\,\mathsf{C}$ a set $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ of arrows from X to Y. If ϕ is such an arrow, X is called the *source* and Y the *target* of ϕ. For each triple of objects X, Y and Z, there must be given a "composition"; that is, a map

$$\mathrm{Hom}_{\mathcal{C}}(X,Y) \times \mathrm{Hom}_{\mathcal{C}}(Y,Z) \to \mathrm{Hom}_{\mathcal{C}}(X,Z)$$

which is written as $\phi \circ \psi$. Moreover, for any object X from the category there is a special arrow $\mathrm{id}_X \in \mathrm{Hom}(X,X)$ called the *identity*. These givens are subjected to the following two axioms, the first asserting that the identity arrows are neutral with respect to composition, and the second that composition is associative:

Identities (identiteter)

i) $\phi \circ \mathrm{id}_X = \phi$ and $\mathrm{id}_Y \circ \phi = \phi$;
ii) $\phi \circ (\psi \circ \rho) = (\phi \circ \psi) \circ \rho$.

$W \xrightarrow{\rho} Z \xrightarrow{\psi} Y \xrightarrow{\phi} X$

where ϕ and ψ and ψ and ρ are two pairs of composable morphisms.

EXAMPLE 18.1 There is a long list of categories one could call *conservative* or *traditional*, whose objects are sets equipped with some extra structure and whose arrows are maps respecting that structure, these are the ones we referred to in the motivating introduction above. However, there are many, many others, and we shall give a few examples.

Alexander Grothendieck (1928–1998)

Stateless–French mathematician

Examples of the conservative categories are legio. Just to mention a few which are central in this text: The category Ab of abelian groups and group homomorphisms, the category Rings of commutative rings with unit and maps of rings, and the categories Mod_A and Alg_A of respectively A-modules and A-algebras with corresponding homomorphisms. And of course, one has the category Sets whose objects are the sets and whose arrows are just the ordinary set-theoretical maps. ☆

EXAMPLE 18.2 *Partial ordered sets:* Any partially ordered set P can be interpreted as a category P. The objects are just the elements in P; that is, $\mathrm{Ob}\,\mathsf{P} = P$, and the arrows are as follows. If $x, y \in \mathrm{Ob}\,\mathsf{P}$, the set $\mathrm{Hom}_{\mathsf{P}}(x,y)$ is either empty or a singleton, and it has an element if and only if $x \leqslant y$. That $x \leqslant x$, ensures that identity maps exist, and the order relation being transitive, ensures that composition is defined and associative. ☆

EXAMPLE 18.3 *Topological spaces up to homotopy:* The category Top of topological space with continuous maps as morphisms was, together with its satellite category, the

homotopy category HomTop, one of the first to be studied. The category HomTop plays an important role in algebraic topology, and was a first and fundamental example of a category whose morphism are not structure preserving maps between sets. Recall that two continuous maps $\phi, \psi\colon X \to Y$ are homotopic if there is a continuous map $\Phi\colon X \times I \to Y$ so that $\Phi(x,0) = \phi(x)$ and $\Phi(x,1) = \psi(x)$. One checks that this is an equivalence relation compatible with composition; *i. e.* $\phi \sim \phi'$ then $\psi \circ \phi \sim \psi \circ \phi'$ and $\phi \circ \psi \sim \phi' \circ \psi$. This means that one may "compose" homotopy classes, and thus topological spaces and homotopy classes of maps, form a category. ☆

Example 18.4 *Local rings; subcategories:* In the category Loc of local rings the objects are local rings A and the arrows the ring homomorphisms that map the maximal ideal into the maximal ideal; that is, rimg maps $\phi\colon A \to B$ such that $\phi(\mathfrak{m}_A) \subseteq \mathfrak{m}_B$, where $\mathfrak{m}_A$ and $\mathfrak{m}_B$ denote the two maximal ideals. This is an example of a *subcategory*. Clearly the objects Ob Loc from a subclass of Ob Rings, and for each pair A and B of objects, the set $\mathrm{Hom}_{\mathsf{Loc}}(A,B)$ is a subset $\mathrm{Hom}_{\mathsf{Rings}}(A,B)$. Moreover, composition in Loc agrees with composition in Rings.

Subcategories (underkategorier)

Notice, that in many instances $\mathrm{Hom}_{\mathsf{Loc}}(A,B)$ will be a proper subset of the set of mappings $\mathrm{Hom}_{\mathsf{Loc}}(A,B)$ (*e.g.* if B is the fraction field of the local domain A), and for most subcategories similarly one expects proper inclusions. This leads to the concept of a *full subcategory* A subcategory C of D is said to *full* if for any two objects X and Y from C, the C-arrows from X to Y coincide with the D-arrows; in other words when it holds that $\mathrm{Hom}_{\mathsf{C}}(X,Y) = \mathrm{Hom}_{\mathsf{D}}(X,Y)$ for all $X, Y \in \mathsf{C}$. ☆

Full subcategories (fulle underkategorier)

Functors

The principle of introducing maps respecting structure along with a new kind of structures, applies naturally also to categories. These new "maps" are called *functors*. In contrast to ordinary maps, they operate on the two levels of a category, both on the objects and on the arrows. So from a formal viewpoint they are not maps; though, one often has a mental picture of them as maps.

Functors (funktorer)

Functors come in two variants. The *contravariant functors* reverse the direction of all arrows, whereas they are kept by the *covariant* ones. The two species are equally important, but for the sake of a short and simple presentation, in what follows we shall only deal with the covariant ones.

Contravariant functors (kontravariacontravariantnet funktorer)
Covariant functors (kovariante funktorer)

Here comes the formal definition. If A and B are two categories a (covariant) functor $F\colon \mathsf{A} \to \mathsf{B}$ is a collection of maps with the following constituents. Firstly, a map $F\colon \mathrm{Ob}\,\mathsf{A} \to \mathrm{Ob}\,\mathsf{B}$ that takes objects to objects, and, secondly, for each pair X and Y of objects from A, a map $F\colon \mathrm{Hom}_{\mathsf{A}}(X,Y) \to \mathrm{Hom}_{\mathsf{B}}(F(X),F(Y))$. The ingredient maps are subjected to rules

- $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$;

❑ $F(\phi \circ \psi) = F(\phi) \circ F(\psi)$.

EXAMPLE 18.5 We have met $\mathrm{Hom}_A(-, M)$ and $-\otimes_A M$, which are functors Mod_A to Mod_A; and we have bases change functor $-\otimes_A B\colon \mathsf{Mod}_A \to \mathsf{Mod}_B$ whenever B is an A-algebra. ☆

EXAMPLE 18.6 From any conservative category C to Sets there is a so-called forgetful functor that just throws away the extra structure and sends an object X to the underlying set and a map to underlying set-theoretical map. A variant are functors forgetting parts of a structure with several layers; for instance, the functor from Rings → Ab that forgets the multiplication; *i. e.* it sends a rings to the underlying abelian group and a map to the underlying homomorphism between the additive structures. ☆

EXERCISE 18.1 Make a list of all categories (up to equivalence) with three objects and not more than nine arrows. ★

Natural transformations

This is the thirds concept mentioned at the top of the chapter are the natural transformations.They are devises to compare different functors, and in fact, they nicely fit into picture of "maps" preserving structure; but this time "maps" between functors.

Given two functors F and G from A to B. A natural transformation is a collection of arrows $\Phi_A\colon F(A) \to G(A)$ in B, one for each object $A \in \mathrm{Ob}\,\mathsf{A}$, which are compatible with the double action of functors, on objects and arrow, That is, they are requested to render commutative all diagrams

$$\begin{array}{ccc} F(A) & \xrightarrow{\Phi_A} & G(A) \\ {\scriptstyle F(\phi)}\downarrow & & \downarrow{\scriptstyle G(\phi)} \\ F(A') & \xrightarrow[\Phi_{A'}]{} & G(A') \end{array}$$

where $\phi\colon A \to A'$ is any arrow in A. Two natural transformations, from F to G and one from G to H can obviously be composed, just compose each Ψ_A and Φ_A. So the class of functors from A to B form a category, the functor category $\mathsf{Func}(\mathsf{A}, \mathsf{B})$.

A natural equivalence between two functors F and G is a natural transformations $\Phi\colon F \to G$ having an inverse $\Psi\colon G \to F$; that is, one so that $\Phi \circ \Psi = \mathrm{id}_G$ and $\Psi \circ \Phi = \mathrm{id}_F$. When F and G are natural equivalent, one writes $F \sim G$.

Equivalent categories (ekvivalente kategorier)

(18.3) Two categories are said to be *equivalent* if there are functors $F\colon \mathsf{A} \to \mathsf{B}$ and $G\colon \mathsf{B} \to \mathsf{A}$ such that $F \circ G \sim \mathrm{id}_{\mathsf{B}}$ and $G \circ F \sim \mathrm{id}_{\mathsf{A}}$.

Equivalence between categories replaces the notion of isomorphism we know from other areas of axiomatic mathematics. It reflects that every object in one category is

isomorphic to one in the other and that these isomorphisms preserve the set homomorphism; *i. e.* given two objects X and Y from A, then F and G induce mutually inverse maps between the arrow-sets $\mathrm{Hom}_{\mathsf{A}}(X, Y)$ and $\mathrm{Hom}_{\mathsf{B}}(F(X), F(Y))$. Every theorem valid in one will as well be valid in the other, with the understanding that proofs are formulated merely in specific categorial terms.

Be aware however, that $\mathrm{Ob}\,\mathsf{A}$ and $\mathrm{Ob}\,\mathsf{B}$ are often far from being bijective.

Example 18.7 A stupid example, the category whose objects are all singletons (or of any singleton of your choice) and all hom-sets also being singletons (*e.g.* $\mathrm{Hom}_s(\{x\}, \{y\}) = \{(x, y)\}$) is equivalent to the tiny toy category with just one object $*$ and $\mathrm{Hom}(*, *) = \{\mathrm{id}_*\}$. ☆

Lecture 19

Solutions

Solutions for exercises in Chapter 1

EXERCISE 1.1 Units are evidently not zero divisors. Le x be a non-zero element in A. Multiplication by x is an additive group homomorphism $A \to A$ which is injective when x is not a zero divisor A. Hence it is surjective as A is finite. It follows that 1 lies in the image; that is, there is a $y \in A$ with $xy = 1$, so x is invertible.

EXERCISE 1.2 We have $72 = 2^3 3^2$. If $72|a^n$, both 2 and 3 must divide a; hence the nilpotent elements are all multiples of 6. If $72|ab$, either 2 or 3 must divide a, so the zero divisors are the multiples of 2 and the multiples of 3. The units are those elements not a multiple of 3 or 2 (for instance, by the previous exercise), that is, the classes of numbers prime to 2 and 3.

EXERCISE 1.3 Let $p_1, \ldots, p_r$ be the different primes that occur in a factorization of n into prime powers. If $n|a^m$, all the p_i's divide a; hence the nilpotens are the multiples of $p_1 \cdot \ldots \cdot p_r$. If $n|ab$, one of the p_i's must divide a. Conversely elements shaped like $a = a' p_i$ kill np_i^{-1}, and the latter being non-zero in $\mathbb{Z}/n\mathbb{Z}$, they are zero-divisors; so the zero divisors are precisely the multiples of the p_i's. Again by Exercise 1.1, the units are the classes of elements not divisible by any of the p_i's.

EXERCISE 1.8 let a/n be an element in A which is written in lowest terms; that is, a and n are without common factors. Then there is a relation $1 = xa + yn$ with $x, y \in \mathbb{Z}$, which gives $1/n = x \cdot a/n + y$; so $1/n \in A$, and consequently $1/p \in A$ for all primes dividing n. Letting S be the set of all primes dividing denominators of elements from A, we see that $A = \mathbb{Z}[p^{-1} | p \in S]$.

EXERCISE 1.10 (*Units in imaginary quadratic extensions*) Write $\eta = i\sqrt{n}$. The elements in $\mathbb{Z}[\eta]$ are all the linear combinations $x = a + b\eta$ where a and b are integers. The squared absolute value of x is given as $|x|^2 = x\bar{x} = a^2 + b^2 n$, and this is always an integer. Now, if $xy = 1$, it follows that $|x|^2|y|^2 = 1$, and $|x|^2$ being an integer, it ensues that $|x| = 1$.

The inverse of x is given as $x^{-1} = \bar{x}|x|^{-2}$, and it equals $\bar{x}$ and lies in $\mathbb{Z}[\eta]$ whenever $|x| = 1$. When $n > 1$, the equation $a^2 + nb^2 = 1$ forces $b = 0$ and $a = \pm 1$, but when $n = 1$, the equation admits the solutions $a = 0$ and $b = \pm 1$ as well. So when $n > 1$ the only units are ± 1 and the group of units is cyclic of order two, but when $n = 1$, they constitute the four-group $\{\pm 1, \pm i\}$.

EXERCISE 1.11 The formula for the sum of a geometric series reads (and is valid in any ring)

$$1 + x + x^2 + \ldots + x^\nu = \frac{1 - x^{\nu+1}}{1 - x},$$

so with $\nu = n - 1$ and $x = -a$ it does the job.

EXERCISE 1.14 (*Units in polynomial rings*) That f is invertible when a is invertible and all the other coefficients are nilpotent, follows from Exercise 1.11 on page 21. So assume that $f(x) = 1 + a_1 x + \ldots + a_n x^n$ is invertible and let $g(x) = 1 + b_1 + \ldots + b_m x^m$ be the inverse (we can safely assume that $a_0 = b_0 = 1$) We aim at showing by induction on i that $a_n^{i+1} b_{m-i} = 0$ for $0 \leqslant i < n + m$. For $i = 0$ this holds since $a_m b_m$ is the only term in fg of degree $n + m$.

Now, it holds that

$$a_n b_{m-i} + a_{n-1} b_{m-(i-1)} + a_{n-2} b_{m-(i-2)} + \ldots = 0$$

which gives

$$a_n^{i+1} b_{m-i} + a_{n-1} a_n^i b_{m-(i-1)} + a_{n-2} a_n^i b_{m-(i-2)} + \ldots = 0$$

By induction, $a_n^i b_{m-(i-1)} = a_n^{i-1} b_{m-(i-2)} = \ldots = 0$, and we conclude that $a_n^{i+1} b_{m-i} = 0$. Hence with $i = m$, it follows that $a_n^{m+1} = 0$. To finish off the exercise, observe that $f(x) - a_n x^n$ will be invertible when $f(x)$ is (Exercise 1.11), so repeating the procedure we obtain eventually that all the a_i's are nilpotent.

EXERCISE 1.16 A homomorphism $k[t, t^{-1}] \to k[u]$. would take the group of units of $k[t, t^{-1}]$ into the group of units of $k[u]$. The latter equals the invertible scalars k^*, and all powers of t being invertible in $k[t, t^{-1}]$, would be mapped into k^*. Consequently the whole of $k[t, t^{-1}]$ would be mapped into k, which is absurd!

EXERCISE 1.18 (*Long division*)

a) We proceed by induction of the degree d of f. Let $n = \deg g$. The claim is trivial if $\deg f < \deg g$, we just take $q = 0$ and $r = f$. So let α be the leading coefficient of f and β that of g. Then $\beta f(t) - \alpha t^{d-n} g(t)$ is of lower degree than f and may by induction we may for some $\gamma \in A$ write $\gamma\beta f(t) - \gamma\alpha t^{d-n} g(t) = q'(t)g(t) + r(t)$ with $\deg r < \deg g$. Then we find

$$af(t) = g(t)(q'(t) + \gamma\alpha t^{d-n}) + r(t)$$

for $a = \gamma\alpha$, and we are through.

b) Induction on the degree, the degree one case being evident. Let a be a root of f, and divide f by $t - a$ to obtain $f(t) = q(t)(t-a) + r(t)$ where r is of degree less than one, hence it must be a constant. Setting $t = a$ gives $r = 0$, and consequently $f(t) = q(t)(t-a)$. If b is another root of f different from a, this relation gives $0 = f(b) = q(b)(b-a)$. It ensues $q(b) = 0$ as A is a domain. The quotient q being of degree $\deg f - 1$ induction applies, and we may infer that q has less than $\deg f - 1$ roots. Consequently f has less than $\deg f$ roots.

EXERCISE 1.23 One has $1 = 4 - 3$; hence it holds that $4 = 4^2 - 12 \equiv 4^2 \mod 12$, so the class of 4 is an idempotent in $\mathbb{Z}/12\mathbb{Z}$. Similarly, $-3 = 12 + (-3)^2$, so -3 is an idempotent in $\mathbb{Z}/12\mathbb{Z}$. They are together with the trivial idempotents the only ones. Indeed, the corresponding decomposition of $\mathbb{Z}/12\mathbb{Z}$ into a direct product is $\mathbb{Z}/12\mathbb{Z} \simeq \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$ with 4 and -3 representing the unit elements. Now, both $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/12\mathbb{Z}$ only have trivial idempotents, and we are done.

One has $1 = 9 - 8$ so 9 and -8 are idempotents in $\mathbb{Z}/36\mathbb{Z}$. The corresponding decomposition of $\mathbb{Z}/36\mathbb{Z}$ in a product is $\mathbb{Z}/36\mathbb{Z} \simeq \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ with -8 being 1 in $\mathbb{Z}/9\mathbb{Z}$ and 9 being 1 in $\mathbb{Z}/8\mathbb{Z}$. The idempotents in $\mathbb{Z}/8\mathbb{Z}$ and $\mathbb{Z}/9\mathbb{Z}$ are the trivial ones, hence we have found all.

Solutions for exercises in Chapter 2

EXERCISE 2.1 If $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ obviously $ab \in \mathfrak{a} \cap \mathfrak{b}$, and sums of elements shaped like ab belong then to $\mathfrak{a} \cap \mathfrak{b}$ as well. If $a, b \in \mathfrak{a} \cap \mathfrak{b}$ obviously $ab \in \mathfrak{a}\mathfrak{b}$ so that $(\mathfrak{a} \cap \mathfrak{b})^2 \subseteq \mathfrak{a}\mathfrak{b}$.

As to the first example, let $\mathfrak{a} = \mathfrak{b} = (x)$ in the polynomial ring $k[x]$. Then $\mathfrak{a}\mathfrak{b} = \mathfrak{a}^2 = (x^2)$, but $\mathfrak{a} \cap \mathfrak{b} = (x)$. For the second take $\mathfrak{a} \subset \mathfrak{b}$; for instance, let $\mathfrak{a} = (x)$ and $\mathfrak{b} = (x, y)$ in the polynomial ring $k[x, y]$. Then $\mathfrak{a} \cap \mathfrak{b} = (x)$ and $(\mathfrak{a} \cap \mathfrak{b})^2 = (x^2)$, whereas $\mathfrak{a}\mathfrak{b} = (x^2, xy)$.

EXERCISE 2.6 We contend that the elements 1, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{6}$ form a basis for the $\mathbb{Z}$-module $\mathbb{Z}[\sqrt{2}, \sqrt{3}]$. That they generate is clear. To see they are linearly independent over $\mathbb{Z}$ it suffices to see they are linearly independent over $\mathbb{Q}$; in other words, it suffices to show that the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ has dimension four over $\mathbb{Q}$. Now, $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is a quadratic extension of $\mathbb{Q}(\sqrt{2})$ (it is easy to see that $\sqrt{3}$ does not belong to $\mathbb{Q}(2)$), hence $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}, \sqrt{3}) = 2 \dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}) = 4$.

A basis for the ideal $(\sqrt{2})$ is obtained by multiplying the basis above by $\sqrt{2}$; the basis elements stay independent as the multiplication map is injective. This gives the basis 2, $\sqrt{2}$, $2\sqrt{3}$, $\sqrt{6}$. In a similar way one finds the basis 3, $\sqrt{3}$, $3\sqrt{2}$, $\sqrt{6}$ for the ideal $(\sqrt{3})$.

EXERCISE 2.9 The proof is *mutatis mutandis* the same as the proof of Gauss' Lemma: Let $f(t) = \sum_i a_i t^i$ and $g(t) = \sum_i b_i t^i$ be two polynomials whose product lies in $\mathfrak{p}A[t]$;

i. e. all the products coefficients belong to $\mathfrak{p}$. Aiming for a contradiction assume that neither f nor g is member of $\mathfrak{p}A[t]$, and let a_{i_0} and b_{j_0} be coefficients of respectively f and g lowest degree that do not lie in $\mathfrak{p}$. Then the coefficient of degree $i_0 + j_0$ of the product equals

$$\sum_{i+j=i_0+j_0} a_i b_j = a_{i_0} b_{j_0} + \dots$$

If $i + j = i_0 + j_0$, either $i = i_0$ and $j = j_0$ or one of the inequalities $i < i_0$ or $j < j_0$ holds, and in the latter case it ensues from the minimality of i_0 and j_0 that the product $a_i b_j$ belongs to $\mathfrak{p}$; hence $a_{i_0} b_{j_0}$ lies there too, which is incompatible with a_{i_0} and b_{j_0} not doing so.

EXERCISE 2.10 Let $\phi\colon A \to B$ be the ring homomorphism and $\mathfrak{p} \subseteq B$ a prime ideal: That $ab \in \phi^{-1}(\mathfrak{p})$ means that $\phi(ab) = \phi(a)\phi(b) \in \mathfrak{p}$. Then either $\phi(a) \in \mathfrak{p}$ or $\phi(b) \in \mathfrak{p}$; that is, $a \in \phi^{-1}(\mathfrak{p})$ or $b \in \phi^{-1}(\mathfrak{p})$.

The most stupid example of the inverse image of a maximal ideal not being maximal, is the zero ideal in $\mathbb{Q}$ (which is maximal): it pulls back to the zero ideal in $\mathbb{Z}$ (which is not maximal).

Consider *e.g.* the ideal $(2)\mathbb{Z}$ in $\mathbb{Z}$ and extend it to the Gaussian integers $\mathbb{Z}[i]$. One has $(1+i)^2 = 2i$, and since i is a unit in $\mathbb{Z}[i]$, it holds that $(2)\mathbb{Z}[i] = (1+i)^2\mathbb{Z}[i]$.

EXERCISE 2.11 Observe that $\mathfrak{a} \nsubseteq \mathfrak{b}$ since the empty set is contained in every set. If $\mathfrak{a}\backslash\mathfrak{b}$ were contained in the union $\mathfrak{p}_1 \cup \dots \cup \mathfrak{p}_r$, it would hold that $\mathfrak{a} \subseteq \mathfrak{b} \cup \mathfrak{p}_1 \cup \dots \cup \mathfrak{p}_r$. From the Prime Avoidance Lemma it would follow that either $\mathfrak{a} \subseteq \mathfrak{b}$, which is not the case, or that $\mathfrak{a} \subseteq \mathfrak{p}_i$ for one i; contradiction.

EXERCISE 2.12 The claim about vector spaces is the following: if V and $V_1, \dots, V_r$ are vector subspaces of the vector space W over k and $V \subseteq V_1 \cup \dots \cup V_r$, then for at least one index i it holds that $V \subseteq V_i$. Certainly one may assume that the union is irredundant. We begin with treating the case $V = W$, that is $W = V_1 \cup \dots \cup V_r$, and we do that by induction on r. So let V' be the subspace generated by $V_2 \cup \dots \cup V_r$ Then $W = \subseteq V_1 \cup V'$; let $v \in V_1$ and $w \in W$, and consider the line $\lambda v + (1-l)v'$. It meets V_1 in v and V' in the v' but has infinitely many points, contradicting that $W = \subseteq V_1 \cup v'$. Hence either $V_1 = W$ and we are through, or $V' = W$ and we are thorugh by induction.

The general case is reduced to this case by assuming that V does not lie in any V_i and choosing points $V\backslash V_i$ and replacing W by the span W' of $v_1, \dots, v_r$ and V_i by $V_i \cap W'$.

EXERCISE 2.13 According to Proposition 2.19 about ideals in quotients the ideal generated by p_i equals $(p_i, p_1 \cdot \dots \cdot p_r)/(p_1 \cdot \dots \cdot p_r)$. By assertion *ii)* in the Isomorphism Theorem (Theorem 2.21 on page 37) it follows that $A/(p_i) \simeq \mathbb{Z}/(p_i, p_1 \cdot \dots \cdot p_r)\mathbb{Z} = \mathbb{Z}/p_i\mathbb{Z} = \mathbb{F}_{p_i}$. It is generally true that the additive group $\mathbb{Z}/n\mathbb{Z}$ is of order n so A has

$n = p_1 \cdot \ldots \cdot p_r$ elements, and hence the ideal (p_i) being the kernel of the canonical map $A \to \mathbb{Z}/p_i\mathbb{Z}$ has $np_i^{-1} = p_1 \cdot \ldots \cdot \widehat{p_i} \cdot \ldots \cdot p_r$ elements.

Finally, let e be the class of $p_1 \cdot \ldots \cdot \widehat{p_i} \cdot \ldots \cdot p_r$ in A. It is killed by p_i hence the ideal e generates (which equals the additive subgroup it generates as all elements in A are classes of integers) is a non-trivial factor of $\mathbb{Z}/p_i\mathbb{Z}$, hence it is isomorphic to $\mathbb{Z}/p_i\mathbb{Z}$ and has p_i elements.

Exercise 2.15 (*Primes in the Gaussian integers*)

a) That $\psi \circ \psi = 1$ amounts to $x^{p-1} = 1$ which is just Fermat's little theorem (or the facts that $\mathbb{F}_p^*$ has $p-1$ elements and that every group is killed by its order), and for the same reason ψ assumes values in μ_2. The kernel of ϕ obviously equals μ_2, and hence its image has $(p-1)/2$ elements, which all lie $\ker \psi$. Now, the equation $x^{(p-1)/2} = 1$ has at most $(p-1)/2$ roots, and consequently $\operatorname{im} \phi = \ker \psi$. Finally, we infer that the image of ψ has $(p-1)/((p-1)/2) = 2$ members, *i. e.* ψ is surjective.

b) This is now a real sweet piece of cake: it suffices to deside when $-1 \in \ker \psi$; that is, when $(-1)^{(p-1)/2} = 1$. And of course, this occurs precisely when $(p-1)/2$ is even; that is, when $p \equiv 1 \mod 4$.

c) The polynomial $x^2 + 1$ is irreducible over $\mathbb{F}_p$ precisely when it does not have a root in $\mathbb{F}_p$; that is, when $p \not\equiv 1 \mod 4$.

d) blabla

Exercise 2.18 Let $\mathfrak{a}$ be a non-zero ideal in A. Following the hint we let $a \in \mathfrak{a}$ be such that $\delta(a_0)$ is the least element in $\delta(\mathfrak{a}\backslash\{0\})$, which is a legitimate definition since W is well ordered so that every non-empty subset has a minimal element. We claim that $\mathfrak{a} = (a_0)$: indeed, for any $a \in A$, we may write $a = a_0 q + r$ with $\delta(r) < \delta(a_0)$, but evidently $r = a - a_0 q$ belongs to $\mathfrak{a}$, hence by minimality of a_0, it holds that $r = 0$ and thence $a = q a_0$.

Exercise 2.20 (*The Eisenstein integers*)

a) The set $\mathbb{Z}[\eta]$ is obviously closed under addition, and using that $\eta^2 = -\eta - 1$, the little calculation

$$\begin{aligned}(n + m\eta)(n' + m'\eta) &= nn' + (nm' + n'm)\eta + mm'\eta^2 = \\ &= nn' - mm' + (nm' + n'm - mm')\eta\end{aligned}$$

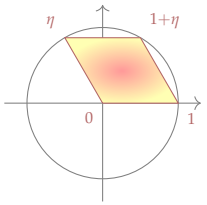
shows it is closed under multiplication as well.

b) Note that $\eta + \bar{\eta} = -1$ and $\eta\bar{\eta} = 1$, so if $x = n + m\eta$ is a unit, the integers n and m satisfy the equation

$$n^2 - nm + m^2 = 1, \tag{19.1}$$

which ensues from the identity $1 = |\eta|^2 = (n + m\eta)(n + m\bar{\eta})$. Assume $n, m \geqslant 0$. If $n > m$ it follows that $n = 1$ and $m = 0$, and symmetrically, $m > n$ implies that $m = 1$ and $n = 0$, and if $n = m$, they are both equal to one. The equality 19.1 also forces n and m to be of the same sign, hence we find the solutions ± 1, $\pm\eta$ and $\pm(1 + \eta)$. The group of units is the group μ_6 of sixth roots of unit; it is cyclic of order 6.

c) Consider a complex number $z = x + y\eta$ with x and y real and set $z_0 = x - [x] + (y - [y])\eta$; then $z = w + z_0$ with $w \in \mathbb{Z}[\eta]$ and z_0 in the convex hull of η, $1 + \eta$, 1 and 0, which is the depicts quadrilateral. The diameter of which is $\sqrt{3}$. Hence the distance of any complex number to $\mathbb{Z}[\eta]$ is less than $\sqrt{3}/2 < 1$. Next, assume given Eisenstein integers a and b and choose $q \in \mathbb{Z}[\eta]$ so that $|ab^{-1} - q| < 1$. Put $r = b(ab^{-1} - q)$. Then $a = bq + r$ and $|r| < |b|$.



d) Since the group of units is of order six, each non-zero element has six associates; hence each non-zero ideal has six generators.

EXERCISE 2.21 Assume that $ab \in \bigcap_{i\in I} \mathfrak{p}_i$, but $a \notin \bigcap_{i\in I} \mathfrak{p}_i$. Then there is an i_0 so that a does not belong to $\mathfrak{p}_{i_0}$, and thus $a \notin \mathfrak{p}_i$ for $i \geqslant i_0$ as the $\mathfrak{p}_i$ form a chain. It follows that $b \in \mathfrak{p}_i$ for $i \geqslant i_0$, and by consequence b lies in $\mathfrak{p}_i$ for all i the $\mathfrak{p}_i$'s forming a chain.

As to the union, let $ab \in \bigcup_{i\in I} \mathfrak{p}_i$ and assume that $a \notin \bigcup_{i\in I} \mathfrak{p}_i$. Then a does not belong to any of the $\mathfrak{p}_i$'s and hence b belongs every one of them.

Finally, the set Σ of prime ideals containing $\mathfrak{a}$ and contained in $\mathfrak{p}$ is trivially non-emty and every descending chain in Σ has a lower bound. Zorn then tells us that there is a minimal element.

EXERCISE 2.22 Let $\{\mathfrak{p}_i\}$ be a saturated chain of prime ideals connecting $\mathfrak{p}$ to $\mathfrak{q}$. Pick an element x in $\mathfrak{q}$ that does not lie in $\mathfrak{p}$ and consider the two sets $S = \{\, \mathfrak{p}_i \mid x \in \mathfrak{p}_i \,\}$ and $T = \{\, \mathfrak{p}_i \mid x \notin \mathfrak{p}_i \,\}$. Every prime ideal from T is strictly contained in every prime ideal from S. Since the original chain is saturated, and the union $\mathfrak{p}'$ of the $\mathfrak{p}_i$'s from T being a prime (Exercise 2.21 above) not containing x, the union $\mathfrak{p}$ lies in T. Similarly, the intersection $\mathfrak{q}'$ of the $\mathfrak{p}_i$'s that belong to S, is an element in S. It holds that $\mathfrak{p}' \subset \mathfrak{q}'$, and because the original chain is saturated, there can be no prime ideal lying between $\mathfrak{p}'$ and $\mathfrak{q}'$.

EXERCISE 2.23

a) The complement of one prime ideal $\mathfrak{p}$ is multiplicatively closed since if neither x nor y lies in $\mathfrak{p}$ neither their product does; and it is saturated since $\mathfrak{p}$ is closed under multilplication. Obviously the intersection of saturated multiplicative sets is a saturated multiplicative set, so the implication in one direction is proved. For the reciproque, assume that S is multiplicative and saturated and let $x \notin S$. Since S is saturate, from $yx \in S$ it would ensue that $x \in S$, and hence the principal ideal (x) is disjoint from S. By the Fundamental Existence Theorem (Theorem 2.49 on page 49) there is then a prime ideal $\mathfrak{p}$ containing x and disjoint from S.

b) Assume that x and y are non-zero divisor, $xya = 0$ would imply that $ya = 0$ since x is a non-zero divisors; hence $a = 0$ since y also is one, and we conclude that xy is a non-zero divisor, and so the non-zero divisors form a multiplicative set. It is saturated because if x is a zero divisor, there is an $a \neq 0$ $xa = 0$, and then obviously $xya = 0$ too, so that xy will be a zero divisor.

c) Follows directly from *a*) and *b*).

Exercise 2.25 Each prime ideal in $A/\mathfrak{p}$ is shaped like $\mathfrak{q}/\mathfrak{a}$ for a prime ideal $\mathfrak{q}$ in A containing $\mathfrak{p}$. Each prime ideal $\mathfrak{q}$ that strictly contains $\mathfrak{p}$ must meet $\{a^n\}$ by the maximality of $\mathfrak{p}$, and consequently, $\mathfrak{q}$ contains a; in other words, the class $[a]$ lies in $\mathfrak{q}/\mathfrak{a}$.

Exercise 2.27

a) Let $\{\mathfrak{a}_i\}$ be the chain, and let $\mathfrak{a} = \bigcup_i \mathfrak{a}_i$. Since A is a PID the ideal A is principal, say generated by a. Since a belongs to the union of the $\mathfrak{a}_i$'s, it belong to one of them; say $\mathfrak{a}_\nu$, thus $\mathfrak{a} = (a) = \mathfrak{a}_\nu$, which evidently ensures that $\mathfrak{a}_i = \mathfrak{a}_\nu$ for $i \geqslant \nu$.

b) Let $x_1, x_2, \ldots$ be irreducibles auch that $a = x_1 \cdot \ldots \cdot x_i a_i$ for each $i \in \mathbb{N}$; *i. e.* $a_i = x_{i+1} a_{i+1}$. Consider the principal ideals (a_i), which form an ascending chain. Hence by *a*) it holds that $(a_{\nu+1}) = (a_\nu)$ for some ν. Thus $a_{\nu+1} = b a_\nu$ and consequently $1 = b x_{\nu+1}$, and $x_{\nu+1}$ is a unit contradicting it is irreducible.

Exercise 2.28 If $\mathfrak{p}$ is the only prime ideal in A, the nil radical, being the intersection of all prime ideals, equals $\mathfrak{p}$. Hence each element in $\mathfrak{p}$ is nilpotent. The ideal $\mathfrak{p}$ is also the sole maximal ideal of A, hence all elements outside $\mathfrak{p}$ are invertible.

Assume then that all elements are either nilpotent or invertible. If $\mathfrak{p}$ is a prime ideal, it consists of nilpotent elements as members are not invertible, and of course, every nilpotent belongs to $\mathfrak{p}$. Hence $\mathfrak{p}$ equals the radical, and it is maximal as all elements outside $\mathfrak{p}$ are invertible.

What we have shown amount to the ring A having just one prime ideal if and only if $\sqrt{(0)}$ is a maximal ideal; that is, if and only if $A/\sqrt{(0)}$ is a field.

Exercise 2.31 Each number in A can be written as $x = n/m \cdot p_1^{\nu_1} \cdot \ldots \cdot p_r^{\nu_r}$ where m and n are integers relatively prime to each p_i, and hence is a unit in A, and where the

ν_i's are non-negative integers. It follows that x is invertible if and only if all the ν_i's are zero; hence each ideal $(p_i)A$ is maximal, and they are the only maximal ideals. Their intersection, the Jacobson radical, equals $(p_1 \cdot \ldots \cdot p_r)$.

And $A/(p_i)$ equals $Z/(p_i) = \mathbb{F}_{p_i}$, since the denominator n in the expression for x above, being relatively prime to p_i, is invertible mod p_i; *i. e.* there is a relation $1 = sn + rp_i$ with $r, s \in \mathbb{Z}$.

EXERCISE 2.34 The Jacobson ideal $J(A)$ is a principal ideal, say (f). If (x_i) are the infinitely many maximal ideals in A, we have $(f) \subseteq (x_i)$ for each i, and hence each x_i (which is irreducible) divides f. This is impossible in view of part **??** of Exercise 2.27.

EXERCISE 2.35 The ideals $\mathfrak{m}_a = (x_1 - a_1, \ldots, x_n - a_n)$ are maximal for every point $(a_1, \ldots, a_n)$ in $\mathbb{C}^n$, and a polynomial f that is contained in all the $\mathfrak{m}_a$'s, vanishes in the entire $\mathbb{C}^n$, hence is 0.

EXERCISE 2.40

a) For each index $j > 1$ there is relation $a_j + b_j = 1$ with $a_j \in \mathfrak{a}_1$ and $b_j \in \mathfrak{a}_j$ because $\mathfrak{a}_1$ and $\mathfrak{a}_j$ are comaximal. Developing the product $\prod_j (a_j + b_j)$, which equals one, one sees that all terms except $b = \prod_j b_j$ belong to $\mathfrak{a}_1$. Denoting their sum by a one finds $1 = a + b$ and it holds that $a \in \mathfrak{a}_1$ and $b \in \prod_{j>1} \mathfrak{b}_j$.

b) First we do the casd $r = 2$. Write $1 = a_1 + a_2$ with $a_1 \in \mathfrak{a}_1$ and $a_2 \in \mathfrak{a}_2$. If $x \in \mathfrak{a}_1 \cap \mathfrak{a}_2$ we see that $x = xa_1 + xa_2$ lies in $\mathfrak{a}_1\mathfrak{a}_2$ since both terms lie there. Induction on r finishes the point. We may assume $\mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r = \mathfrak{a}_2 \cap \ldots \cap \mathfrak{a}_r$. By *a*) $\mathfrak{a}_1$ and $\mathfrak{a}_2 \cdot \ldots \cdot \mathfrak{a}_r$ are comaximal and the claim follows by the case $r = 2$.

c) Induction on r: Let $\mathfrak{b}'_i = \prod_{j \neq i, i<r} \mathfrak{a}_i$, then $\mathfrak{b}'_1 + \ldots + \mathfrak{b}'_{r-1} = A$. Now after *a*) $\mathfrak{a}_r$ and $\mathfrak{b}_r$ are comaximale so that $\mathfrak{a}_r + \mathfrak{b}_r = A$. These two relations give $\mathfrak{a}_r = \mathfrak{a}_r\mathfrak{b}'_1 + \ldots + \mathfrak{a}_r\mathfrak{b}'_{r-1} = \mathfrak{b}_1 + \ldots + \mathfrak{b}_{r-1}$, hence $\mathfrak{b}_1 + \ldots + \mathfrak{b}_{r-1} + \mathfrak{b}_r = A$.

EXERCISE 2.41 It holds that $30 = 2 \cdot 3 \cdot 5$ and $6 + 10 - 15 = 1$ hence $[6]$, $[10]$ and $[-15]$ are orthogonal and equal one mod 5, 3 and 2 respectively. Similarly, $105 = 3 \cdot 5 \cdot 7$ and $15 + 21 - 35 = 1$ so that $[15]$, $[21]$, $[-35]$ will be a complete orthogonal set of idempotens; they are congruent one mod 7, 5 and 3 respectively.

EXERCISE 2.43 It is trivial that $\sqrt{\mathfrak{a}} \subseteq \sqrt{\mathfrak{a} + \mathfrak{n}}$. To prove the reverse inclusion, let x be an element in A so that $x^n = \sum_{1 \leqslant i \leqslant r} a_i y_i$ with $a_i \in \mathfrak{a}$ and $y_i \in \mathfrak{n}$. Then each y_i is nilpotent, and they being finite in number, there is an $m \in \mathbb{N}$ so that $y_i^m = 0$. Appealing to the binomial theorem, one may express a power x^N as a sum of terms, each having as factor a monomial $y_1^{\nu_1} \ldots y_r^{\nu_r}$ of degree N; so choosing $N \geqslant rm$ each term will vanish.

To the second question, observe that if $\mathfrak{n}$ is a locally nilpotent ideal in a ring A, then $\mathfrak{n}$ is contained in every prime ideal of A.

EXERCISE 2.46 A monomial $x^\mu y^\nu$ is of degree $\mu - \nu$ hence of degree zero if and only if $\mu = \nu$; *i. e.* it equals $(xy)^\mu$. In other words a polynomial is of degree zero precisely

when it is shaped like $f(xy)$ with f a polynomial . Hence R_0 equals the polynomial ring $k[xy]$.

If $n > 0$, the homogenous piece R_n equals $x^n k[x,y]$ whereas $R_n = y^n k[x,y]$ when $n < 0$.

Exercise 2.47 (*Homogeneous prime ideals*) Let $\mathfrak{p}$ be a homogenous prime and x and y two elements. Let $xy \in \mathfrak{p}$ and assume that $x \notin \mathfrak{p}$; we aim at showing that y lies in $\mathfrak{p}$ by contradiction, so assume as well that $y \notin \mathfrak{p}$. Let ν and μ be the highest degrees of the homogeneous component of x respectively y not belonging to $\mathfrak{p}$. Decompose $x = x'' + x_n + x'$ and $y = y'' + y_\mu + y'$ where x'' and y'' recollect the homogeneous terms of degree superior to ν respectively μ. So x'' and y'' both belong to $\mathfrak{p}$, and by definition $(x - x'')(y - y'') \in \mathfrak{p}$. Replacing x with $x - x''$ and y with $y - y''$ we may assume that x_ν and y_μ are the highest homogeneous term of x and y.

Now $x_\nu y_\mu$ is the term of highest degree of the product xy and so lies in $\mathfrak{p}$ because $\mathfrak{p}$ is homogeneous, but by construction neither x_ν nor y_μ lies there; contradiction.

The reverse implication is obvious.

Exercise 2.49 We shall write x for $(x_1, \ldots, x_r)$.

a) If the polynomial f is homogenous of degree d, all monomials occurring in f are of degree d so they are shaped like $M(x) = x_1^{\nu_1} \cdot \ldots \cdot x_r^{\nu_r}$ with the exponents satisfying $\nu_1 + \ldots + \nu_r = d$, and for such obviously $M(\alpha \cdot x) = \alpha^d \cdot x$ (regardless of k being finite or not).
For the reverse implication, write $f = \sum_i f_i$ with f_i homogeneous of degree i. Then $f(\alpha \cdot x) = \sum_i f_i(\alpha \cdot x) = \sum_i \alpha^i \cdot f_i(x)$. On the other hand, $f(\alpha \cdot x) = \alpha^d \cdot f(x)$ by assumptions. Equating the two expression we infer the relation

$$\sum_i (\alpha^d - \alpha^i) f_i(x) = 0.$$

Now, non-zero homogeneous polynomials of different degrees are linearly independent, so were f not homogeneous, at least one term of degree less than d, say i, would not vanish, and we could have deduced that $\alpha^d - \alpha^i = \alpha^i(\alpha^{d-i} - 1) = 0$. Hence $\alpha^{d-i} - 1 = 0$ with $d - i > 0$ and valid for all $\alpha \in k^*$. This equation obviously has only finitely many solutions, so k being infinite would be absurd. Consequently f is homogenous.

b) Assume that the ideal $\mathfrak{a}$ is invariant. We shall show that all homogeneous components of a member f of $\mathfrak{a}$ lie in $\mathfrak{a}$, and we intend to do that by induction on the number of homogeneous components. So let $f = \sum_{0 \leqslant i \leqslant n} f_i$ be the expansion of f into homogeneous components. Now, for all $\alpha \in k^*$ it holds that $f^\alpha = \sum_{i<n} \alpha^i f_i + \alpha^n f_n$ so that $f^\alpha - \alpha^n f = \sum_{0 \leqslant i < n} (\alpha^i - \alpha^n) f_i$ belongs to $\mathfrak{a}$. By induction it follows that $\alpha^i(1 - \alpha^{n-i}) f_i \in \mathfrak{a}$ and since k^* infinite we may choose

α so that $\alpha^{n-i} \neq 1$. Hence it holds that $f_i \in \mathfrak{a}$ for $i < n$, and consequently also $f_n \in \mathfrak{a}$.

If the ideal $\mathfrak{a}$ is homogeneous, it is generated by homogeneous elements and hence is invariant by a).

EXERCISE 2.51 (*Homogenization of polynomials*) We rely on Exercise 2.49. Identities between polynomial being formal, they can be check over any extension of the ground field, so we may safely assume that k is infinite: then, for $\alpha \in k^*$ we find

$$f^H(\alpha x_0, \dots, \alpha x_r) = (\alpha x_0)^d f(\alpha x_1/\alpha x_0, \dots, \alpha x_r/\alpha x_0) = \\ = \alpha^d x_0^d f(x_1/x_0, \dots, x_r/x_0) = \alpha^d f^H(x_1, \dots, x_r).$$

The second claim is trivial.

EXERCISE 2.52 (*Dehomogenization of polynomials*) We begin with analyzing the case when g does not have x_0 as factor. At least one of the homogeneous terms in g will then not have x_0 as factor and will survive unchanged when we put $x_0 = 1$; so g^D will be of the same degree d as g. A homogeneous term $x_0^{\nu_0} \cdot \dots \cdot x_r^{\nu_r}$ of g becomes $x_1^{\nu_1} \cdot \dots \cdot x_r^{\nu_r}$ in g^D, and in $(g^D)^H$ it reappear transformed into

$$x_0^d \cdot (x_1/x_0)^{\nu_1} \cdot \dots \cdot (x_r/x_0)^{\nu_r} = x_0^{\nu_0} \cdot x_1^{\nu_1} \cdot \dots \cdot x_r^{\nu_r},$$

because $d = \nu_0 + \nu_1 + \dots + \nu_r$. So in this particular case, when g does not have x_0 as factor, it holds true that $g = (g^D)^H$. The final remark is that if $g = x_0^s h$ with h without x_0 as factor, we find $g^D = h^D$, and consequently $(g^D)^H = h$; so the factor x_0^ν disappears, and has to be reintroduced to get the desired equality $g = x_0^s (g^D)^H$. It is then obvious to find examples with s arbitrary between 0 and d; *e.g.* just take $g = x_0^s x_1^{d-s}$.

EXERCISE 2.53 Let a be of degree one and assume that $a = bc$. Let b_μ and c_ν be the terms of highest degree in b and c respectively. Then $b_\mu c_\nu \neq 0$ and $\nu + \mu = 1$. Both μ and ν are non-negative, so one equals one and the other zero, say $\mu = 1$ and $\nu = 0$. It follows that $c \in A_0$ which is a field, hence c is invertible, and we conclude that a irreducible.

EXERCISE 2.58 (*Distinguished open sets*)

a) The closed set $V((f))$ is composed of the prime ideals $\mathfrak{p}$ so that $f \in \mathfrak{p}$ and clearly equals the complement of $D(f)$; hence $D(f)$ is open.
b) It obviously holds that $V(\mathfrak{a}) = \bigcap_{f \in \mathfrak{a}} V((f))$ hence the complement of $V(\mathfrak{a})$ satisfies $V(\mathfrak{a})^c = \bigcup_{f \in \mathfrak{a}} D(f)$.
c) If $\mathfrak{a}$ is generate by the f_i's, then $V(\mathfrak{a}) = \bigcap_i V((f_i))$ and $V(\mathfrak{a})^c = \bigcup_i D(f_i)$
d) Assume that the family $\{D(f_i)\}_{i \in I}$ cover Spec A. Consider the ideal $\mathfrak{a} = (f_i | i \in I)$ generated by the f_i's. Since $V(\mathfrak{a}) = \varnothing$, it holds that $\mathfrak{a} = A$. Write $1 = a_1 f_1 + \dots + a_r f_r$; and then Spec $A = D(f_1) \cup \dots \cup D(f_r)$ since no proper ideal can contain all the $f_1, \dots, f_r$.

Solutions for exercises in Chapter 3

EXERCISE 3.8 Let $\mathfrak{m} = (\pi)$ be the maximal ideal in A and let $k = A/\mathfrak{m}$. Choose a prime ideal $\mathfrak{p}$ in $A[t]$.

a) We follow the hint and consider the evaluation map $A[x] \to K$ that sends $f(x)$ to $f(\pi^{-1})$. Clearly $\pi x - 1$ lies in its kernel. Next, let $f(x)$ belong to the kernel and assume it is primitive. By standard long division in $K[x]$ one finds a relation $f(x) = g(x)(\pi x - 1) + r$ where $g \in K[x]$ and $r \in K$. Putting $x = \pi^{-1}$, yields $r = 0$, so that $f(x) = g(x)(\pi x - 1)$. The content of $\pi x - 1$ being one, one finds $1 = c_f = c_g$, and g belongs to $A[x]$.

b) Assume that $\mathfrak{p} \cap A = 0$. Then $\mathfrak{p}K[x]$ is proper and generated by an irreducible $g(x)$ which we may assume lies in A. The ideal $(g(x), \pi)$ is not proper, since $(g(x))$ is maximal, and it follows that $(g(x))$ in $k[t]$ is the entire ring, hence $g(x)$ is inverible mod π. If $\mathfrak{p} \cap A = \mathfrak{m}$ it holds that $\pi \in \mathfrak{p}$, and consequently $\mathfrak{p}/(\pi)$ is a proper prime ideal in $k[x]$; hence it is generated by an irreducible polynomial whose lift to $A[x]$ will be irreducible and together with π it will generate $\mathfrak{p}$.

EXERCISE 3.10 A factorization of q would persist over any field extension of k, so we may assume that k is algebraically closed. Suppose then that $\sum_{1 \leqslant i \leqslant n} x_i^2$ is not irreducible; since it is homogeneous of degree two, it will have two factors both being linear, say $f = \sum_i a_i x_i$ and $g = \sum_i b_i x_i$. All the a_i's (and the b_i's) must be different from zero, since if if it happened that $a_i = 0$, there would be no x_i^2 term in q. Absorbing $\sqrt{a_i}$ in x_i, we may assume that each $a_i = 1$ at the price of changing the coefficients of q, but no cross term will be introduced. Each cross term $x_i x_j$ in the product fg has the coefficient $a_i + a_j$. For any pair i, j of indices there is a third index k different from both. So we find $a_i = -a_j$, $a_j = -a_k$ and $a_i = -a_k$ which give $a_i = -a_i$, and hence $a_i = 0$ since the characteristic is not two.

EXERCISE 3.13 Let $f = f_1^{\nu_1} \cdot \ldots \cdot f_r^{\nu_r}$ we a factorization of f into irreducibles and such that no two f_i's are associates. Then $f = f^g = f_1^g \cdot \ldots \cdot f_r^g$. The ring A is assumed to be factorial, so each f_i^g is associated to a uniquely defined f_j^g. Hence G permutes the principal ideals (f_i), and thus maps into the symmetric group S_r. By assumption, G has no finite quotients, hence the image in S_r is trivial and each ideal (f_i) is invariant. It holds that $f_i^g = \chi g f_i$ with $\chi(g)$ a unit. When G acts trivially on the units, we find $(f_i^{gh} = (\chi(g) f_i)^h = \chi(g)^h \chi(h) f_i$; that is $\chi(gh) = \chi(g)\chi(h)$ because $\chi(g)^h = \chi(g)$.

EXERCISE 3.14 One finds $(2, i\sqrt{2k})^2 = (4, 2i\sqrt{2k}, -2k)$, but since $k = 2r + 1$ one has $2 = 2k - 4$ and $2 \in \mathfrak{p}^2$.

Assume then that f generates $\mathfrak{p}$; then $2 = af^2$ and it follows that $4 = N(2) = N(a)N(f)^2$. There are two possible cases which turn out to be impossible: either $N(f) = 2$ and $N(a) = 1$ which is impossible as $N(f) = x^2 + 2ky^2$, or $N(a) = 4$ and

$N(f) = 1$; that is, $a = \pm 2$ and $f = 1$ which is impossible since $\mathfrak{p}$ is a proper ideal.

In $\mathbb{Z}[i\sqrt{2k}]$ one has $2(k+2) = (2 + i\sqrt{2k}) \cdot (2 - i\sqrt{2k})$, and *e.g.* 2 is irreducible since $N(2) = 4$ and $x^2 + 2ky^2$ is larger then 4 unless $x = \pm 2$ and $y = 0$. Moreover, from $2a = (2 + i\sqrt{2k})$ it ensues that $4N(a) = 4 + 2k$ hence $2N(a) = 2 + k$ which is not the case since k is odd. It follows that 2 is not associated to either factor to the right.

EXERCISE 3.18 (*The ring of real trigonometric polynomials*)

a) You cannot get rid of a Y^2 term just by multiplication, so no polynomial $p(X)$ only depending on X can be a multiple of $X^2 + Y^2 - 1$, hence $\mathbb{R}[x]$ is polynomial ring. Clearly 1 and y generate $\mathbb{R}[x, y]$ over $\mathbb{R}[x]$, and if $f(x) + g(x)y = 0$, it follows that $f = g = 0$; indeed, a polynomial $f(X) + g(X)Y$ can not be a multiple of $X^2 + Y^2 - 1$, just look at the highest term in Y of a potensial product.

b) this is just Proposition 3.36 above.

EXERCISE 3.21 Answers: $1 + \sqrt{2}$, $2 + \sqrt{3}$ and $9 + 4\sqrt{5}$.

Solutions for exercises in Chapter 4

EXERCISE 4.1 Define $A \to B$ by sending an element a to $a \cdot 1_B$. The module axioms ensure this will be a map of rings; *e.g.* it repects multiplcation since aa' goes to $aa' \cdot 1_B$ and $aa' \cdot 1_B = a \cdot (a' \cdot 1B) = (a \cdot 1_B) \cdot (a' \cdot 1_B)$.

EXERCISE 4.2 Each $\phi\colon A \to M$ gives an element $\phi(1) \in M$. For each $m \in M$ one may define $\phi\colon A \to M$ by $a \mapsto am$. The two construction are evidenly mutually inverse.

EXERCISE 4.29 The hint says everything: Since $\det \overline{\Phi} \neq 0$, the determinant $\det \Phi$ does not belong to the maximal ideal of the local ring A; thence $\det \Phi$ is invertible, and by the adjunction formula Φ is invertible; indeed, $\Phi^{-1} = \det \Phi^{-1} \Phi^{\dagger}$.

In the first calculus courses we learned that a continuous function that does not vanish at a point, does not vanish in a vicinity. Here an analogy to that is playing and is applied to the determinant; working in a local ring is in a way working in an unspecified neigbourhood.

EXERCISE 4.30 (*Demystisufying Nakayama's lemma*) Let $e_1, \dots, e_r$ and $f_1, \dots, f_s$ be bases for E and F respectively and let the matrix Φ of ϕ be the $s \times r$-matrix (a_{ij}).

a) Since $\overline{\Phi}$ is surjective it holds that $r \geqslant s$, and one of the maximal minor (that is an $s \times s$-minor) of $(\overline{a}_{ij})$ is non-zero, After permuting basis elements if required, we may assume that the minor in question is minor (a_{ij}) with $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant s$; in other words the matrix to the restriction $\phi|_{E'}$ of ϕ to the submodule $E' = A_{e_1} \oplus \dots A_{e_s}$. By the previous exercise we conclude that $\phi|_{E'}$ is an isomorphism; hence ϕ is surjective.

b) Recall that a module of finite presentation a finite generated module so that the relations is finitely generated as well; in other words M is a cokernel of a map

between finite free modules; it lives in the sequence

$$E \longrightarrow F \longrightarrow M \longrightarrow 0$$

Solutions for exercises in Chapter 5

Solutions for exercises in Chapter 6

Exercise 6.12 Start with an expression $\tau = \sum_{1 \leqslant i \leqslant s} e_i \otimes f_i$, and proceede by induction on s. If the e_i's are linearly independent, there is nothing to prove, if not, one of the e_i's is a linear combinations of the others, say $e_s = \sum_{1 \leqslant j \leqslant s-1} \alpha_j e_j$. Substituting this in the expression for τ yields $\tau = \sum_{i \leqslant s-1} e_i \otimes f_i - \sum_{j \leqslant s-1} \alpha_j e_j \otimes f_s = \sum_{i \leqslant s-1} e_i \otimes (f_i - \alpha_i f_s)$, and we are through by induction.

Solutions for exercises in Chapter 7

Exercise 7.3 Consider the localization $\Sigma^{-1}A[t_1, \ldots, t_n]$, which is a subring of the fraction field of $A[t_1, \ldots, t_n]$. It consists of fractions shaped like $b^{-1}f(t_1, \ldots, t_n)$ where $b \in A \backslash \{0\}$ and $f \in A[t_1, \ldots, t_n]$. All these are elements in $K[t_1, \ldots, t_n]$. On the other hand, if* $f = \sum_\alpha a_\alpha b_\alpha^{-1} t^\alpha \in K[t_1, \ldots, t_n]$ and $b = \prod_\alpha b_\alpha$, it holds that $bf \in A[t_1, \ldots, t_n]$ and so $f = b^{-1}(bf) \in \Sigma^{-1}A[t_1, \ldots, t_n]$.

*Here we use multi-index notation

Exercise 7.5 As $1/2 = 5 \cdot 1/10$ and $1/5 = 2 \cdot 1/10$, clearly $\mathbb{Z}[1/2, 1/5] \subseteq \mathbb{Z}[1/10]$, and because $1/10 = 1/2 \cdot 1/5$, the reverse inclusion also holds true.

Exercise 7.6 Let $S \subseteq \mathbb{Z}$ be the subset $S = \{\, n \mid 1/n \in A \,\}$. It is clearly multiplicatively closed. If $a/m \in A$ with $(a, m) = 1$, we may find integers x and y such that $xa + ym = 1$, which upon divison by m gives $m^{-1} = xam^{-1} + y$; hence $m \in S$ and ensues that $A = S^{-1}\mathbb{Z}$.

Exercise 7.7 The product of two units is obviously a unit. If $S = A^*$, the identity map id_A takes elements in S to units, and induces a map $\phi \colon A_S \to A$ such that $\phi \circ \iota_S = \mathrm{id}_A$. It ends $\iota(s)^{-1}a$ to $s^{-1}a$ so it is clearly an isomorphism.

Exercise 7.8 This is just Exercise 7.3 applied with $A = k[x_1, \ldots, x_r]$ and $t_i = x_{i+r}$.

Exercise 7.9 The group of units A^* is saturated since if xy is invertibel both x and y are. And of course, if $xy \in \{1\}$, by definition x and y are units.

Exercise 7.10 That the both the odd and the even numbers form multiplicative sets is obvious. In the case of the even integers, the saturation will be the multiplicative set $\mathbb{Z}\backslash\{0\}$. Indeed, if $n \neq 0$, obviously $2n$ is even, and hence n belongs to the saturation. In the case of odd integers the saturation will be generated by -1 and all odd primes: any odd numbers is a product of such. If $[a] = 1$ and $[b] = 1$ clearly $[ab] = 1$ (where $[x]$ denotes the class of x in $\mathbb{Z}/p\mathbb{Z}$). The saturation will be the set of integers invertible mod

p; that is, those on the form $ap + b$ where $(b, p) = 1$. Indeed, $[ab] = 1$, obviously $[a]$ is invertible in $\mathbb{Z}/p\mathbb{Z}$.

EXERCISE 7.11 (T) he saturated multiplicative sets are those generated by -1 and an arbitrary set of primes. Indeed, if S is saturated, and $n \in S$ any prime factor p of n belongs to S. Hence, the all elements of S are products of primes belonging to S, and of course $-1 \in S$ as well ($(-1)(-1) = 1 \in S$). In a general A, the same argument gives that the saturated multiplicative sets are the sets generated by the units A^* and any set of irreducible (equivalently prime) elements from A.

EXERCISE 7.12 Consider the set $S = \bigcup x + \mathfrak{a}$ where the union extends over all $x \in A$ whose class $[x] \in A/\mathfrak{a}$ is invertible. It is clearly multiplicative and saturated, it contains $1 + \mathfrak{a}$, and saying $(x + a)(y + a') = 1 + aa'$ is precisely to say that x (and y) is invertible mod $\mathfrak{a}$. Hence S is saturation of $1 + \mathfrak{a}$.

Solutions for exercises in Chapter 8

EXERCISE 8.11 The A-algebra B is free from torsion being an integral domain, and hence it is free according to Theorem 8.30. Every ideal $\mathfrak{a}$ is a torsion free A-module as well, hence free of the same rank as B as $\mathfrak{a} \otimes_A K = B \otimes_A K$ where K is the fraction field of A.

EXERCISE 8.14 The salient case is the 2×2-case. As a preparation consider any matrix

$$D = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

in $\mathrm{Gl}(2, k[x, x^{-1}])$ with $f_{11} \neq 0$ and let $c_D = \deg f_{11}$. Moreover if $f_{21} \neq 0$, let $d_D = \deg f_{21}$. We contend that if $f_{21} \neq 0$ and $d_D \leqslant c_D$, there is a matrix C in $\mathrm{Gl}(2, k[x])$ so that either $c_{CD} < d_D$ or has vanishing upper left corner. Indeed, write $f_{11} = \sum_{i \leqslant c} a_i x^i$ and let $g = \sum_{d_D \leqslant i \leqslant c_D} a_i x^{i - d_D}$. Then if α is the leading coefficient of f_{21}, the matrix

$$C = \begin{pmatrix} 1 & -\alpha^{-1} g \\ 0 & 1 \end{pmatrix}$$

will do.

Now chose D among the matrices shaped like CDC' so that $f_{11} \neq 0$ and such that f_{11} is of minimal degree. We claim that $f_{21} = 0$. If not, after the observation above we have $d_D > c_D$. Switching rows of D to obtain D', $c_{C'} > d_{C'}$ and by the process above, we obtain a matrix D'' with $c_{D''} < c_{D'} = c_D$; contradiction. So D is diagonal.

We can assume that $\det D = 1$; and as the group of units in $k[x, x^{-1}]$ equals $\{ \alpha x^a \mid a \in \mathbb{Z}, \alpha \in k^* \}$, it takes the form

$$\begin{pmatrix} x^a & p(x, x^{-1}) \\ 0 & x^{-a} \end{pmatrix}$$

It suffices to successively get rid of each term of p by multiplying D by a matrix from $\mathrm{Gl}(2, k[x^{-1}])$ from the right: so consider a matrix

$$D = \begin{pmatrix} x^a & x^s \\ 0 & x^{-a} \end{pmatrix}$$

If $s > a$ use the right matrix below and if $s < a$ the left one will do:

$$\begin{pmatrix} 1 & x^{s-a} \\ 0 & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & -x^{a-s} \end{pmatrix}$$

The general case of an $r \times r$-matrix is done ny induction on r; so we assume that D is diagonal except for the first row and the first column; *i. e.* it has the shape

$$\begin{pmatrix} x^{a_1} & * & \dots & \dots & * \\ * & x^{a2} & 0 & \dots & 0 \\ * & 0 & x^{a_3} & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ * & 0 & \dots & \dots & x^{a_r} \end{pmatrix}$$

Finally, by the 2×2-case one successively forcesthe i-th member of the first row and the i-th member of the first column pairwise to zero.

Solutions for exercises in Chapter 9

Exercise 9.1 Consider each addend M_i of the sum $M = \bigoplus_{i \in I} M_i$ as a submodule of M. Suppose that the index set I is infinite, equip the index set I with a linear order and let $N_j = \bigoplus_{i<j} M_i$. Then $\{N_j\}_{j \in I}$ is an ascending chain which does not stabilize; hence M is not Noetherian. Submodules of Noetherian modules are Noetherian, hence if M is Noetherian, so is each M_i.

Exercise 9.2 The ring of integers $\mathbb{Z}$ is a PID and PID's are Noetherian. In the module $\mathbb{Z}_{p^\infty}$ one has the sub $\mathbb{Z}$-modules $(p^{-i}) = \{\, [ap^{-1}] \mid a \in \mathbb{Z} \,\}$. They form an ascending chain of submodules, but it does not stabilize: if $p^{-i-1} = ap^{-i} + b$ with $a, b \in \mathbb{Z}$, we would have $p^{-1} = a + bp^i \in \mathbb{Z}$, which is absurd.

Now, $\mathbb{Z}$ is not Artinian since *e.g.* $\{(p^i)\}_{i \in \mathbb{N}}$ is descending chain that does not stabilize (p being any integer except ± 1).

Every element x in $\mathbb{Z}_{p^\infty}$ lies in some $(p^{-\nu})$ with $\nu > 0$; indeed, write x as a finite sum $x = \sum_i a_i p^{-i} = bp^{-\nu}$ where $\nu = \min i$. Hence the only submodules of $\mathbb{Z}_{p^\infty}$ are the cyclic modules $(p^{-i})\mathbb{Z}_{p^\infty}$. The submodule $(p^{-i})\mathbb{Z}_{p^\infty}$ is killed by p^i but not by p^{i-1} and it is therefore isomorphic to $\mathbb{Z}/p^i\mathbb{Z}$, in which there clearly is no infinite descending chain.

EXERCISE 9.3 The first assertion is clear: any B-submodule of M is an A-submodule, so any ascending chain of B-submodules is a chain of A-submodules and stabilizes as A is Noetherian. Examples abound. A simple one is furnished by the extension $\mathbb{Z} \subseteq \mathbb{Q}$. Obviously $\mathbb{Q}$ is Noetherian as $\mathbb{Q}$-module (every finite dimensional vector space over a field is), but clearly $\mathbb{Q}$ is not a Noetherian $\mathbb{Z}$-module: if it were, it would be finitely generated by the Main Theorem, and the common denominator of the elements from a finite generating set would serve at a denominator for all rational numbers.

Finally, suppose that ϕ is surjective. Then any A-submodule N is a B-module: if $x \in B$, is of shape $x = \phi(y)$ it holds that $xN = yN$, and the assertion follows since chains of B-modules stabelizes.

EXERCISE 9.4 Simple module are Noetherian (the two sole submodules are both finitely generated), and the claim in the exercise follows directly from Proposition 9.7 on page 231.

EXERCISE 9.5 The implication *ii)* $\Rightarrow$ *i)* is obvious. For the reverse, we mimic the argument in the Main Theorem: if no maximal element is found in Σ, any element has a strict subset lying in Σ, and with this one recursively constructs a strictly descending infinite chain of submodules from Σ; by consequence, M is not Artinian.

EXERCISE 9.6 Let n be an integer from A such that $1/n \notin A$, and consider the principal ideals $(n^{-i}t)\mathbb{Q}[t]$. They form a strictly increasing chain of ideals showing that A is not Noetherian: that the chain is increasing is clear, and if $n^{-i+1t} \in (n^{-i}t)\mathbb{Q}[t]$, one has $n^{-(i+1)}t = f(t)n^{-i}f$ for some polynomials $f(t)$. Cancelling $n^{-i}t$ gives $f(t) = n^{-1}$, which contradicts the assumption that $f(0) \in A$. Hence it is also strictly increasing.

EXERCISE 9.17 Let x be the generator of the maximal ideal.

a) Assume that a and b are elements that do not belong to $\bigcap_i \mathfrak{m}^i$ but whose product does. Then we may write $a = \alpha x^i$ and $b = \beta x^j$ with i and j maximal and consequently α and β will be units. It holds that $ab = \gamma x^{i+j+1}$ for some γ, hence $\alpha\beta x^{i+j} = \gamma x^{i+j+1}$ from which it ensues that $(\alpha\beta - \gamma x)x^{i+j} = (0)$. Now, $\alpha\beta - \gamma x$ is a unit ($\alpha\beta$ is a unit and $\gamma x \in \mathfrak{m}$) so that $x^{i+j} = 0$ and x is nilpotent.
Assume next that $\mathfrak{p}$ is a prime ideal contained in $\bigcap_i \mathfrak{m}^i$, but which does not contain x. Any element $a \in \mathfrak{p}$ is shaped like $a = \gamma x$, and since $x \notin \mathfrak{p}$, it follows that $\gamma \in \mathfrak{p}$. Hence $\mathfrak{p} = x\mathfrak{p}$, and an obvious induction shows that $\mathfrak{p} \subseteq \bigcap_i \mathfrak{m}^i$.

b) If $\bigcap_i \mathfrak{m}^i = (0)$, the set of natural numbers i so that $\mathfrak{a} \subseteq \mathfrak{m}^i$ is not empty (1 lies there) and is not the entire $\mathbb{N}$ (the ideal $\mathfrak{a}$ is non-zero); hence there is a maximal i such that $\mathfrak{a} \subseteq \mathfrak{m}^i$. We claim that $\mathfrak{a} = (x^i)$, and it suffices to show that $x^i \in \mathfrak{a}$: since $\mathfrak{a} \nsubseteq \mathfrak{m}^{i+1}$, there is an element $a \in \mathfrak{a}$ not in $\mathfrak{m}^{i+1}$, which must be shaped like $a = \alpha x^i$ with α not in $\mathfrak{m}$; hence α is a unit and $x^i = \alpha^{-1}a \in \mathfrak{a}$.

c) When A is Noetherian, Krull's Intersection Theorem yields that $\bigcap_i \mathfrak{m}^i = (0)$ (or even Nakayama Classic suffices since $x(\bigcap_i \mathfrak{m}^i) = \bigcap_i \mathfrak{m}^i$ by a)). If $\bigcap_i \mathfrak{m}^i = 0$, all

ideals are principal by b), in particular they are finitely generated.

EXERCISE 9.20 To ease the notation, we shall identify N with its image in M. Let $M_{\nu-1} \subset M_\nu$ be one of the inclusions in a composition series $\{M_i\}$ for M. Consider the intersection $N \cap M_\nu$. As there is no submodule lying strictly between $M_{\nu-1}$ and M_ν, there are two possibilities: Either $M_\nu \cap N = M_\nu$, and in that case $M_{\nu-1} \cap N = M_{\nu-1} \subset M_i = M_i \cap N$, and $M_\nu \cap N / M_{\nu-1} \cap N$ is simple. Or $M_\nu \cap N \subset M_{\nu-1}$, in which case $M_\nu \cap N = M_{\nu-1} \cap N$ and a repetition appears.

Solutions for exercises in Chapter 10

EXERCISE 10.2 Consider $k[x_1, \ldots, x_n]$ as the graded ring $A[x_1, \ldots, x_r]$ where we let $A = k[x_{r+1}, \ldots, x_n]$ be the part of degree zero and each x_i with $i \leqslant r$ is of degree one. Then $\mathfrak{p}^m$ consists of elements whose lowest term is of degree m or more.

Develop f in homogeneous components $f = f_0 + f_1 + \ldots$. If $f_0 = 0$, the class of f mod $\mathfrak{p}^m$ is nilpotent—in fact f^m is of degree at least m so that $f^m \in \mathfrak{p}^m$. So assume that $f_0 \neq 0$, but that $fg = 0 \bmod \mathfrak{p}^m$; that is, all its homogenous components are of degree at least m, and that $g \notin \mathfrak{p}^m$; *i. e.* $\deg g < m$. Develop g in homogeneous components: $g = g_s + g_{s+1}$ with $g_s \notin \mathfrak{p}^m$; *i. e.* of degree less than m. We find that

$$fg = f_0 g_s + f_1 g_s + f_0 g_{s+1} + \text{ higher terms } .$$

For degree reasons—that is, since $s < m$, it follows that $f_0 g_s = 0$, and hence $g_s = 0$; contradiction.

EXERCISE 10.4 Note that $(x^2, xy) = (x^2, x(y + a))$, changing coordinates $y \to y + a$ the general case follows from the case $a = 0$, which is just Example 10.4. The ideals $(x, , y + a)$ are different for different a.

EXERCISE 10.5 Localizing in x, we find $\mathfrak{a}A_x = A_x$, due to x^3 being a generator, and inverting y, we obtain $\mathfrak{a}A_y = (x)A_y$. Hence we conclude that

$$\mathfrak{a} = (x) \cap \mathfrak{q}$$

where $\mathfrak{q}$ is (x, y)-primary. A good guess would be

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

The inclusion $\subseteq$ is clear, and we know that $\mathfrak{a}$ is monomial, så it suffices to see that any monomial from theright hand side belongs to $\mathfrak{a}$. It must have x as a factor, hence is either y^3x, y^2x^2 or x^3 as a factor; and that's it!

One could also resort to the JCO-algoritm;

$$\begin{aligned}\mathfrak{a} &= (x^3, y^2x^2, \underline{y^3x}) \\ &= (x) \cap (x^3, \underline{y^2x^2}, y^3) \\ &= (x) \cap (x^3, y^2) \cap (x^2, y^3)\end{aligned}$$

and coalescing the two last ideals results in (x^3, x^2y^2, y^3) (again, enough to check monomials).

EXERCISE 10.6 The only challenge is to determine the primary decomposition of the square

$$\mathfrak{a}^2 = (y^2z^2, y^2x^2, x^2z^2, xyz^2, xzy^2, yzx^2).$$

Localizing in x we find that $\mathfrak{a}^2A_x = (y,z)^2A_x$, and since $(y,z)^2$ is primary and by symmetry we conclude that

$$\mathfrak{a}^2 = (x,y)^2 \cap (x,z)^2 \cap (y,z)^2 \cap q$$

where $\mathfrak{q}$ is (x,y,z)-primary. Note that xyz lies in the intersection of the three squares to the right; it is not a member of $\mathfrak{a}^2$, but x^2yz, xy^2z and xyz^2 are, so get rid of xyz we try with $\mathfrak{q} = (x^2, y^2, z^2)$. We claim that

$$\mathfrak{a}^2 = (x,y)^2 \cap (x,z)^2 \cap (y,z)^2 \cap (x^2, y^2, z^2),$$

and indeed, this holds true. The inclusion $\subseteq$ is easy, for the other, it suffices to verify it for monomials. A monomial lying in the right hand side must contain a square of one of the variables, say it contains x^2. Lying in $(y,z)^2 = (y^2, yz, z^2)$ it must contain one of the monomials y^2x^2, yzx^2 or z^2x^2, but all these belong to $\mathfrak{a}^2$.

EXERCISE 10.13

$$\begin{aligned}
&= (\underline{x^\alpha y}, z^\beta x, y^\gamma z) \\
&= (x^\alpha, \underline{z^\beta x}, y^\gamma z) \cap (\underline{z^\beta x}, y) \\
&= (x^\alpha, z^\beta, \underline{y^\gamma z}) \cap (x, \underline{y^\gamma z}) \cap (z^\beta, x) \cap (x, y) \\
&= (x^\alpha, z^\beta, y^\gamma) \cap (x^\alpha, z) \cap (x, y^\gamma) \cap (x, z) \cap (z^\beta, x) \cap (x, y) \\
&= (x^\alpha, z^\beta, y^\gamma) \cap (x^\alpha, z) \cap (x, y^\gamma) \cap (z^\beta, x)
\end{aligned}$$

Solutions for exercises in Chapter 11

EXERCISE 11.1 The maximal ideal of $A_{(x)}$ is the principal ideal $(x)A_{(x)}$. According to Exercise 9.17 on page 244 the powers $(x^i)A_{(x)}$ are all non-zero ideals in $A_{(x)}$, and obviously, none of these are prime. So there is just one chain in $A_{(x)}$, namely $(0) \subseteq (x)A_{(x)}$ and the dimesion is one.

EXERCISE 11.2 Let $K = k(z)$, and $S = \{ f(z) \mid f \text{ poly } \}$: Then $S^{-1}k[x,y,z] = K[x,y]$. Hence $R_\mathfrak{p} = K[x,y]_{(x,y)}$. We argued in Paragraph 11.4 on page 289 that $\dim K[x,y] = 2$, so $\dim R_\mathfrak{p} \leqslant 2$. And there is the obvious chain $(0) \subseteq (y) \subseteq (x,y)$, so $\dim R_\mathfrak{p} = 2$.

EXERCISE 11.3 Indeed, if $\mathfrak{q} \subset \mathfrak{p}$ is a given pair, there is minimal prime $\mathfrak{q}_0$ contained in $\mathfrak{q}$ and a maximal ideal $\mathfrak{m}$ containing $\mathfrak{p}$. Fix a saturated chain ascending from $\mathfrak{q}_0$ to $\mathfrak{q}$ and

one ascending fra $\mathfrak{p}$ to $\mathfrak{m}$. Then any saturated chain from $\mathfrak{q}$ to $\mathfrak{p}$ may be embedded in one from $\mathfrak{q}_0$ to $\mathfrak{m}$ by concatenating it with the two fixed chains, and all such chains have the same length.

Solutions for exercises in Chapter 12

EXERCISE 12.1 An element $x = a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ has a minimal equation

$$x^2 - 2ax + a^2 - b^2 d = 0$$

and it is integral if and only if the coefficients are integers; that is, $n = 2a \in \mathbb{Z}$ and $a^2 - b^2 d \in \mathbb{Z}$. It follows that $4b^2 d \in \mathbb{Z}$ which yields that $2b \in \mathbb{Z}$ (write $b = zy^{-1}$ with $z, y \in \mathbb{Z}$; then $4z^2 d = y^2$, and each factor in y must be a factor in $2z$ as d is square-free). It follows that $(2a)^2 \equiv (2b)^2 d \mod 4$.

If $d \equiv 1 \mod 4$, it follows that $(2a)^2 \equiv (2b)^2$ and $2a$ and $2b$ have the same parity. If they both are odd, one has $a + b\sqrt{d} = 2^{-1}(2a + 2b\sqrt{d}) = w + 2^{-1}(1 + \sqrt{d})$ with $w \in \mathbb{Z}[\sqrt{d}]$, and if they are even clearly both a and b are integers.

If $d \not\equiv 1 \mod 4$, it ensues, since squares are either 1 or 0 mod 4, that $(2a)^2 \equiv (2b)^2 \equiv 0 \mod 4$ and hence that $2a$ and $2b$ both are even, so that $a, b \in \mathbb{Z}$.

EXERCISE 12.4 Let $f \in B[x]$ be an element and write $f = b_n x^n + \ldots + b_1 x + b_0$. Since B is supposed to be integral over A, the extension $C = A[b_0, \ldots, b_b]$ is a finite faithful module over A, and it esnues that $C[x]$ is a finite faithful module over $A[x]$; indeed, it has the same genrators as C. And of course $f \in C[x]$, hence it is integral over A.

EXERCISE 12.8

a) Define $g(xy^{-1}) = g(x)g(y)^{-1}$.

b) The crucial observation is that $yh(x) = h(y)x$ for all h; hence we find

$$h(y)\prod_{g\neq e} hg(x) = h(y)x\prod_{g\neq h,e} g(x) = yh(x)\prod_{g\neq h,e} g(x) = y\prod_{g\neq e} g(x).$$

Then as in the hint $x/y = \left(x\prod_{g\neq e} g(x)\right)\left(y\prod_{g\neq e} g(x)\right)^{-1}$, and it follows that $K^G = L$.

c) The polynomial $P(t) = \prod_{g\in G}(t - g(x))$ in $B[t]$ where t is a variable, is clearly invariant under G since G just permutes the factors, hence each coefficient is invariant and lies in A (different powers of t are linearly independent over B). The polynomial P is monic and $P(x) = 0$ since $e \in G$.

d) If $x \in L$ is integral over A, it is *a priori* integral over B, so if B is integrally closed in K it lies in B, and being invariant, it therefore belongs to A.

EXERCISE 12.9 Let $\mathfrak{q}$ and $\mathfrak{q}'$ be primes both lying over $\mathfrak{p}$ and assume that $\mathfrak{q}$ is not equal to any of of the translates $g(\mathfrak{q}')$. By prime avoidance there is then an $x \in \mathfrak{q}$ not

lying in any of the $g(\mathfrak{q})'$. Consider $\prod_{g\in G} g(x)$. It is invariant and lies in $A\cap\mathfrak{q}$, but $A\cap\mathfrak{q}=\mathfrak{p}=A\cap\mathfrak{q}'$, so $\prod_{g\in G} g(x)$ also lies in $\mathfrak{q}'$. It follows that $g(x)\in\mathfrak{q}'$ for some g, hence $x\in g(\mathfrak{q}')$. Contradiction.

EXERCISE 12.11 By 12.29 if $\mathfrak{m}$ is a maximal ideal in B, the intersection $\mathfrak{m}\cap A$ is maximal is maximal in A; hence $\bigcap_{\mathfrak{n}\subseteq A}\mathfrak{n}\subseteq\bigcap_{\mathfrak{m}\subseteq B}\mathfrak{n}\cap A$ blabla

EXERCISE 12.12 Let $\mathfrak{q}_1$ and $\mathfrak{q}_2$ be two prime ideals in B lying over the same prime ideal $\mathfrak{p}$ in A. A power x^{p^ν} of each element x in $\mathfrak{q}_1$ lies in A, hence it belongs to $\mathfrak{q}_1\cap A$, which equals $\mathfrak{q}_2\cap A$. It follows that $x\in\mathfrak{q}_2$, and we may conclude that $\mathfrak{q}_2\subseteq\mathfrak{q}_1$. By symmetry the reverse inclusion holds as well, and $\mathfrak{q}_1=\mathfrak{q}_2$. It follows that π is injective. Now, B is integral over A and π is therefore surjective and closed, hence it is a homeomorphism.

Solutions for exercises in Chapter 13

EXERCISE 13.4 According to Proposition 3.30 on page 77 the maximal ideal $\mathfrak{m}$ is of the form $(f(x),g(y))$ with f and g irreducible. Now, there is an isomorphism $k[x,y]/(f(x),g(y))\simeq k[x]/(f(x))\otimes_k k[y]/(g(y))=E_f\otimes_k E_g$, and the tensor product $E_f\otimes_k E_g$ is a field precisely when E_f and E_g are linearly disjoint.

EXERCISE 13.7

a) Assume there only are finitely many irreducibles $f_1,\dots,f_r$ and consider $g=1+f_1\cdot\ldots\cdot f_r$. Clearly none of the f_i's divides g, but because $k[t]$ is a UFD, g is a product of irreducibles, contradiction.

b) Assume that $k(t_1,\dots,t_n)$ is finitely generated and let $f_1,\dots,f_r$ be generators. If g_i is the denominator of f_i, the deominator of every element is of the form $g_1^{\nu_1}\cdot\ldots\cdot g_n^{\nu_n}$ which have only finitely many different irreducible factors, but according to *a*) there are infinitely many irreducible elements which of course can occur as denominator (simply as $1/f$).

c) We keep the notation in the hint, but choose a basis e_i so that $e_1=1$. Each monomial $f^\alpha=f_1^{\alpha_1}\cdot\ldots\cdot f_r^{\nu_r}=\sum c_{\alpha,i}e_i$ where $c_{\alpha,i}$ is a polynomial in the a_{ij}'s and the b_{ijl}'s. This follows by induction of the degree of f:

$$f_j\cdot f^\alpha=\sum_{i,s}a_{ij}c_{\alpha,s}e_ie_s=\sum_l(\sum_{i,s}a_{ij}c_{\alpha,s}b_{isl})e_l$$

Hence every element of L is of the same form, and it belongs to K when all terms vanish except the the one corresponding to e_1.

d) Let $\mathfrak{m}$ be a maximal ideal in $k[x_1,\dots,x_n]$. Then $L=k[x_1,\dots,x_n]/\mathfrak{m}$ is of finite type over k, and if $\operatorname{trdeg}_k L=r$, it will be finite over a subfield K shaped like $K=k(t_1,..,t_r)$. Hence K will be of finite type, which is not the case when $r>0$.

EXERCISE 13.9 Let $A=k[x,y]$ with constituting relation $xy=y(y-1)=0$ Then $\dim A=1$, and in A_y one has $x=y-1=0$; so $A_y\simeq k$ and is of dimension 0.

Exercise 13.12 If $A' \subseteq A$ and $B' \subseteq B$ are subalgebras, one has the sequence of inclusions $A' \otimes_k B' \subseteq A' \otimes_k B \subseteq A \otimes_k B$; the first because A' is flat over k and the second because B is (remember, all algebras over fields are flat). Assume then that $f = \sum_{i \in I} a_i \otimes b_i$ and $g = \sum_{j \in J} c_j \otimes d_j$ are two elements in $A \otimes_k B$. The idea is to replace A by the subalgebra $A' = k[a_i, c_j | i \in I, j \in J]$ and B by $B' = k[b_i, d_j | i \in I, j \in J]$, which both are of finite type over k and both f and g are contained in their tensor product. The proposition gives that $A' \otimes_k B'$ is a domain (resp. reduced) when A and B are. Since $f \in A'$ and $g \in B'$, it follows that $fg \neq 0$ (resp. $f^n \neq 0$) in $A' \otimes_k B'$ when f and g are non-zero, and hence $fg \neq 0$ (resp. $f^n \neq 0$) in $A \otimes_k B$ in view of the inclusions above.

Exercise 13.13

a) That the map extends is just saying that the assignment $e \otimes f$ is bilinear and it is a ring map as $e' \otimes f' \cdot e \otimes f = ee' \otimes ff'$ which is sent to $ee'ff' = (ef) \cdot (e'f')$.

b) Both E and F being algebraic, EF will be algebraic; indeed, if $x_1, \ldots, x_r \in E$ and $y_1, \ldots, y_s \in F$, it follows that $k[x_1, \ldots x_r, y_1 \ldots y_s]$ is integral over k (the integral closure is a ring) and hence it is a field by the basic lemma 12.28 on page 327. It follows that the union of all such extensions, which is a directed union, is a field, and it must be equal to the compositum EF. This shows that each element in EF is shaped like a finite sum $\sum_i x_i y_i$ with $x_i \in E$ and $y_i \in F$ (not the same as the ones above) and consequently ϕ is surjective.
Each element z in the tensor product may be written as a finite sum $z = \sum e_i \otimes f_i$ with e_i's linear independent. If E an F are linearly disjoint, they stay independent over F in EF, and $\phi(z) = \sum f_i e_i$ is non-zero unless all the f_i's vanish, and then of course $z = 0$. On the other hand, assume that ϕ is injective and that the e_i's are linearly independent over k; they can then be extended to a k-basis for E. If there is a relation $\sum g_j e_i = 0$, it follows as ϕ is injective, that $\sum g_i \otimes f_i = 0$. If $\{f_j\}$ is a basis for f, each g_i is of shape $g_i = \sum_j a_{ij} f_j$ so that

$$0 = \sum_i e_i \otimes g_i = \sum_{i,j} a_{ij} e_i \otimes f_j$$

but by Proposition 6.21 on page 155, the decomposable tensors $f_i \otimes f_j$ form a basis for $E \otimes F$, and it follows that $a_{ij} = 0$.

c) If $E \otimes F$ is a field the map ϕ is an isomorphism since algebra homomorphisms from fields automatically are injective. The implication the other way is trivial.

Solutions for exercises in Chapter 14

Exercise 14.1

a) Consider the short exact sequence

$$0 \longrightarrow (\mathfrak{a} : x) \longrightarrow A \longrightarrow A/\mathfrak{a} \longrightarrow 0 \tag{19.2}$$

where the map $A \to A/\mathfrak{a}$ sends an element a to the class $[xa]$ mod $\mathfrak{a}$. Flatness of B over A yields the equality $(\mathfrak{a} : x)\otimes_A B = (\mathfrak{a} : x)B$ so that tensorizing sequence (19.2) by B results in the exact sequence

$$0 \longrightarrow (\mathfrak{a} : x)B \longrightarrow B \longrightarrow B/\mathfrak{a}B \longrightarrow 0,$$

and hence $(\mathfrak{a} : x)B = (\mathfrak{a}B : x)$.

b) Assume that $x \in \mathfrak{b}$, but $x \notin \mathfrak{a}$. Then $(\mathfrak{a} : x)$ is a proper ideal in A, and there is a maximal ideal $\mathfrak{m}$ containing it. By assumption there is an index i so that $\mathfrak{m}A_i$ is proper. Now, we assumed that $\mathfrak{b}A_i \subseteq \mathfrak{a}A_i$, and therefore $(\mathfrak{a}A_i : x) = A_i$, but by in vie of a and the flatness of A_i, it holds that $(\mathfrak{a}A_i : x) = (\mathfrak{a} : x)A_i \subseteq \mathfrak{m}A$, contradiction.

c) For each of the finite number of indices such that $\mathfrak{b}A_i$ differ from $\mathfrak{a}A_i$ chose a finite set of generators for $\mathfrak{a}A_i$ and one may chose them to lie in $\mathfrak{a}$. Let $\mathfrak{c}$ be the ideal in A generated all these together with a finite generator set for $\mathfrak{b}$. Then $\mathfrak{a}A_i \subseteq \mathfrak{c}A_i$ for all i, and by b, it follows that $\mathfrak{a} \subseteq \mathfrak{c}$. By construction $\mathfrak{c} \subseteq \mathfrak{a}$, so $\mathfrak{a} = \mathfrak{c}$, and $\mathfrak{a}$ is finitely generated.

d) Let $a \in \mathfrak{a}$ be a non-zero element. Then $(a) \subseteq \mathfrak{a}$ with equality at least for all but the finitely many indices i so that $(a)A_i$ is proper. Then c ensures that $\mathfrak{a}$ is finitely generated.

EXERCISE 14.3 Assume that there is a relation

$$a_n(x)f(x)^n + a_{n-1}(x)f(x)^{n-1} + \ldots + a_0(x) = 0$$

with the a_i's from the field $k(x)$; we may assume that n is minimal and that the $a_n(x)$ and $a_{n-1}(x)$ are polynomials with no common factor. Substituting x^d for x and using that $f(x^d) = f(x) - x$ we find

$$a_n(x^d)f(x)^n + (-na_n(x^d)x + a_{n-1}(x^d))f(x)^{n-1} + \ldots + a_0x^d = 0$$

Eliminating the dominating term of the two relations, we deduce the equality

$$a_n(x^d)a_{n-1}(x) = -na_n(x^d)a_n(x)x + a_{n-1}(x^d)a_n(x),$$

which, since $a_n(x)$ and $a_{n-1}(x)$ are relatively prime, implies that $a_n(x^d)$ divides $a_n(x)$; but is absurd unless a_n is constant, say c, which must be non-zero. But then $a_{n-1}(x) = -ncx + a_{n-1}(x^d)$ which is impossible since k is of characteristic zero.

Solutions for exercises in Chapter 15

EXERCISE 15.1 The function v is well defines since if $x \in K$ has the two representations $x = ab^{-1} = cd^{-1}$, it holds that $ad = cb$; thus $v_0(a) + v_0(d) = v_0(c) + v_0(b)$, or equivalently $v_0(a) - v_0(b) = v_0(c) - v_0(d)$. The first axiom follows readily from the definition of n and that v_0 complies to the first axiom.

Assume then we are given two elements $x = ab^{-1}$ and $y = cd^{-1}$ from K with say $v(y) \geqslant v(x)$; that is, $v_0(c) - v_0(d) \geqslant v_0(a) - v_0(b)$; or in other words, $v_0(cb) = v_0(c) + v_0(b) \geqslant v_0(a) + v_0(d) = v_0(ad)$. We have $x + y = (ad + bc)/bd$ and find

$$v(x+y) = v_0(ad+bc) - v_0(bd) \geqslant v_0(ad) - v_0(bd) = v_0(a) - v_0(b) = v(x).$$

Exercise 15.2 Since A is a UFD, each element $x \in A$ can be factorized as $x = p^s z$ with z relatively prime to p in an unambiguous manner, and we may define $v_p(x) = s$. Clearly $v_p(xy) = v_p(x) + v_p(y)$, and if $x = p^s z$ and $y = p^t w$ with say $s \geqslant t$, it holds that $x + y = p^t(p^{s-t}z + w)$ so that $v_p(x+y) \geqslant t$ and the second requirement is also fulfilled. To see that v_p is unique, it suffices to see that if v is another valuation of our kind, then $v(x) = 0$ whenever x is relatively prime to p; so assume that $x \in A$ is such that $(x, p) = 1$. Thence there is a relation $1 = ax + bp$; from *iii)* of Lemma 15.2 we have $v(x) = v(a^{-1}(1 - bp)) = -v(a) \leqslant 0$, and since $v(x) \geqslant 0$, it follows that $v(x) = 0$.

Exercise 15.4 (DVR's *are maximal subrings*) This is just the fact that an over ring B of A is of the form $A_{\mathfrak{p}}$ where $\mathfrak{p}$ is a non-zero prime ideal, but when A is DVR, the maximal $\mathfrak{m}$ ideal is the only no-zero prime ideal, $A_{\mathfrak{m}} = A$.

Exercise 15.5 Clearly x belongs to the intersection $\bigcap_i A_i$ with each A_i being a valuation ring, it hods that x lies in each A_i and hence x^{-1} belongs to each A_i as well. If x lies in the union $\bigcup_i A_i$, it lies in one of the A_i's and hence x^{-1} lies in then same A_i (the only use of the hypothesis that the union be directed is to ensure that the union is a ring).

Exercise 15.6 (*Chevalley's lemma*) Assume that neither $\mathfrak{a}A[x]$ nor $\mathfrak{a}A[x^{-1}]$ is proper. Then there are relations of minimal degrees

$$\begin{aligned} 1 &= a_0 + a_1 x + \ldots + a_n x^n \\ 1 &= b_0 + b_1 x^{-1} + \ldots + b_m x^{-m} \end{aligned}$$

where a_i's and b_i's belong to A. The situation is symmetric in x and x^{-1} so we may certainly assume that $n \geqslant m$, which enables us to eliminate the x^n-term: just multiply the last equality by $a_n b_m^{-1} x^n$ to obtain an expression of $a_n x^n$ in terms of lower powers of x, which can be substituted into the first equality and gives a relation of degree less than n. (Note that if $n > m$, allready multiplying the last equality by x^m, gives a contradiction).

Exercise 15.7 (*Existence of places*) Say that a local subring B of K *dominates* another local subring B' if $B' \subseteq B$ and $\mathfrak{m}_B \cap B' = \mathfrak{m}_{B'}$. This a partial order on the set of local subrings of K. If $\{B_i\}$ is an ascending chain in this order, the union $\bigcup_i B_i$ will be a local subsring of K. The union is directed and so is a ring, and we assert that $\mathfrak{m} = \bigcup_i \mathfrak{m}_{B_i}$ is maximal ideal. Indeed, since the maximal ideals $\mathfrak{m}_{B_i}$ form a chain $\mathfrak{m}$ is an ideal (if $j \leqslant i$,

it holds that $\mathfrak{m}_{B_j} \subseteq \mathfrak{m}_{B_i} \cap B_j \subseteq \mathfrak{m}_{B_i}$); and it is clearly maximal since if $x \notin \mathfrak{m}$, it holds that x is invertible in each B_i to which it belongs.

So consider the set Σ of local subrings B such that $\mathfrak{m}_B \cap A = \mathfrak{p}$. It is not empty having the localization $A_\mathfrak{p}$ as a member, and by what we just did, any chain in Σ has an upper bund in Σ (namely its union). Thus, by Zorn's lemma there is a maximal element B in Σ. We contend that B is a valuation ring. Indeed, pick a non-zero element $x \in K$. Citing Chevalley's lemma from the previous exercise we know that either $\mathfrak{m}_B B[x]$ or $\mathfrak{m}_B B[x^{-1}]$ is a proper ideal; by symmetry we may assume that the former is. Choose a maximal ideal $\mathfrak{n}$ in $B[x]$ containing $\mathfrak{m}_B B[x]$. Then $B[x]_\mathfrak{n}$ is a local ring whose maximal ideal $\mathfrak{n}B[x]_\mathfrak{n}$ intersects B in a proper ideal containing $\mathfrak{m}$, hence the intersection equals $\mathfrak{m}$ since $\mathfrak{m}$ is maximal. By maximality of B, it follows that $B = B[x]_\mathfrak{n}$ and hence $x \in B$.

EXERCISE 15.12 We proceed by induction on ν; the case $\nu = 1$ is clear since $\mathfrak{p} = \mathfrak{p}A_\mathfrak{p} \cap A$ for any prime ideal. It will suffice to establish the inclusion $\mathfrak{p}^{(\nu)} \subseteq \mathfrak{p}^\nu$, the other one is trivial. So let $a \in \mathfrak{p}^\nu A_\mathfrak{p} \cap A = (x^\nu)A_\mathfrak{p} \cap A$; this means that $a = s^{-1}bx^\nu$ with $s \notin \mathfrak{p}$ and $b \in A$, or equivalently $tsa = tbx^\nu$ for some $t \notin \mathfrak{p}$. It holds that that $a \in \mathfrak{p}$, and we may write $a = cx$ with $c \in A$; hence $tscx = tbx^\nu$, and cancelling x (which is a non-zero divisor) we find $tsc = tbx^{\nu-1}$. It follows that $c \in \mathfrak{p}^{(\nu-1)}$, which by induction equals $\mathfrak{p}^{\nu-1}$. Thus $c = dx^{\nu-1}$ for some $d \in A$, and consequently $a = dx^\nu$.

EXERCISE 15.17

a) One easily sees that $\mathfrak{m} = (2, \sqrt{d}-1)$ so that the quotient map $\mathbb{Z}[\sqrt{d}] \to \mathbb{Z}[\sqrt{d}]/\mathfrak{m}$ factors by $\mathbb{F}_2[\sqrt{d}]/(\sqrt{d}-1) = \mathbb{F}_2[t]/(t^2-d, t-1)$, but $d \equiv 1 \mod 2$, so that $(t^2-d, t-1) = (t^2-1, t-1) = (t-1)$ and hence $\mathbb{F}_2[\sqrt{d}]/(\sqrt{d}-1) = \mathbb{F}_2$.

b) We relay on the isomorphism $\mathfrak{m}/\mathfrak{m}^2 \simeq \mathfrak{m}A_\mathfrak{m}/\mathfrak{m}^2A_\mathfrak{m}$ from xxxx, and the fact that $\dim_{A/\mathfrak{m}} \mathfrak{m}A_\mathfrak{m}/\mathfrak{m}^2A_\mathfrak{m}$ equals the minimal number of generators for $\mathfrak{m}A_\mathfrak{m}$. So let us compute $\mathfrak{m}^2$; and using that $d \equiv 1 \mod 4$ we find that $(2, \sqrt{d}-1)^2 = (4, 2\sqrt{d}-2, d+1-\sqrt{d}) = (4, 2(\sqrt{d}-1))$; in other words, $\mathfrak{m}/\mathfrak{m}^2 \simeq 2\mathbb{F}_2$ as a vectorspace.

c) If Clearly $\mathfrak{m} \cap \mathbb{Z} = (2)$, and by Lying–Over any other maximal ideal $\mathfrak{n}$ intersects $\mathbb{Z}$ in an ideal different from (2), Hence $2 \notin \mathfrak{n}$ and by consquent $\mathbb{Z}[\sqrt{d}]_\mathfrak{n} = \mathbb{Z}[(1+\sqrt{d})/2]_\mathfrak{n}$ and latter is a DVR since $\mathbb{Z}[(1+\sqrt{d})/2]$ is normal.

EXERCISE 15.29 We contend that $\mathfrak{m}$ is generated by any element x so that $v(x) = (0, \dots, 0, 1)$. The maximal ideal $\mathfrak{m}$ corresponds to the minimal isolated subgroup (0), hence $\mathfrak{m}$ equals the set of elements y so that the first non-zero coordinate is positive. It follows that $v(xy^{-1}) > 0$.

Solutions for exercises in Chapter 16

EXERCISE 16.3 Let x and y be homogenous elements with $\deg x = a > 0$ and $\deg y = -b < 0$. Then $\deg x^b y^a = 0$; so $x^b y^a \in R_0 = k$ and is thus invertible. It follows that

x is invertible. That every non-zero homogenous element is invertible has the effect that $I = \{ \deg x \mid x \in R \text{ homogenous } \}$ is an *ideal*: the only pertinent point is that I is closed under sign change as $-\deg x = \deg x^{-1}$. Let now r be the positive generator of the I. Then $r = \deg w$ for some w. If $x \in R$ is any homogeous element it holds that $\deg x = a \deg w$ for some a: hence $x^{-1}w^a \in k$ and $x = \alpha w^a$ for some scalar α. That is $R = k[w, w^{-1}]$.

Exercise 16.4

a) It suffices to see thatfor all maximal ideals $\mathfrak{m}$ the localized complex $(C_\bullet)_\mathfrak{m}$ is exact. So we may assume that R is local with; let the maximal ideal be $\mathfrak{m}$. If all the x_{ij}'s belonged to $\mathfrak{m}$, the three minors would belong to it too, but they do not since $\mathfrak{a} = R$. After as series of elementary row and column operations and possible a renaming, we may assume that x_{00} is invertible, and can then resort to the argument in the Example.

b) Localize in x. Then $\mathfrak{a}_x = R_x$ and the localized Hilbert-Burch $(C_\bullet)_x$ complex is exact. So $\ker M$ is killed by a power of x, and as it is contained in a free module, it follows that $\ker M = 0$ since x is a non-zero divisor.

c) Let x, y be the regular sequence in $\mathfrak{a}$. Since x is non-zero divisor, $H^2 C_\bullet = 0$. Now y is non-zero divisor in $R/(x)R$ lying in $\mathfrak{a}/(x)R$. Hence $H^2(C_\bullet/xC_\bullet) = 0$. The long exact sequence derived from the short exact sequence of complexes

$$0 \longrightarrow C_\bullet \longrightarrow C_\bullet \longrightarrow C_\bullet/xC_\bullet \longrightarrow 0,$$

has the relevant part

$$0 = H^2(C_\bullet/xC_\bullet) \longrightarrow H^1(C_\bullet) \xrightarrow{x} H^1(C_\bullet).$$

Hence multiplication by x is injective, and since a power of x kills $H^1(C_\bullet)$, it ensues that $H^1(C_\bullet) = 0$.

Exercise 16.5 Any pair of minors does not have common factors (all three are irreducible). Hence any two form a regular sequence in $\mathfrak{a}$, and the Hilbert–Burch complex is a resolution of A. It follows that the Hilbert polynomial is given as

$$\chi_A(t) = \binom{t+3}{3} - 3\binom{t+1}{3} + 2\binom{t}{3} = 3t + 1.$$

Solutions for exercises in Chapter 17

Solutions for exercises in Chapter 18

Lecture 20

Index