ALGEBRA II

P. Stevenhagen

2026

TABLE OF CONTENTS ALGEBRA II

This version is based on the English translation, by Reinie Erné, of the Dutch original. The latest version will be made available on https://websites.math.leidenuniv.nl/algebra/.
Publication date of this version: September 2026.

Each subsequent edition will, hopefully, contain fewer typos and inaccuracies than the current one—please send any comments to the author at psh@math.leidenuniv.nl.

Author's address:

Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden

11 Rings

In Algebra I, we studied sets G endowed with a single binary operation $G \times G \to G$ inducing a *group structure* on G. In these course notes,[1] we study sets R endowed with two binary operations called addition and multiplication that induce a *ring structure* on R. The sets of integers $\mathbf{Z}$ and real numbers $\mathbf{R}$ are well-known examples, and so is the ring $\mathbf{Z}/n\mathbf{Z}$ that we encountered already in 6.9. We repeat the formal definition that was given in 6.8.

11.1. Definition. *A ring is an abelian group R written additively, endowed with a binary operation $R \times R \to R$ written multiplicatively that satisfies the following three conditions:*

(R1) *The set R contains a unit element 1 for the multiplication.*

(R2) *For any three elements $x, y, z \in R$, we have the associative property*

$$x(yz) = (xy)z.$$

(R3) *For any three elements $x, y, z \in R$, we have the distributive properties*

$$x(y+z) = xy + xz \qquad \text{and} \qquad (x+y)z = xz + yz.$$

If we, moreover, have $xy = yx$ for all $x, y \in R$, then R is called a commutative ring.

The underlying additive group of a ring R is abelian by definition. This is not a restriction because the commutativity of the addition is a *consequence* of the other axioms. After all, by the distributive properties, we have

$$(x+y)(1+1) = x(1+1) + y(1+1) = x + x + y + y,$$
$$(x+y)(1+1) = (x+y)\cdot 1 + (x+y)\cdot 1 = x + y + x + y;$$

the identity $x + y = y + x$ follows immediately.

The multiplicative structure of a ring is rarely that of a group. Multiplication is associative, and the unit element $1 \in R$ is unique, but the standard example $R = \mathbf{Z}$ shows that ring elements do not always have a multiplicative inverse and that left or right multiplication by a ring element does not generally give a bijection from the ring to itself. The identity $0 \cdot x = (0+0) \cdot x = 0 \cdot x + 0 \cdot x$ shows that the zero element has product 0 with every ring element.

The *zero ring* is the ring obtained by endowing the trivial group $R = \{0\}$ with the multiplication $0 \cdot 0 = 0$; notation: $R = 0$. In the zero ring, we have $0 = 1$. If R contains an element $x \neq 0$, then we have $1 \cdot x = x \neq 0 = 0 \cdot x$, and so $0 \neq 1$.

Exercise 1. Show that the zero ring is the only ring where multiplication is a group operation.

A *subring* R' of a ring R is a subset $R' \subset R$ with $1 \in R'$ on which we define the structure of a ring by restricting the ring operations on R. In other words, R' is an additive subgroup of R that contains $1 \in R$ and is closed under multiplication within R.

▶ Units

An element $x \in R$ is called a *unit* if there exists an element $y \in R$ with $xy = yx = 1$. If such an element y exists, it is uniquely determined by x (Exercise 11). We then write $y = x^{-1}$ and call y the (multiplicative) *inverse* of x. The set of units of R is the *group of units* R^* of R.

11.2. Lemma. *The group of units R^* of R is a group under multiplication.*

Proof. The product xy of two units is a unit: $y^{-1}x^{-1}$ is a two-sided inverse of xy. So multiplication defines a binary operation on R^*. The element $1 \in R^*$ is the unit element, the associativity is axiom (R2), and inverses of units—which are also units—exist by the definition of R^*. □

11.3. Definition. *A division ring or skew field is a ring R with group of units $R^* = R \setminus \{0\}$. A commutative division ring is called a field.*

Hamilton's *quaternion algebra* $\mathbf{H}$, defined in 8.7, is an example of a non-commutative division ring (Exercise 19). The better-known division rings $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ of, respectively, rational, real, and complex numbers are examples of fields.

The ring $\mathbf{Z}$ of integers is a commutative ring with group of units $\mathbf{Z}^* = \{\pm 1\}$. It is a subring of each division ring mentioned above.

For an integer $n \neq 0$, the quotient $\mathbf{Z}/n\mathbf{Z}$ is a *finite* commutative ring with group of units

$$(\mathbf{Z}/n\mathbf{Z})^* = \{\bar{a} \in \mathbf{Z}/n\mathbf{Z} : \mathrm{GCD}(a, n) = 1\}.$$

As in 6.12, we conclude that $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ is a field for every prime p. Note that for $n = 1$, $\mathbf{Z}/n\mathbf{Z}$ is the zero ring and that this is not a field.

▶ Examples of Rings

11.4. Rings of functions. Let X be an arbitrary set and R be a ring. Then the set $\mathrm{Map}(X, R)$ of R-valued functions on X has the structure of a ring if we define sums and products of functions pointwise by

$$(f + g)(x) = f(x) + g(x) \qquad \text{and} \qquad (f \cdot g)(x) = f(x) \cdot g(x).$$

For commutative R, the ring of functions $\mathrm{Map}(X, R)$ is commutative.

Rings of functions are very common. Usually, X has an additional structure, and instead of the ring of *all* R-valued functions, some subring of interesting functions is studied. Consider, for example, the ring $C(X) \subset \mathrm{Map}(X, \mathbf{R})$ of continuous real-valued functions on a topological space X or the ring $C^\infty(0, 1)$ of infinitely differentiable functions on the open unit interval $(0, 1)$.

Exercise 2. Verify that $C(X)$ and $C^\infty(0, 1)$ are sub*rings*.

11.5. Polynomial rings. For every ring R, we can define the polynomial ring $R[X]$ in one variable X over R in the usual way. The elements of $R[X]$ are formal expressions

$\sum_{k\geq 0} r_k X^k$ with coefficients $r_k \in R$ that are different from 0 for only finitely many integers $k \geq 0$. These expressions, called *polynomials*, are added coordinate-wise:

$$\textstyle\sum_{k\geq 0} r_k X^k + \sum_{k\geq 0} s_k X^k = \sum_{k\geq 0}(r_k + s_k)X^k.$$

In particular, every polynomial can be written uniquely as a sum of polynomials with exactly one coefficient different from 0, called *monomials*. We define multiplication for monomials by setting $r_k X^k \cdot s_\ell X^\ell = r_k s_\ell X^{k+\ell}$. It follows from the distributive property (R3) that the n-th coefficient c_n of the product polynomial

$$\textstyle\sum_{k\geq 0} c_k X^k = (\sum_{k\geq 0} r_k X^k) \cdot (\sum_{k\geq 0} s_k X^k)$$

is equal to $c_n = \sum_{k=0}^{n} r_k s_{n-k}$. A straightforward verification shows that this makes $R[X]$ into a ring. In most applications, R is commutative, in which case so is $R[X]$. The polynomial ring $R[X]$ contains R as the subring of *constant polynomials*.

For every element $r \in R$, a polynomial $f \in R[X]$ gives a value $f(r)$ resulting from the *evaluation* of f in $X = r$. Sometimes, for example for $R = \mathbf{R}$, we can use this to view $R[X]$ as a subring of the ring of functions $\mathrm{Map}(R, R)$. However, this is not possible in general because polynomials in $R[X]$ are not fixed by their values on R for all R. For example, the polynomial $X^p - X \in \mathbf{F}_p[X]$ is not the zero polynomial, but by Fermat's little theorem 6.18, the corresponding function $\mathbf{F}_p \to \mathbf{F}_p$ is the zero function.

There are several ways to vary the construction of the polynomial ring $R[X]$ from R. For example, if we admit expressions $\sum_{k\geq 0} r_k X^k$ where infinitely many coefficients may differ from 0, then the definitions above for the sum and product lead to the *ring of power series* $R[[X]]$ over R. The *Taylor series* from analysis are known examples in the case $R = \mathbf{R}$.

If, in the definition of a polynomial, we also allow negative powers of X, then we obtain the ring of *Laurent polynomials*

$$R[X, X^{-1}] = \{\textstyle\sum_{k\in\mathbf{Z}} r_k X^k : \quad r_k = 0 \text{ for almost all } k \in \mathbf{Z}\}.$$

The calculation rules are the same as for usual polynomials, but $r_k X^k \cdot s_\ell X^\ell = r_k s_\ell X^{k+\ell}$ now holds for all $k, \ell \in \mathbf{Z}$. Finally, we have the ring $R((X))$ of *Laurent series* over R; for $R = \mathbf{C}$, this ring is widely used in complex analysis. Here, we consider expressions $\sum_{k\in\mathbf{Z}} r_k X^k$ where only finitely many coefficients r_k with $k < 0$ differ from 0. Note that with this definition, all coefficients of the product of two Laurent series are still given by *finite* sums. The ring $R((X))$ contains both $R[X, X^{-1}]$ and $R[[X]]$ as subrings in a natural way.

In an abstract ring R, or even in $R = \mathbf{Z}$, "infinite sums" have no meaning. So the power series and Laurent series defined above are formal objects, which do not induce functions $R \to R$. To define known notions from analysis such as *limit* and *convergence*, we need an additional (topological) structure on R.

We obtain polynomials in multiple variables by iterating the construction of polynomials in one variable. The polynomial ring $(R[X])[Y]$ in the variables X and Y, which can also be viewed as the polynomial ring $(R[Y])[X]$ (why?), is denoted by

$R[X,Y]$. More generally, we have the polynomial ring $R[X_1, X_2, \ldots, X_n]$ in n variables, and, likewise, we can construct the ring of power series $R[[X_1, X_2, \ldots, X_n]]$ in n variables.

A definition of Laurent series in more than one variable is more problematic (Exercise 71).

11.6. Group rings. For a ring R and a group G, we define the *group ring* $R[G]$ in a way somewhat reminiscent of the construction of the polynomial ring. As a set, $R[G]$ consists of the expressions $\sum_{g\in G} r_g g$, where $r_g \in R$ is 0 for almost all $g \in G$. Addition is done component-wise:

$$(\textstyle\sum_{g\in G} a_g g) + (\sum_{g\in G} b_g g) = \sum_{g\in G}(a_g + b_g)g.$$

Since only finite sums can occur, multiplication is given by combining the rule $a_g g \cdot b_h h = a_g b_h gh$ with the distributive laws. It is easy to check that this makes $R[G]$ into a ring with unit element $1 \cdot e$ and that for $R \neq 0$, the group ring $R[G]$ is commutative if and only if R is commutative and G is abelian. The multiplicative notation for G used here avoids confusion between the addition in $R[G]$ and the group operation in G. If G is written additively, then there is the notation $\sum_{g\in G} a_g[g]$ for ring elements that clearly differentiates between $[g_1 + g_2]$ and $[g_1] + [g_2]$.

Exercise 3. Show that $R[X]$ can be viewed as a subring of the group ring $R[\mathbf{Z}]$ and that $R[X, X^{-1}]$ can be identified with $R[\mathbf{Z}]$.

11.7. Matrix rings. A common type of ring, which is usually not commutative, is the *matrix ring* of dimension $n \geq 0$ over an arbitrary ring R. For $n \in \mathbf{Z}_{\geq 0}$ and matrices $A = (a_{ij})_{i,j=1}^n$ and $B = (b_{ij})_{i,j=1}^n$ in the set $\mathrm{Mat}_n(R)$ of $n \times n$ matrices with coefficients in R, we define the sum "coefficient-wise" by setting

$$A + B = (a_{ij} + b_{ij})_{i,j=1}^n.$$

The *matrix product* $AB = (c_{ij})_{i,j=1}^n$ has coefficients $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$ given by what is also called the "row-column rule." A direct (but slightly tiring) verification shows that this defines the structure of a ring on $\mathrm{Mat}_n(R)$. We have $\mathrm{Mat}_0(R) = 0$, and $\mathrm{Mat}_1(R)$ can be identified with R. For $n \geq 2$ and $R \neq 0$, the ring $\mathrm{Mat}_n(R)$ is non-commutative. The group of units of $\mathrm{Mat}_n(R)$ is the group $\mathrm{GL}_n(R)$ of invertible $n \times n$ matrices.

In linear algebra, R is a field such as $\mathbf{R}$, $\mathbf{C}$, or $\mathbf{F}_p$. The elements of $\mathrm{Mat}_n(R)$ then correspond to the R-linear maps $R^n \to R^n$ from the n-dimensional "standard vector space" over R to itself. Under this identification, the definition of the matrix product is very natural: it is the *composition* of the corresponding maps.

11.8. Endomorphism rings. For every abelian group A, the set $\mathrm{End}(A)$ of group homomorphisms $A \to A$, usually called *endomorphisms*, is a ring in a natural way. The sum and the product of two endomorphisms $f, g \in \mathrm{End}(A)$ are defined as

$$\begin{aligned}(f + g)(a) &= f(a) + g(a),\\ (f \cdot g)(a) &= (f \circ g)(a) = f(g(a)).\end{aligned}$$

We already know (Exercise 4.41) that for abelian A, the map $f+g$ is indeed a homomorphism and that under this addition, $\mathrm{End}(A)$ is an abelian group with the trivial homomorphism $A \to 0 \subset A$ as zero element. The identity $1 = \mathrm{id}_A$ is a unit element for the multiplication, and a distributive law such as $f(g+h) = fg + fh$ follows easily by a simple expansion:

$$[f(g+h)](a) = f(g(a)+h(a)) = f(g(a)) + f(h(a)) = (fg)(a) + (fh)(a).$$

We conclude that $\mathrm{End}(A)$ is a ring. It is generally not commutative. The group of units $\mathrm{End}(A)^*$ of $\mathrm{End}(A)$ is nothing but the automorphism group $\mathrm{Aut}(A)$ of A already defined in §4.

Exercise 4. Show that $\mathrm{End}(\mathbf{Z}^n)$ can be identified with the matrix ring $\mathrm{Mat}_n(\mathbf{Z})$.

Endomorphism rings are very general because, as we will see in 11.12, the elements of a ring R can be viewed as endomorphisms of the underlying additive group.

▶ Zero Divisors

In fields or in rings such as $\mathbf{Z}$, we know that the product of two non-zero elements is never 0. In arbitrary rings, however, this phenomenon can occur. For matrices, this is a fact known from linear algebra, but there are many more examples. For instance, if we take two non-zero functions $f, g : [0,1] \to \mathbf{R}$ that differ from 0 on disjoint pieces of $[0,1]$, then fg is the zero function even though f and g are not. If g is an element of order n in a group G, then in the group ring $\mathbf{Z}[G]$, we have the identity $(1-g)(1+g+g^2+\ldots+g^{n-2}+g^{n-1}) = 1-g^n = 0$.

11.9. Definition. *An element $x \in R$ is called a zero divisor if there exists an element $y \neq 0$ in R such that we have $xy = 0$ or $yx = 0$.*

In every ring $R \neq 0$, we have the *trivial zero divisor* $0 \in R$. For non-commutative rings, we distinguish in 11.9 between *left zero divisors* (the case $xy = 0$) and *right zero divisors* (the case $yx = 0$). An element $x \in R$ is a left zero divisor if and only if the map $\lambda_x : R \to R$ given by $\lambda_x(r) = xr$, which is an endomorphism of the additive group of R, is *not* injective. For right zero divisors, we have the same statement with λ_x replaced by the map $\rho_x : R \to R$ given by $\rho_x(r) = rx$. Left zero divisors are not always right zero divisors (Exercise 28).

Exercise 5. Show that a unit of a ring is never a zero divisor.

The behavior of rings with non-trivial zero divisors differs in many aspects from that of well-known rings such as $\mathbf{Z}$ and $\mathbf{R}$. For example, for $R = \mathbf{Z}/8\mathbf{Z}$, the linear polynomial $\overline{4}X$ and the quadratic polynomial $X^2 - \overline{1}$ each have *four* distinct zeros in R. Also, from the equality $\overline{4} \cdot \overline{1} = \overline{4} \cdot \overline{3} \in \mathbf{Z}/8\mathbf{Z}$, we *cannot* conclude that $\overline{1} = \overline{3}$, as we could in the case of multiplicative groups.

▶ Integral Domains

A commutative ring $R \neq 0$ without non-trivial zero divisors is called an *integral domain*. In integral domains, we have the following property, known from the real numbers:

$$xy = 0 \quad \Longrightarrow \quad x = 0 \quad \text{or} \quad y = 0.$$

By replacing y with $y - z$, we obtain the implication

$$xy = xz \quad \Longrightarrow \quad x = 0 \quad \text{or} \quad y = z,$$

which suggests that in integral domains, we can "divide out" elements different from 0. This is true in fields and, therefore, in all subrings of fields, which are integral domains. In fact, *every* integral domain R can be viewed as a subring of a field, namely of the field of fractions $Q(R)$ of R. The construction of $Q(R)$ from R is the ring-theoretic analog of the construction of $\mathbf{Q}$ from $\mathbf{Z}$. The definition is somewhat subtle because the representation of fractions is not unique—as is the case for ordinary fractions.

11.10. Field of fractions. For an integral domain R, define an equivalence relation on the product set $R \times (R \setminus \{0\})$ by setting

$$(a, b) \sim (c, d) \Longleftrightarrow ad = bc.$$

The transitivity of this relation needs to be verified: if $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, we have $adf = bcf = bde$, and by the commutativity of R, this gives $d(af - be) = 0$. Since R has no non-trivial zero divisors and $d \neq 0$ holds, the relation $af = be$ follows, which gives $(a, b) \sim (e, f)$.

We denote the equivalence classes of (a, b) evocatively by $\frac{a}{b}$. The set $Q(R)$ of equivalence classes now becomes a field if we define the addition and multiplication of "fractions" in the familiar way by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad \text{and} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

We can verify that these are *well-defined* operations: they do not depend on the choice of the representatives. Once we know this, it follows easily that the ring axioms are satisfied. Since every element $\frac{a}{b} \in Q(R)$ different from the zero element $\frac{0}{b} = \frac{0}{1}$ has an inverse $\frac{b}{a} \in Q(R)$, the field of fractions $Q(R)$ is a field. When we identify an element $r \in R$ with the class $\frac{r}{1} \in Q(R)$, R becomes a subring of $Q(R)$.

If, above, we take R to be the polynomial ring $K[X]$ over a field K, then $Q(R)$ is the field $K(X)$ of *rational functions* with coefficients in K.

The formation of the field of fractions of an integral domain is a special case of the central concept of *localization* in commutative algebra. We can allow other subsets of R than only $S = R \setminus \{0\}$ as "denominators" in the *localized ring* $S^{-1}R$. With an appropriate adaptation of the equivalence relation on the set of "fractions" $\frac{r}{s} = (r, s) \in R \times S$, the construction works for arbitrary commutative rings R and not only for integral domains. See Exercise 72.

▶ Homomorphisms and Ideals

As in group theory, maps in ring theory are called *homomorphisms*, and there is an analogous notion of an isomorphism.

11.11. Definition. *A map $f : R \to R'$ between rings is called a ring homomorphism if for all $x, y \in R$,*

$$f(x+y) = f(x) + f(y) \qquad \text{and} \qquad f(xy) = f(x)f(y)$$

hold and also $f(1_R) = 1_{R'}$. A bijective ring homomorphism is called a ring isomorphism.

As in group theory, rings R and R' are called isomorphic if there exists a ring isomorphism $R \xrightarrow{\sim} R'$; notation: $R \cong R'$.

The identity $f(1_R) = 1_{R'}$ in (11.11) means that f sends the unit element for the multiplication in R to the unit element for the multiplication in R'. If R' is not the zero ring, this condition excludes the "zero map" $f : R \to 0 \subset R'$, which does satisfy the first two conditions, from being a homomorphism. Under an injective homomorphism $f : R \to R'$, we can view R as a subring of R'. More generally, under a ring homomorphism $f : R \to R'$, the image $f[R]$ is a subring of R'.

Cayley's theorem stated in 5.8 says that elements of a group G can be seen as permutations of the set G and that this makes G into a subgroup of $S(G)$. The ring-theoretic analog is obtained by viewing ring elements as endomorphisms of the underlying additive group.

11.12. Theorem. *Let R be a ring and R^+ be the underlying abelian additive group. For $x \in R$, denote by $\lambda_x : R^+ \to R^+$ the left multiplication $r \mapsto xr$. Then*

$$\begin{aligned} f : R &\longrightarrow \mathrm{End}(R^+) \\ x &\longmapsto \lambda_x \end{aligned}$$

is an injective homomorphism, and R is isomorphic to a subring of $\mathrm{End}(R^+)$.

Proof. By the distributive law $x(r_1 + r_2) = xr_1 + xr_2$, λ_r is an endomorphism of the additive group R^+. The distributivity also gives

$$(\lambda_x + \lambda_y)(r) = xr + yr = (x+y)r = \lambda_{x+y}(r),$$

and the associativity gives $(\lambda_x \cdot \lambda_y)(r) = x(yr) = (xy)r = \lambda_{xy}(r)$. Since $\lambda_1 = \mathrm{id}_{R^+}$, the map f is now a homomorphism, and since $\lambda_x(1) = x$, the maps λ_x and λ_y are different for $x \neq y$. We conclude that R is isomorphic to the subring $f[R] \subset \mathrm{End}(R^+)$. □

Exercise 6. Prove: R is isomorphic to $\mathrm{End}_R(R^+) = \{f \in \mathrm{End}(R^+) : f(r_1 r_2) = f(r_1)r_2 \text{ for all } r_1, r_2 \in R\}$.

In group theory, we saw that the subgroups $H \subset G$ that can serve as kernels of homomorphisms form a well-defined class, that of the *normal* subgroups. For a ring R, the underlying additive group R^+ is abelian, so all subgroups of R^+ are normal. A

ring homomorphism $f : R \to R'$ is, in particular, a homomorphism on the underlying additive group, so the kernel

$$\ker(f) = \{r \in R : f(r) = 0\}$$

is a subgroup of R^+. However, the multiplicative property $f(xy) = f(x)f(y)$ implies that not all subgroups of R^+ can serve as kernels of ring homomorphisms; only the so-called *ideals* of R can.

11.13. Definition. *An ideal I of a ring R is a subgroup of the additive group R^+ with the following property:*

$$\text{for } r \in R \text{ and } x \in I, \text{ we have } rx \in I \text{ and } xr \in I.$$

It is not difficult to see that the kernel of a ring homomorphism $f : R \to R'$ is an ideal. After all, for $r \in R$ and $x \in \ker(f)$, we have $f(rx) = f(r)f(x) = f(r) \cdot 0 = 0$ and therefore $rx \in \ker(f)$. Likewise, we have $xr \in \ker(f)$. Conversely, the following analog of 4.14 for rings shows that the ideals $I \subset R$ are exactly the subgroups of R^+ for which the quotient group R/I, which, as we know, has the cosets $x + I$ of I in R as elements, inherits the structure of a ring from R.

11.14. Theorem. *Let R be a ring and $I \subset R$ be an ideal. Then the multiplication*

$$(x + I)(y + I) = xy + I$$

defines the structure of a ring on the quotient R/I. This makes the canonical map $R \to R/I$ into a ring homomorphism with kernel I.

Proof. It suffices to show that the mentioned multiplication on R/I is well defined. This means that for elements $x \equiv x' \bmod I$ and $y \equiv y' \bmod I$, the products xy and $x'y'$ must be congruent modulo I. If, as in (6.10), we write

$$xy - x'y' = x(y - y') + (x - x')y',$$

then it follows from the hypotheses $y - y' \in I$ and $x - x' \in I$ and the ideal property 11.13 that $xy - x'y'$ indeed lies in I.

The ring axioms for R/I now follow immediately from those for R, and the canonical map $R \to R/I$ is almost by definition a ring homomorphism with kernel I. □

Every ring $R \neq 0$ has two trivial ideals, the zero ideal $\{0\}$ and the ring R itself. The corresponding quotient rings are R and the zero ring. If R is a field, or more generally a division ring, then there are no other ideals. After all, every ideal $I \subset R$ different from $\{0\}$ then contains a unit $x \in R^*$ and therefore also $1 = xx^{-1}$. By the ideal property 11.13, we then have $r \cdot 1 = r \in I$ for all $r \in R$, and so $I = R$.

For a ring R without non-trivial ideals, every ring homomorphism $f : R \to R'$ to a ring $R' \neq 0$ is injective because $\ker(f)$ is the zero ideal and 4.4 remains valid in the context of rings.

An example of a non-trivial ideal is $n\mathbf{Z} \subset \mathbf{Z}$ for an integer $n > 1$. The corresponding quotient ring is the ring $\mathbf{Z}/n\mathbf{Z}$ of residue classes modulo n from 6.9. More generally, for a commutative ring R and $x \in R$, we can look at the *principal ideal*

$$(x) = xR = \{xr : r \in R\}$$

generated by x. This is the smallest ideal that contains x (why?), and it is equal to R if and only if x is a unit. The elements of (x) are the elements of R *divisible* by x; they are the *multiples* of x. By 6.2, for $R = \mathbf{Z}$, *all* ideals are principal. An integral domain with this pleasant property is called a *principal ideal domain*, sometimes abbreviated to PID. We will study them in more detail in the next section.

Exercise 7. Show that every commutative ring $R \neq 0$ without non-trivial ideals is a field.

For every finite subset $S = \{s_1, s_2, \dots s_n\}$ of a commutative ring R, we can define the ideal $(S) = (s_1, s_2, \dots s_n) \subset R$ *generated by* S as

$$(S) = \{\textstyle\sum_{s\in S} r_s s : \quad r_s \in R \text{ for all } s\}.$$

This is a subgroup of R^+ that satisfies the ideal property 11.13 and clearly forms the smallest ideal that contains S. Ideals generated by a finite set are called *finitely generated*. For infinite S, we define (S) as above, but with the additional condition that in the sums $\sum_{s\in S} r_s s$, almost all r_s are 0.

In non-commutative rings, it makes sense to speak not only of ideals but also of left and right ideals. A *left ideal* of R is an additive subgroup of R^+ mapped to itself under left multiplication by R. Likewise, we have the notion of a *right ideal*. If R is non-commutative, then an ideal of R in the sense of 11.13, which is both a left ideal and a right ideal, is called a *two-sided ideal*. We can only construct quotient rings R/I for two-sided ideals.

Exercise 8. Show that every element $x \in R$ generates a left ideal Rx and a right ideal xR. Determine Rx and xR for $R = \mathrm{Mat}_2(\mathbf{R})$ and $x = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

▶ Isomorphism and Homomorphism Theorems

The fundamental isomorphism theorem 4.10 from group theory holds unchanged in the ring-theoretic context.

11.15. Isomorphism theorem. *Let $f : R \to R'$ be a ring homomorphism with kernel I. Then the map*

$$\overline{f} : R/I \xrightarrow{\sim} f[R]$$

given by $x + I \mapsto f(x)$ is a ring isomorphism.

Proof. By 4.10, $\overline{f}$ is an isomorphism of additive groups. Since f is a ring homomorphism, $\overline{f}$ is also a ring homomorphism and so a ring isomorphism. □

11.16. Example. Let R be a commutative ring and $a \in R$ be an element. Then the *evaluation map* $\phi_a : R[X] \to R$ on the polynomial ring $R[X]$ in the "point" $a \in R$ is

defined by $\phi_a(f) = f(a)$. It is easy to check (Exercise 35) that for commutative R, this is a ring homomorphism. Let us show that the kernel $\ker \phi_a$, which consists of the polynomials in $R[X]$ having a as a *zero*, is equal to the principal ideal $(X - a)$.

Since $X - a \in \ker \phi_a$, we have $(X - a) \subset \ker \phi_a$. Conversely, every polynomial $f = \sum c_i X^i$ with $f(a) = 0$ can be written as

$$f = f - f(a) = \textstyle\sum_i c_i X^i - \sum_i c_i a^i = \sum_i c_i (X^i - a^i).$$

Each term $X^i - a^i$ is a multiple of $X - a$:

$$(*) \qquad X^i - a^i = (X^{i-1} + aX^{i-2} + a^2 X^{i-3} + \ldots + a^{i-2}X + a^{i-1})(X - a).$$

It follows that f itself also lies in $(X - a)$, and we obtain $\ker \phi_a = (X - a)$. If we now apply the isomorphism theorem for ϕ_a, then since ϕ_a is surjective, we obtain an isomorphism

$$\begin{aligned} R[X]/(X - a) &\xrightarrow{\sim} R \\ f + (X - a) \cdot R[X] &\longmapsto f(a). \end{aligned}$$

Apparently, every polynomial $f \in R[X]$ can be written as $f = q \cdot (X - a) + f(a)$ with $q \in R[X]$. Since the commutativity of R is not needed for the equality $(*)$, the above is also true in non-commutative rings. It is a special case of *division with remainder* in polynomial rings, which we will discuss further in 12.1.

Exercise 9. Show that the map $\mathbf{Z}[X] \to \mathbf{Z}/3\mathbf{Z}$ given by $f \mapsto (f(0) \bmod 3)$ is a ring homomorphism with kernel $(3, X)$, and prove that $(3, X)$ is not a principal ideal.

Many of the theorems proved in §8 generalize effortlessly to rings. In most cases, it suffices to note that the constructed maps are not only group homomorphisms but also ring homomorphisms. Let us give, for example, the ring-theoretic analog of the homomorphism theorem 8.4. For the analogs of 8.1, 8.2, and 8.5, we refer to Exercises 51–53.

11.17. Homomorphism theorem. *Let $f : R \to R'$ be a ring homomorphism and $I \subset \ker(f)$ be an ideal of R. Then there exists a unique ring homomorphism $\overline{f} : R/I \to R'$ such that f is the composition*

$$R \xrightarrow{\pi} R/I \xrightarrow{\overline{f}} R'$$

of the quotient map $\pi : R \to R/I$ and $\overline{f}$. □

▶ Chinese Remainder Theorem

In a commutative ring R, as in $\mathbf{Z}$, we can look at sums, products, and intersections of ideals. The *intersection* $I \cap J$ of ideals I and J of R is also an ideal, and the same holds for the *sum* $I + J = \{i + j : i \in I \text{ and } j \in J\}$. If we have $I + J = R$, then, analogously to the case $R = \mathbf{Z}$ in 6.3.2, we call the ideals I and J *relatively prime*. The *product* $I \cdot J$ (or IJ) of two ideals is defined as the ideal *generated* by the elements $i \cdot j$ with $i \in I$ and $j \in J$. The set $\{i \cdot j : i \in I \text{ and } j \in J\}$ is not generally an ideal

of R (Exercise 44). Since all products $i \cdot j$ are contained in $I \cap J$, we always have the inclusion

$$I \cdot J \subset I \cap J. \tag{11.18}$$

For $R = \mathbf{Z}$, this means that for $a, b \in \mathbf{Z}$, the product ab is divisible by the least common multiple $\mathrm{LCM}(a, b)$.

The Chinese remainder theorem 6.15 can be generalized to arbitrary commutative rings. First note that the Cartesian product $R_1 \times R_2$ of two rings with coefficient-wise binary operations is a ring. The projection $R_1 \times R_2 \to R_1$ onto the first coordinate is a surjective ring homomorphism. However, for $R_2 \neq 0$, the map $R_1 \to R_1 \times R_2$ given by $r_1 \mapsto (r_1, 0)$ is not a ring homomorphism in the sense of 11.11: the unit element $1 \in R_1$ is not sent to the unit element $(1, 1) \in R_1 \times R_2$. This is sometimes called a *non-unitary ring homomorphism*.

11.19. Chinese remainder theorem. *Let I and J be relatively prime ideals of a commutative ring R. Then we have $I \cdot J = I \cap J$, and the natural map*

$$\begin{array}{rccc} \psi: & R/(I \cdot J) & \xrightarrow{\sim} & R/I \times R/J \\ & (x + I \cdot J) & \longmapsto & (x + I, x + J) \end{array}$$

is a ring isomorphism.

Proof. Because of (11.18), it suffices to prove the inclusion $I \cap J \subset I \cdot J$ to get the equality $I \cdot J = I \cap J$. By the assumption, there exist elements $i \in I$ and $j \in J$ with $i + j = 1$. If we now write $x = x(i + j) = ix + xj$ for $x \in I \cap J$, then ix and xj each lie in $I \cdot J$, and therefore so does x.

The natural map $f : R \to R/I \times R/J$ is a ring homomorphism with kernel $I \cap J = I \cdot J$, so if we show that f is surjective, we obtain an isomorphism ψ as a direct consequence of 11.15. With $i = 1 - j$ as above, we have $f(i) = (0, 1)$, and likewise $f(j) = (1, 0)$. It follows that $f(r_1 j + r_2 i) = (r_1 + I, r_2 + J)$, so f is surjective. □

Arithmetic with ideals of a commutative ring R is very similar to arithmetic with elements of R. There are obvious definitions for sums and products of more than two ideals and distributive laws such as $(I + J) \cdot K = I \cdot K + J \cdot K$ (verify!). For the n-tuple product $I \cdot I \cdot \ldots \cdot I$, we also write I^n.

Exercise 10. Prove: $(I + J)^2 = I^2 + I \cdot J + J^2$.

At the end of the next section, we will see that arithmetic with ideals is, in some ways, *easier* than arithmetic with elements. See also Exercise 42.

EXERCISES.

In the exercises below, R always denotes a ring.

11. Let $x \in R$ be given, and let $y, z \in R$ satisfy $xy = zx = 1$. Prove: $y = z$. Conclude that a ring element has at most one (multiplicative) inverse.
12. Prove: $\mathrm{Map}(X, R)^* = \mathrm{Map}(X, R^*)$.
13. Let R be an integral domain. Prove: $R[X]^* = R^*$.
14. Let $R = \mathbf{Z}/4\mathbf{Z}$. Prove: for all $r \in R[X]$, we have $1 + 2r \in R[X]^*$. Is every unit $u \in R[X]^*$ of the form $u = 1 + 2r$?
15. Let R be an integral domain. Prove: $R[X, X^{-1}]^* = \{uX^k : u \in R^*, k \in \mathbf{Z}\} \cong R^* \times \mathbf{Z}$.
16. Prove: $\sum_{k \geq 0} r_k X^k \in R[[X]]^* \iff r_0 \in R^*$. Conclude that the ring $R((X))$ of Laurent series over R is a field if and only if R is a field.
17. Let p be a prime and $R_p \subset \mathbf{Q}$ be the set of fractions $\frac{m}{n}$ with $p \nmid n$. Prove that R_p is a subring of $\mathbf{Q}$, and determine R_p^*.
18. Show that every element of a finite ring R is a unit or a zero divisor. Conclude: every finite integral domain is a field.
19. Let $\mathbf{H} = \mathbf{R} + \mathbf{R} \cdot i + \mathbf{R} \cdot j + \mathbf{R} \cdot k$ be *Hamilton's quaternion algebra*, defined in 8.7. Prove that $\mathbf{H}$ is a non-commutative division ring.
[Hint: $(a + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2$.]
20. Prove that the identity $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$ for $a, b \in R$ and $n > 0$ ("Newton's binomial formula") holds in every commutative ring R. Show that, conversely, a ring in which Newton's binomial formula holds is commutative.
21. Show that the *center* $Z(R) = \{x \in R : \ rx = xr \text{ for all } r \in R\} \subset R$ of R is a *subring* of R. Prove: $Z(R[X]) = (Z(R))[X]$.
22. Show that for every ring R, there exists exactly one homomorphism $f : \mathbf{Z} \to R$. The non-negative generator of $\ker f$ is called the *characteristic* $\mathrm{char}(R)$ of R. Show that the characteristic of an integral domain is 0 or a prime. What is the characteristic of the zero ring?
23. Let $R' \subset R$ be a subring. Give a proof or a counterexample for
    a. R is a field $\implies R'$ is a field;
    b. R' is a field $\implies R$ is a field;
    c. R is an integral domain $\implies R'$ is an integral domain;
    d. R' is an integral domain $\implies R$ is an integral domain.
24. For what values of $n \in \mathbf{Z}_{\geq 0}$ is the determinant map $\det : \mathrm{Mat}_n(\mathbf{R}) \to \mathbf{R}$ a ring homomorphism?
25. Let $A \in \mathrm{Mat}_n(\mathbf{R}) \setminus \{0\}$ be a non-invertible matrix. Prove: A is both a left and a right zero divisor.
26. Show that the matrix ring $\mathrm{Mat}_n(\mathbf{R})$ has no non-trivial (two-sided) ideals.
27. Determine the center of the matrix ring $\mathrm{Mat}_n(\mathbf{R})$.

28. Let $V = \{(a_i)_{i=0}^{\infty} : a_i \in \mathbf{R}\}$ be the $\mathbf{R}$-vector space of $\mathbf{R}$-valued sequences with component-wise addition and scalar multiplication. Let R be the ring of $\mathbf{R}$-linear maps $V \to V$. Define $s, t, u \in R$ by

$$\begin{aligned}(a_0, a_1, a_2, a_3, \ldots) &\overset{s}{\longmapsto} (0, a_0, a_1, a_2, \ldots),\\ (a_0, a_1, a_2, a_3, \ldots) &\overset{t}{\longmapsto} (a_1, a_2, a_3, a_4, \ldots),\\ (a_0, a_1, a_2, a_3, \ldots) &\overset{u}{\longmapsto} (a_0, 0, 0, 0, \ldots).\end{aligned}$$

Prove the following statements:

a. $st \neq 1 \in R$ and $ts = 1 \in R$;
b. $us = 0 \in R$ and $tu = 0 \in R$;
c. s is a right zero divisor in R but not a left zero divisor; t is a left zero divisor in R but not a right zero divisor.

29. Let $R \neq 0$ be a commutative ring with group of units $R^* = \{1\}$.

a. Prove: R has characteristic 2.
b. Prove that the map $f : R \to R$ given by $f(r) = r^2$ is a ring homomorphism.
c. Prove that the map f in part (b) is injective. Is f necessarily surjective?

30. A *Boolean ring* is a ring where for every element x, we have the identity $x^2 = x$. Prove that every Boolean ring is either the zero ring or a commutative ring of characteristic 2. Show that $\mathbf{F}_2$ is the only Boolean ring that is a field.
[George Boole (1815–1864) was an English mathematician.]

31. Let X be a set. Then by Exercise 4.43, the symmetric difference $A \,\Delta\, B = (A \cup B) \setminus (A \cap B)$ defines the structure of an abelian group on the power set $\mathcal{P}(X)$. Show that the multiplicative operation $A \cdot B = A \cap B$ makes $\mathcal{P}(X)$ into a Boolean ring and makes the group isomorphism $\mathcal{P}(X) \xrightarrow{\sim} \mathrm{Map}(X, \mathbf{Z}/2\mathbf{Z})$ from 4.43 into a ring isomorphism to the ring of functions $\mathrm{Map}(X, \mathbf{Z}/2\mathbf{Z})$.

32. Let X and Y be sets and $f : X \to Y$ be a map. Then f induces the natural maps $f_* : \mathcal{P}(X) \to \mathcal{P}(Y)$ and $f^* : \mathcal{P}(Y) \to \mathcal{P}(X)$ defined by

$$f_*(A) = f[A] = \{f(x) : x \in A\}, \qquad f^*(B) = f^{-1}[B] = \{x \in X : f(x) \in B\}$$

for $A \subset X$ and $B \subset Y$.

a. Prove: f_* is a ring homomorphism if and only if f is a bijection.
b. Is f^* a ring homomorphism? Give a proof or a counterexample.

33. Let R be the ring of differentiable functions $\mathbf{R} \to \mathbf{R}$. Which of the following maps $R \to \mathbf{R}$ are homomorphisms of additive groups? Are they ring homomorphisms?

$$f \longmapsto f(0), \qquad f \longmapsto f'(0), \qquad f \longmapsto \textstyle\int_0^1 f(x)dx.$$

34. Show that by restriction, a ring homomorphism $f : R_1 \to R_2$ induces a group homomorphism $g = f|_{R_1^*} : R_1^* \to R_2^*$ on the group of units. Is g surjective if f is? Is f surjective if g is?

35. Show that the evaluation map $\phi_a : \sum c_i X^i \mapsto \sum c_i a^i$ gives a ring homomorphism $R[X] \to R$ if and only if a is contained in the center $Z(R)$ of R.

36. Let R be a commutative ring. Show that for $n \geq 1$ and $a = (a_1, a_2, \ldots, a_n) \in R^n$, the evaluation map $\phi_a : R[X_1, X_2, \ldots, X_n] \to R$ given by $\phi_a(f) = f(a)$ is a surjective ring homomorphism with kernel $(X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n)$.

37. Define $T \subset \mathrm{Mat}_2(R)$ by $T = \{\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) : a, b, d \in R\}$. Prove:

    a. T is a subring of $\mathrm{Mat}_2(R)$, and for $R \neq 0$, T is not commutative;
    b. $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in T^* \iff a \in R^*$ and $d \in R^*$;
    c. T^* is abelian $\iff R^* = \{1\}$.

    Does there exist a non-commutative ring R for which the group of units R^* is abelian?

38. Let K be a field.

    a. Can $K[X, X^{-1}]$ be identified with the field of fractions $Q(K[X])$ of $K[X]$?
    b. View the ring of power series $K[[X]]$ as a subring of the ring $K((X))$ of Laurent series over K. Prove: $Q(K[[X]]) = K((X))$.

39. A *ring without unit element*, called a *rng* by some, is an abelian group R written additively, endowed with a multiplicative operation $R \times R \to R$ that satisfies (R2) and (R3) but not (R1). In which of the cases below are we dealing with such a ring?

    a. the subset $2\mathbf{Z} \subset \mathbf{Z}$ of even numbers,
    b. an ideal $I \subset R$ that is not equal to $\{0\}$ or R,
    c. the subset $R \subset C(\mathbf{R})$ of continuous functions $\mathbf{R} \to \mathbf{R}$ with bounded support, (The *support* of $f \in C(\mathbf{R})$ is the set $\{x \in \mathbf{R} : f(x) \neq 0\}$.)
    d. an arbitrary abelian group R endowed with the "zero multiplication" $xy = 0$ for all $x, y \in R$.

40. Let R' be a ring without unit element. Define on $R = \mathbf{Z} \times R'$ the operations

$$(n_1, r_1) + (n_2, r_2) = (n_1 + n_2, r_1 + r_2),$$
$$(n_1, r_1) \cdot (n_2, r_2) = (n_1 n_2, r_1 r_2 + n_1 r_2 + n_2 r_1).$$

    (Use the obvious definition of nr for $n \in \mathbf{Z}$ and $r \in R'$.) Show that R is a ring and that R' can be identified with an ideal of R.

41. Let I and J be relatively prime ideals of a commutative ring R and $m, n \in \mathbf{Z}_{>0}$ be positive numbers. Prove that I^m and J^n are relatively prime.

42. Let I and J be as in the previous exercise, and suppose that we have $I \cdot J = K^n$, with K an ideal and $n \in \mathbf{Z}_{>0}$. Prove that there exist ideals I_0 and J_0 with $I = I_0^n$ and $J = J_0^n$.
    [Hint: look at $I_0 = I + K$, the "GCD" of I and K.]

43. Let i and j be relatively prime integers, and suppose that we have $i \cdot j = k^n$ with $k \in \mathbf{Z}$ and $n \in \mathbf{Z}_{>0}$. Prove: there exist $i_0, j_0 \in \mathbf{Z}$ and $\varepsilon \in \mathbf{Z}^* = \{\pm 1\}$ with $i = \varepsilon i_0^n$ and $j = \varepsilon j_0^n$. *Give an example of a commutative ring R where the identities $ai + bj = 1$ and $i \cdot j = k^n$ hold $(a, b, i, j, k \in R)$ but i and j are not products of a unit and an n-th power in R.

44. Define $I, J \subset \mathbf{Z}[X]$ by $I = (2, X)$ and $J = (3, X)$. Show that $\{i \cdot j : i \in I \text{ and } j \in J\}$ is not an ideal in $\mathbf{Z}[X]$ and, in particular, is not equal to $I \cdot J = (6, X)$.

45. Let I and J be ideals of a commutative ring R. Show that $\{xy : x \in I, y \in J\}$ is an ideal of R if I or J is a principal ideal.

46. For ideals I and J of a commutative ring R, prove the inclusion $(I+J)\cdot(I\cap J)\subset IJ$, and show that we have equality for $R=\mathbf{Z}$. *Do we always have equality?

47. Let X be a finite set, and take $R=\mathrm{Map}(X,\mathbf{C})$. For $x\in X$, define the ideal $I_x=\{f\in R : f(x)=0\}$. Prove:
    a. We have $R/I_x\cong\mathbf{C}$ for every $x\in X$.
    b. There is a ring isomorphism $R\cong\mathbf{C}^n$ for some $n\in\mathbf{Z}_{\geq 0}$.

48. State and prove a generalization of 11.19 for the case of pairwise relatively prime ideals $I_1, I_2, \ldots, I_n$.

49. Let $f: R_1\to R_2$ be a ring homomorphism and $I_2\subset R_2$ be an ideal of R_2. Prove that $I_1=f^{-1}[I_2]$ is an ideal of R_1 and that R_1/I_1 is isomorphic to a subring of R_2/I_2.

50. Let $R=\mathrm{Map}(\mathbf{Z}_{>0},\mathbf{Q})$ be the ring of $\mathbf{Q}$-valued functions on $\mathbf{Z}_{>0}$. We call $f\in R$ a *Cauchy function* if for all $\varepsilon\in\mathbf{Q}_{>0}$, there exists a number N such that we have $|f(m)-f(n)|<\varepsilon$ for all $m,n>N$. A function $f\in R$ is called a *zero function* if we have $\lim_{n\to\infty} f(n)=0$.

    Prove:
    a. The set C of Cauchy functions forms a subring $C\subset R$.
    b. The set I of zero functions is an ideal of C but not of R.
    c. We have $C/I\cong\mathbf{R}$.

51. Let $I\subset R$ be an ideal. Prove: every ideal of R/I is of the form J/I for an ideal $J\supset I$ of R, and for such a J, there is a natural isomorphism

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

    [This is the ring analog of 8.1. It allows us to calculate R/J "step-by-step": first take R/I, and then quotient by J/I.]

52. Let $R'\subset R$ be a subring and $I\subset R$ be an ideal. Prove the ring analog of 8.2:
    a. The intersection $R'\cap I$ is an ideal of R', and $R'+I=\{r+s : r\in R', s\in I\}$ is a subring of R.
    b. There is a natural isomorphism $R'/(R'\cap I)\xrightarrow{\sim}(R'+I)/I$.

53. The *commutator ideal* $[R,R]$ of a ring R is the left ideal generated by the elements of the form $yx-xy$ with $x,y\in R$. Prove that $[R,R]$ is a two-sided ideal and $R_{\mathrm{comm}}=R/[R,R]$ is a commutative ring with the property that every homomorphism $R\to R'$ to a commutative ring R' factors through R_{comm}, that is, can be written as a composition of homomorphisms $R\to R_{\mathrm{comm}}\to R'$.
    [This is the ring analog of 8.5.]

54. The two-sided ideal (S) generated by a subset $S\subset R$ is defined as the smallest two-sided ideal of R that contains S. Show that such an ideal always exists, and give an "explicit description" of (S) similar to the one we gave for commutative R.

*55. Prove that every left ideal of $\mathrm{Mat}_n(\mathbf{R})$ is of the form $L_V=\{M : \ker(M)\supset V\}$ and that every right ideal is of the form $R_V=\{M : \mathrm{im}(M)\subset V\}$, with $V\subset\mathbf{R}^n$ a subspace.

56. An *idempotent* of a ring R is an element $e\in R$ for which we have $e^2=e$. The trivial idempotents of a ring are the elements 0 and 1. Prove that if e is an idempotent of a

commutative ring R, then $1-e$ is also an idempotent and there is an isomorphism

$$R \xrightarrow{\sim} R/eR \times R/(1-e)R.$$

Show that the idempotents of a commutative ring R correspond bijectively to the "decompositions" of R as products $R = R_1 \times R_2$ and that the set of idempotents of R forms a *ring* under the "usual" multiplication and an "idempotent summation" given by

$$x \oplus y = (x-y)^2.$$

57. Determine the idempotents of the ring $\mathbf{Z}/1000\mathbf{Z}$.

58. Define a sequence of integers recursively by $x_0 = 5$ and, for $k \geq 1$,

$$x_k = [\text{remainder of } x_{k-1}^2(3-2x_{k-1}) \text{ when dividing by } 10^{2^k}].$$

All remainders are taken in the interval $[0, 10^{2^k})$, so $x_1 = 25$, $x_2 = 625$, $x_3 = 12890625$, etc. Prove: $(x_k \bmod 10^{2^k})$ is an idempotent of $\mathbf{Z}/10^{2^k}\mathbf{Z}$ for all $k \geq 1$. Does this also hold if we begin with $x_0 = 6$?

59. Let G be a group. Define the *augmentation map* $f : \mathbf{Z}[G] \to \mathbf{Z}$ on the group ring $\mathbf{Z}[G]$ by $f(\sum_{g\in G} a_g g) = \sum_{g\in G} a_g$. Show that f is a surjective ring homomorphism and that the two-sided ideal $I_G = \ker(f)$ is generated by the set $\{g-1 : g \in G\}$.
[The ideal I_G is called the *augmentation ideal* of $\mathbf{Z}[G]$.]

60. Let G and I_G be as in the previous exercise. Prove that the map

$$\begin{aligned} G_{\text{ab}} = G/[G,G] &\longrightarrow I_G/I_G^2 \\ g \bmod [G,G] &\longmapsto g-1 \bmod I_G^2 \end{aligned}$$

is an isomorphism of abelian groups. Here, G_{ab} is the abelianization from 8.5.

61. Let G be a cyclic group. Show that the augmentation ideal I_G is a principal ideal.

62. Let R be a ring for which the underlying additive group R^+ is cyclic. Prove: $R \cong \mathbf{Z}/n\mathbf{Z}$ (as rings) for some $n \in \mathbf{Z}_{\geq 0}$.

63. Determine all (isomorphism types of) rings R for which R^+ is a Klein four-group.

64. The sequence $(F_n)_{n=0}^{\infty}$ of *Fibonacci numbers* is defined recursively by $F_0 = 0$, $F_1 = 1$, and $F_{n+2} = F_{n+1} + F_n$ (for $n \geq 0$).
    a. Construct a ring R that contains $\mathbf{Z}$ as a subring and contains an element ϑ with the property that for all positive integers n, we have $\vartheta^n = F_{n-1} + F_n \cdot \vartheta$.
    b. Does there exist a ring R as in (a) such that there is only one $\vartheta \in R$ with the stated property?

65. Let R be the set of all 2×2 matrices $\left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$ with $a \in \mathbf{Z}/4\mathbf{Z}$, $b \in 2\mathbf{Z}/4\mathbf{Z}$, $c \in \mathbf{Z}/2\mathbf{Z}$. Two elements of R are added component-wise and multiplied as matrices, where the necessary operations on the entries of the matrices are induced by the ring operations on $\mathbf{Z}$. This way, R becomes a ring (verify; this is not part of the exercise...). Let $I = \{x \in R : x^2 = 0\}$.
    a. Prove that I is a two-sided ideal of R with $\#I = 4$ and that the ring R/I is isomorphic to the product ring $(\mathbf{Z}/2\mathbf{Z}) \times (\mathbf{Z}/2\mathbf{Z})$.

b. Prove: there do not exist any elements $y \in R$ with $I = Ry$, but there is an element $z \in R$ with $I = zR$.

66. Let R be a ring. The opposite ring R^{opp} is made from R by defining a multiplication "$\star$" on the additive group of R by $x \star y = yx$. Show that R^{opp} is a ring, and construct a ring R for which R is *not* isomorphic to R^{opp}.

67. *(Goursat's lemma for rings)* Let R_1 and R_2 be rings and $A \subset R_1 \times R_2$ be a subring. Prove that there exist subrings $A_1 \subset R_1$ and $A_2 \subset R_2$, as well as an ideal I_1 of A_1, an ideal I_2 of A_2, and a ring isomorphism $\phi : A_1/I_1 \xrightarrow{\sim} A_2/I_2$, such that we have

$$A = \{(a_1, a_2) \in A_1 \times A_2 : \phi(a_1 + I_1) = a_2 + I_2\}.$$

(Compare with Exercise 8.37.)

68. Prove that the ideals of $R_1 \times R_2$ are of the form $I_1 \times I_2$, with $I_i \subset R_i$ an ideal.
[So there is no "Goursat's lemma" for ideals.]

69. Suppose that R is a ring with the property that every element of R has finite order in the additive group of R. Prove that there is a positive integer m such that for every $a \in R$, we have $m \cdot a = 0$. Construct an abelian group that is not the additive group of a ring.

70. Let $A_X = \mathbf{Z}((X))$ be the ring of Laurent series in X over $\mathbf{Z}$ and $A_{X,Y} = A_X((Y))$ be the ring of Laurent series in Y over A_X. Prove: $X - Y \in A_{X,Y}^*$.

71. Also define $A_Y = \mathbf{Z}((Y))$ and $A_{Y,X} = A_Y((X))$ as in the previous exercise, and let B be the additive group of all formal expressions $\sum_{i,j \in \mathbf{Z}} a_{ij} X^i Y^j$ with $a_{ij} \in \mathbf{Z}$, under component-wise addition.

    a. Show that the additive groups of the rings $A_{X,Y}$ and $A_{Y,X}$ can be viewed as subgroups of B, that the intersection $C = A_{X,Y} \cap A_{Y,X}$ in B is a subring of both $A_{X,Y}$ and $A_{Y,X}$, and that the two resulting multiplications on C coincide.
    b. Prove: $X - Y \in C$ and $X - Y \notin C^*$.
    c. Does there exist a subgroup D of B that contains both $A_{X,Y}$ and $A_{Y,X}$, on which we can define a multiplication such that D is a ring containing both the rings $A_{X,Y}$ and $A_{Y,X}$ as subrings (with the same multiplication)?

72. *(Localization)* Let R be a commutative ring and $S \subset R$ be a subset that contains 1 and is closed under multiplication. Define an equivalence relation on $R \times S$ by

$$\frac{r}{s} \sim \frac{r'}{s'} \iff (rs' - r's)s'' = 0 \quad \text{for an element } s'' \in S.$$

    a. Show that this is an equivalence relation.

    Denote the equivalence class of $(r, s) \in R \times S$ by $\frac{r}{s}$, and define an addition and multiplication on the set $S^{-1}R$ of equivalence classes by the well-known "fraction formulas" from 11.10.

    b. Show that this makes $S^{-1}R$ into a commutative ring and the map $R \to S^{-1}R$ given by $r \mapsto \frac{r}{1}$ into a ring homomorphism.
    c. Prove: $S^{-1}R = 0 \iff 0 \in S$.
    d. Give an example where $S^{-1}R$ is not the zero ring and the map $R \to S^{-1}R$ is *not* injective.

12 Principal Ideal Domains

Principal ideal domains are integral domains in which all ideals are principal. The ring $\mathbf{Z}$ is an example, and many properties of $\mathbf{Z}$ turn out to hold for arbitrary principal ideal domains. Among these is the fundamental theorem 6.7 on the *unique prime factorization* of integers.

▶ Division with Remainder

In 6.2, we proved that $\mathbf{Z}$ is a principal ideal domain using the division with remainder 6.1. This procedure can be used in many situations where a suitable notion of the "size" of elements exists. In the polynomial ring $R[X]$, the *degree* is such a notion. The degree $\deg(f)$ of a polynomial $f = \sum a_k X^k \in R[X]$ is the greatest index $k \in \mathbf{Z}_{\geq 0}$ with $a_k \neq 0$. Such a k exists for all polynomials except the zero polynomial; we take $\deg(0) = -1$. For f of degree $n \geq 0$, the coefficient a_n is called the *leading coefficient* of f. If it equals 1, then we call f *monic*. For rings R without non-trivial zero divisors, the leading coefficient of a product fg in $R[X]$ is the product of the leading coefficients of f and g, and we have the well-known rule

$$\deg(fg) = \deg(f) + \deg(g)$$

for the product of polynomials different from 0.

12.1. Division with remainder for polynomials. *Let R be a ring and $f, g \in R[X]$ be polynomials. Suppose that $g \neq 0$ has a leading coefficient in R^*. Then there exist polynomials $q, r \in R[X]$ with*

$$f = qg + r \qquad \text{and} \qquad \deg(r) < \deg(g).$$

Proof. We carry out the proof by induction on the degree of f. For $\deg(f) = -1$, we have $f = 0$, and the choice $q = r = 0$ suffices. Now, suppose that division with remainder is possible for polynomials f of degree $\deg(f) < n \in \mathbf{Z}_{\geq 0}$. For f of degree n, we now informally write $f = a_n X^n + \ldots$, and likewise $g = b_m X^m + \ldots$ with $m = \deg(g)$ and $b_m \in R^*$. If $m > n$, we can take $q = 0$ and $r = f$, and there is nothing to prove. If $m \leq n$, we look at the difference $f - a_n b_m^{-1} X^{n-m} g$. Since $a_n b_m^{-1} X^{n-m} g$ has degree n and leading coefficient a_n just like f, the degree of this difference is less than n. So by the induction hypothesis, we have

$$f - a_n b_m^{-1} X^{n-m} g = q_1 g + r \qquad \text{with} \qquad \deg(r) < \deg(g)$$

for some $q_1, r \in R[X]$. For $q = a_n b_m^{-1} X^{n-m} + q_1$, we now have $f = qg + r$; this is the division with remainder for f itself. □

Exercise 1. Show that q and r in 12.1 are uniquely determined by f and g.

The ring R in 12.1 is completely arbitrary and may have non-trivial zero divisors or even be non-commutative. In all these cases, the *quotient* q and the *remainder* r are

uniquely determined by f and g, and the proof of 12.1 describes how to calculate q and r from f and g using *long division*: subtract a multiple of g from f for which the "leading term" cancels out, and continue in this way until you are left with a remainder of degree less than that of g.

Exercise 2. Calculate q and r in 12.1 for $R = \mathbf{Z}$ and $f = 4X^4+3X^3+2X^2+X$ and $g = X^2+2X+3$.

If in 12.1, we take g to be a linear polynomial $X - a$, then the remainder r is a constant polynomial with value $f(a)$. For commutative R, this follows by substituting $X = a$. However, the argument in 11.16 shows that, as an "R-linear combination" of expressions $X^i - a^i$, the polynomial $f - f(a)$ is always of the form $q \cdot (X - a)$.

12.2. Corollary. *Let R be a ring. Then for every $f \in R[X]$ and $a \in R$, there exists a polynomial $q \in R[X]$ with*

$$f = q \cdot (X - a) + f(a). \qquad \square$$

12.3. Corollary. *Let R be an integral domain, and let $a_1, a_2, \ldots, a_n \in R$ be n distinct zeros of $f \in R[X]$. Then we have*

$$f = q \cdot (X - a_1)(X - a_2) \ldots (X - a_n) \qquad \text{with } q \in R[X].$$

For $f \neq 0$, the number of zeros of f in R is at most $\deg(f)$.

Proof. We apply induction on n. For $n = 1$, we have a special case of 12.2. For $n > 1$, we write $f = q_1 \cdot (X - a_n)$ with $q_1 \in R[X]$ using 12.2. Since R is commutative, substituting a zero a_i with $i \neq n$ gives the relation $q_1(a_i) \cdot (a_i - a_n) = 0$. Since R has no non-trivial zero divisors, it follows that all $n-1$ zeros a_i with $i < n$ are zeros of the polynomial q_1. By induction, we now have $q_1 = q \cdot (X - a_1)(X - a_2) \ldots (X - a_{n-1})$, which shows that $f = q_1 \cdot (X - a_n)$ is of the required form. In particular, we see that for an integral domain R, a polynomial $f \in R[X] \setminus \{0\}$ with n distinct zeros has degree at least n. $\square$

12.4. Corollary. *Let R be an integral domain and $H \subset R^*$ be a finite subgroup of the group of units of R. Then H is cyclic.*

Proof. Recall that for every positive divisor $d|n$, a *cyclic* group $C = \langle y \rangle$ of order n contains exactly one subgroup of order d, generated by $x = y^{n/d}$, and that the elements of C of order d are the powers x^i of x with exponent $i \in \{1, 2, \ldots d\}$ relatively prime to d. In particular, for every $d|n$, C contains exactly $\varphi(d)$ elements of order d, with φ Euler's φ-function. Taking the sum over all $d|n$ gives $\sum_{d|n} \varphi(d) = \#C = n$, Gauss's formula.

Now, for $H \subset R^*$ in 12.4 of order n and $d|n$ a positive divisor, let $\psi(d)$ be the number of elements of H of order d. If $x \in H$ has order d, then the d different powers $x, x^2, x^3, \ldots, x^d = 1$ of x are zeros of $X^d - 1$ in R. By 12.3, there are then no other zeros of $X^d - 1$ in R, so the elements of H of order d are exactly the $\varphi(d)$ powers x^i of

x with exponent i relatively prime to d. We conclude that $\psi(d)$ is equal to $\varphi(d)$ if H contains an element of order d and equal to 0 otherwise. We now have

$$n = \#H = \textstyle\sum_{d|n} \psi(d) \leq \sum_{d|n} \varphi(d) = n,$$

and it follows that $\psi(d) = \varphi(d)$ for all $d|n$. In particular, we have $\psi(n) = \varphi(n) > 0$, so H contains an element of order n and is cyclic. □

12.5. Corollary. *The group of units F^* of a finite field F is cyclic.* □

The only finite fields we have encountered are the fields $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ corresponding to the primes $p \in \mathbf{Z}$. A number $x \in \mathbf{Z}$ for which $(x \bmod p)$ is a generator of $\mathbf{F}_p^* = (\mathbf{Z}/p\mathbf{Z})^*$ is called a *primitive root* modulo p. We already mentioned the existence of such elements in §7 and §9 (in connection with the structure of the group $(\mathbf{Z}/p^k\mathbf{Z})^*$) and in §10 (in connection with the construction of non-abelian groups of order pq).

For every integer $n > 0$, the complex zeros of $X^n - 1$ form the subgroup of n-th *roots of unity* in $\mathbf{C}^*$. This is a cyclic group of order n generated by $e^{2\pi i/n}$. The only non-trivial finite subgroup of $\mathbf{R}^*$ (or $\mathbf{Z}^*$) is the cyclic sign group $\{\pm 1\}$ of order 2.

12.6. Corollary. *Let K be a field. Then $K[X]$ is a principal ideal domain.*

Proof. We imitate the proof of 6.2. Let $I \subset K[X]$ be an ideal. For $I = 0$, the zero element 0 is a generator of I. For $I \neq 0$, we choose a non-zero polynomial $g \in I$ of minimal degree. If $f \in I$ is any other polynomial, then by 12.1, there exist polynomials $q, r \in K[X]$ with $f = qg + r$ and $\deg(r) < \deg(g)$. Since $r = f - qg \in I$, by the minimality of $\deg(g)$, we have $r = 0$, so $f = qg$. Consequently, we have $I = (g)$. □

The fact that domains having 'division with reminder' are principal ideal domains is a generality that can be formalized as in Exercise 38.

▶ Unique Factorization

In Theorem 12.11, we will prove that in an arbitrary principal ideal domain R, every element $x \neq 0$ can be decomposed as an essentially unique product of prime elements. For $R = \mathbf{Z}$, this is Theorem 6.7. For $R = K[X]$ from 12.6, it means that every polynomial $f \neq 0$ can be decomposed essentially uniquely as a product of *irreducible polynomials* in $K[X]$.

We speak of "essentially" unique decompositions because the factors in a decomposition can always be multiplied by units of the ring. For $R = \mathbf{Z}$, we could always choose the primes *positive*, in which case every number $x \neq 0$ has a unique decomposition as a product of a unit in $\mathbf{Z}^* = \{\pm 1\}$ and finitely many primes. For $R = K[X]$, the group of units $R^* = K^*$ consists of the constant polynomials (Exercise 11.13), so that irreducible polynomials in a decomposition are fixed up to multiplication by a non-zero constant.

For the record, let us mention that in a commutative ring R, an element $x \in R$ is called a *divisor* of $y \in R$ if we have $y = xz$ for some $z \in R$. We also say that y is a *multiple* of x. The units of R divide 1 and therefore every other element of R.

12.7. Definition. *An element p of an integral domain R is called irreducible if it is not a unit and satisfies the property*

$$p = xy \in R \quad \Longrightarrow \quad x \in R^* \quad \text{or} \quad y \in R^*.$$

Note that, up to sign, the irreducible elements of $\mathbf{Z}$ are exactly the prime numbers. In polynomial rings we usually speak of *irreducible polynomials*. By definition, the units in an integral domain are not irreducible, just like 1 is not taken to be a prime number in $\mathbf{Z}$.

12.8. Lemma. *Let R be a principal ideal domain and $x \in R$ be different from 0. Then there exists a factorization*

$$x = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t$$

of x as a product of a unit $u \in R^*$ and finitely many irreducible elements $p_i \in R$.

Proof. We give a proof by contradiction. Suppose that there exists an element $x \neq 0$ of R that does *not* have a decomposition of the mentioned type. Then x is not a unit because if it were, $x = x$ would be a decomposition (with $t = 0$ irreducible elements). Also, x is not irreducible because if it were, $x = 1 \cdot x$ would give a decomposition. It now follows from Definition 12.7 that there exists an identity $x = yz$ in R where y and z are *not* units. If y and z both have a decomposition of the desired type, then combined, those give a decomposition of x; after all, we only need to multiply the units from both decompositions to obtain a new unit in the decomposition of x. We conclude that y or z, say y, also does not have a decomposition of the desired type.

It follows from $x = yz$ that we have an inclusion of ideals $(x) \subset (y)$. This inclusion is not an equality. Otherwise, it would follow from $y \in (x)$, say $y = wx$, that $x = zwx$ and therefore, since R is an integral domain, that $zw = 1$ and $z \in R^*$.

If we apply the argument for x to $x_1 = y$, we find a strict inclusion of ideals $(x_1) \subsetneq (x_2)$ for an element x_2 that again has no decomposition. Continuing in this way, we obtain an infinitely long strictly increasing chain

$$(x) \subsetneq (x_1) \subsetneq (x_2) \subsetneq (x_3) \subsetneq (x_4) \subsetneq \ldots$$

of principal ideals in R generated by elements x_i without decomposition in R.

We conclude the proof by showing that such a chain cannot exist. Namely, let $I = \bigcup_{k \geq 1} (x_k)$ be the union of the ideals (x_i). Then I is also an ideal of R (why?), and we have $I = (x_\infty)$ for some $x_\infty \in R$. By the definition of I, there is a value of k with $x_\infty \in (x_k)$; this gives $I = (x_\infty) = (x_k)$ for this k, in contradiction with the assumption. □

In concrete principal ideal domains R, it is often immediately clear that we can find a decomposition as in 12.8 in finitely many steps. After all, if x is not irreducible and not a unit, then we can write $x = yz$ with y and z not units and continue inductively by decomposing y and z. In many cases, this process is clearly finite because the elements y and z in such a splitting are "smaller" than x. For example, in $\mathbf{Z}$, the absolute value decreases, while in $K[X]$, the degree of the polynomials decreases.

Let us now prove that in principal ideal domains, the decomposition found in 12.8 is unique up to multiplication of the irreducible elements p_i by units. For this, we need the *prime property* of irreducible elements in a principal ideal domain that was formulated in 6.6.

12.9. Definition. *A prime element of an integral domain R is an element $p \neq 0$ that is not a unit and satisfies the prime property*

$$p|xy \qquad \Longrightarrow \qquad p|x \quad \text{or} \quad p|y.$$

The prime elements of an integral domain are always irreducible. After all, if for a prime element p, we have the equality $p = yz$, then p is a divisor of y or z. If we have, for example, $y = pw$, then we find that $p = yz = pwz$ and $wz = 1$, so z is a unit and p is irreducible.

At the end of this section, we will encounter integral domains with the complicating property that they contain irreducible elements that are *not* prime. They form the historical reason for considering *ideals* of rings.

12.10. Lemma. *Every irreducible element of a principal ideal domain is prime.*

Proof. We imitate the proof of 6.6. Suppose that p is an irreducible element of a principal ideal domain R and that p is a divisor of xy but not of x. We must show that, in this case, p divides y.

Since p is not a divisor of x, we have $x \notin (p)$, which means that the ideal $(p)+(x) = (p, x) \subset R$ is strictly greater than (p). Since it is a principal ideal, it has a generator $z \in R$. Then we have $p = zw$ for some $w \in R$, and since the inclusion $(p) \subsetneq (z)$ is strict, w is *not* a unit of R. By the irreducibility of p, we now have $z \in R^*$, so $(z) = (p, x) = R$. In particular, the element $1 \in (p, x) = (p) + (x)$ can be written as $1 = ap + bx$ with $a, b \in R$. The element $y = (ap + bx)y = apy + bxy$ is now a sum of elements apy and bxy that are each divisible by p. It follows that y itself is also divisible by p. □

Exercise 3. Prove: if p is an irreducible element of a principal ideal domain R, then R/pR is a field.

We can now prove our main result for principal ideal domains.

12.11. Theorem. *Let R be a principal ideal domain. Then every element $x \neq 0$ of R can be written as a product*

$$x = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t$$

of a unit $u \in R^*$ and finitely many irreducible elements $p_i \in R$; this notation is unique up to the order and multiplication by units.

Proof. By 12.8, it suffices to prove the uniqueness of the decomposition. So suppose that we have two decompositions

$$up_1p_2 \ldots p_s = vq_1q_2 \ldots q_t,$$

with $u, v \in R^*$ and all p_i and q_j irreducible. We want to prove that $s = t$ holds and that the irreducible elements in the two decompositions are the same up to units. We carry out the proof by induction on s. For $s = 0$, we have $t = 0$ because no irreducible element q_i divides the unit u; otherwise, q_i would be a unit. In that case, we have $u = v$, and we are done.

For $s > 0$, we note that by the prime property of p_s proved in 12.9, the element p_s divides one of the elements q_i, say $q_t = p_s w$. Since p_s is not a unit, by the irreducibility of q_t, the element w is a unit. Since we are in an integral domain, we can "cross out" p_s to obtain the equality $up_1p_2 \dots p_{s-1} = vwq_1q_2 \dots q_{t-1}$. By induction, we now have $s - 1 = t - 1$, and we have the same irreducible elements on both sides up to units and the order. So this also holds for the original decomposition. This proves the *uniqueness* of the decomposition. □

For $R = \mathbf{Z}$, the ambiguity of the decomposition caused by the units is limited to a simple sign issue. For $R = K[X]$, the group of units K^* consists of the constant polynomials. A polynomial with leading coefficient equal to 1 is called *monic*. Every irreducible polynomial in $K[X]$ is monic after multiplication by a suitable constant. We can therefore formulate Theorem 12.11 for $K[X]$ as follows.

12.12. Corollary. *Let K be a field. Then, up to the order, every polynomial $f \neq 0$ in $K[X]$ can be written uniquely as a product $f = c \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t$ of a constant $c \in K^*$ and finitely many monic irreducible polynomials $p_i \in K[X]$.* □

In the next section, we elaborate on the explicit *calculation* of the factorization in 12.12. As for integers, in practice, this is a non-trivial problem.

▶ Prime Ideals

The proofs of the main ingredients 12.8 and 12.9 in the proof of 12.11 show that divisibility issues in commutative rings can easily be formulated in terms of *inclusions of ideals*. After all, x divides y if and only if the principal ideal (x) contains the principal ideal (y). The adage "dividing is containing" for ideals sometimes leads to a conflicting notion of "large" and "small." For example, in $\mathbf{Z}$, a "small" number $n > 0$ corresponds to a set-theoretic "large" ideal $n\mathbf{Z}$ whose *index* n in $\mathbf{Z}$ is small.

Theorem 12.11 also looks more transparent in terms of ideals. The reason is that multiplication by units becomes invisible when we switch from elements to ideals. In integral domains, we also have a converse of this statement.

12.13. Lemma. *For elements x, y in an integral domain R, we have*

$$(x) = (y) \qquad \Longleftrightarrow \qquad x = uy \quad \text{for some} \quad u \in R^*.$$

Proof. If there exists a unit u with $x = uy$, we also have $y = u^{-1}x$, so in addition to $(x) \subset (y)$, we have $(y) \subset (x)$. It follows that $(x) = (y)$.

If $(x) = (y)$ holds, then there exist $u, v \in R$ with $x = uy$ and $y = vx$. This gives $x = uvx$ and $(1 - uv)x = 0$. For $x = 0$, we have $y = 0$ and can take $u = 1$. For $x \neq 0$, we have $uv = 1$ because R is an integral domain, and therefore $u \in R^*$. □

The condition in 12.13 that R is an integral domain cannot be left out (Exercise 36). Elements as in 12.13 that generate the same ideal in an integral domain R are said to be *associated* in R.

12.14. Definition. *An ideal I in a commutative ring R is called a prime ideal if it is not equal to R and if for all $x, y \in R$, we have*

$$xy \in I \qquad \Longrightarrow \qquad x \in I \quad \text{or} \quad y \in I.$$

We obtain a more compact version of Definition 12.14 by writing the implication above in the quotient ring R/I as $(\overline{xy} = 0 \Rightarrow \overline{x} = 0 \text{ or } \overline{y} = 0)$. This leads to the equivalence

(12.15) $$I \subset R \text{ is a prime ideal } \Longleftrightarrow R/I \text{ is an integral domain}$$

for an ideal I in a commutative ring R. In particular, the zero ideal (0) is prime if and only if R is an integral domain. If in 12.14, we take I to be the non-zero principal ideal (p) generated by an element $p \neq 0$ of an integral domain R, then by 12.9, we have

(12.16) $$(p) \text{ is a prime ideal } \Longleftrightarrow p \text{ is a prime element.}$$

Combining this with 12.13 and 12.16, we see that Theorem 12.11 has the following compact formulation in terms of ideals.

12.17. Theorem. *Every ideal $I \neq 0$ of a principal ideal domain R can be written as a product of prime ideals that is unique up to the order.* □

In a principal ideal domain R, we call elements $a, b \in R$ *relatively prime* if we have $(a) + (b) = R$. More generally, as we did in 6.3.3 for $R = \mathbf{Z}$, we can define a *greatest common divisor* $d = \mathrm{GCD}(a, b)$ and a *least common multiple* $k = \mathrm{LCM}(a, b)$ of elements $a, b \in R$ by the equalities

(12.18) $$(a) + (b) = (d) \qquad \text{and} \qquad (a) \cap (b) = (k).$$

By 12.13, this fixes d and k up to multiplication by units. For $R = \mathbf{Z}$ and $R = K[X]$, we can choose unique generators of ideals by requiring that they be, respectively, *non-negative* and *monic* (or 0).

▶ Gaussian Integers

We now give a classic number-theoretic application of the theory in this section. We take the ring R in 12.11 to be the ring $\mathbf{Z}[i]$ of *Gaussian integers*. This is the subring of the field $\mathbf{C}$ of complex numbers defined by $\mathbf{Z}[i] = \{a+bi \in \mathbf{C} : a, b \in \mathbf{Z}\}$. We easily see that this is a subring of $\mathbf{C}$. It can be visualized as the set of "standard lattice points" in the complex plane. We define the *norm* $N : \mathbf{Z}[i] \to \mathbf{Z}$ by

$$N(\alpha) = \alpha\overline{\alpha}, \qquad \text{that is,} \qquad N(a+bi) = (a+bi)(a-bi) = a^2 + b^2.$$

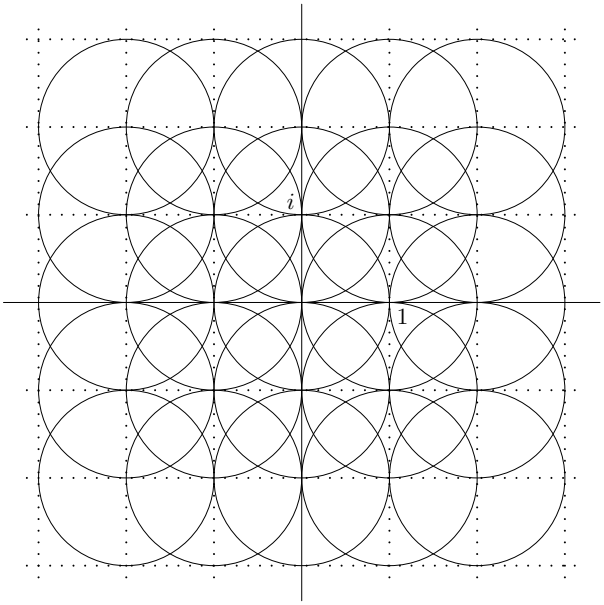
This norm is the *square* of the usual absolute value on $\mathbf{C}$, and like the usual absolute value, it satisfies the multiplicative property $N(xy) = N(x)N(y)$. We can extend it to a multiplicative map $N : \mathbf{Q}(i) \to \mathbf{Q}$ on the field of fractions $\mathbf{Q}(i) = \{a+bi \in \mathbf{C} : a, b \in \mathbf{Q}\}$ of $\mathbf{Z}[i]$.

12.19. Theorem. *The ring* $\mathbf{Z}[i]$ *is a principal ideal domain; the group of units* $\mathbf{Z}[i]^*$ *is cyclic of order* 4 *and is generated by* i.

Proof. We must show that for any two numbers α and $\beta \neq 0$ in $\mathbf{Z}[i]$, there exist numbers $q, r \in \mathbf{Z}[i]$ that satisfy $\alpha = q\beta + r$ and $N(r) < N(\beta)$. If we write this identity as

$$\frac{\alpha}{\beta} - q = \frac{r}{\beta} \qquad \text{with} \quad N(\frac{r}{\beta}) < 1,$$

then we see that we must approximate the fraction $\frac{\alpha}{\beta} \in \mathbf{Q}(i)$ so closely by a number $q \in \mathbf{Z}[i]$ that the difference $\frac{\alpha}{\beta} - q$, which is indeed of the form $\frac{r}{\beta}$ with $r \in \mathbf{Z}[i]$, is a complex number of absolute value less than 1. If we draw the elements of $\mathbf{Z}[i]$ as lattice points in the plane, this means that the open disks with radius 1 about these lattice points must cover the entire complex plane. We can easily see this in a figure.



Those who do not find proofs based on figures convincing can observe that for every number $z = x + yi \in \mathbf{C}$, there exist integers x_0 and y_0 with $|x - x_0| \leq 1/2$ and $|y - y_0| \leq 1/2$. Then $q = x_0 + y_0 i$ is an element of $\mathbf{Z}[i]$, and we have

$$|z - q|^2 = |(x - x_0) + (y - y_0)i|^2 = |x - x_0|^2 + |y - y_0|^2 \leq (1/2)^2 + (1/2)^2 < 1.$$

Now that we know that $\mathbf{Z}[i]$ admits division with remainder just like $\mathbf{Z}$ and $K[X]$, it follows as in 12.6 or Exercise 38 that $\mathbf{Z}[i]$ is a principal ideal domain.

An element $\alpha = a+bi \in \mathbf{Z}[i]$ that divides 1 has a norm $N(\alpha) = a^2+b^2$ that divides $N(1) = 1$. The four integer solutions of $N(a + bi) = a^2 + b^2 = 1$ give the elements of the group of units $\mathbf{Z}[i]^* = \{\pm 1, \pm i\}$. By 12.4, this group is cyclic, with generator i. □

Exercise 4. Determine a quotient q and a remainder r for $\alpha = 10 + 3i$ and $\beta = 3 + 4i$. Are q and r unique?

Exercise 5. Prove: $N(\alpha)$ is prime in $\mathbf{Z} \Rightarrow \alpha$ is prime in $\mathbf{Z}[i]$. Does the converse hold?

It follows from 12.11 and 12.19 that every element $\alpha \in \mathbf{Z}[i]$ can be decomposed as a product of prime elements. Since every number $\alpha \in \mathbf{Z}[i]$ divides the integer $N(\alpha) = \alpha\overline{\alpha}$, every prime element of $\mathbf{Z}[i]$ divides a positive integer. If we write this integer as a product of primes, then it follows that every prime element of $\mathbf{Z}[i]$ divides a prime number in $\mathbf{Z}$. To find them all, it therefore suffices to decompose the primes from $\mathbf{Z}$ in $\mathbf{Z}[i]$.

12.20. Theorem. *Let $p \in \mathbf{Z}$ be prime. Then p can be decomposed as follows in $\mathbf{Z}[i]$:*

1. *For $p = 2$, we have $2 = -i \cdot (1+i)^2$.*
2. *For $p \equiv 1 \bmod 4$, we have $p = \pi\overline{\pi}$, with π and $\overline{\pi}$ non-associated and prime in $\mathbf{Z}[i]$.*
3. *For $p \equiv 3 \bmod 4$, p is a prime element of $\mathbf{Z}[i]$.*

Proof. If p is a prime that is not irreducible in $\mathbf{Z}[i]$, then it can be written as $p = \alpha\beta$ with $\alpha, \beta \notin \mathbf{Z}[i]^*$. It then follows from $N(\alpha)N(\beta) = N(p) = p^2$ that we must have $N(\alpha) = N(\beta) = p$. So the question is for which p there exists an element $\pi \in \mathbf{Z}[i]$ of norm $N(\pi) = \pi\overline{\pi} = p$.

For primes $p \equiv 3 \bmod 4$, the equation $N(a+bi) = a^2 + b^2 = p$ has no integer solutions. After all, squares in $\mathbf{Z}$ are in the residue classes 0 mod 4 and 1 mod 4, so a sum of two squares is not congruent to 3 mod 4. We conclude that a prime $p \equiv 3 \bmod 4$ is irreducible in $\mathbf{Z}[i]$ and therefore a prime element of $\mathbf{Z}[i]$ by 12.10.

For primes $p \equiv 1 \bmod 4$, the group $\mathbf{F}_p^* = (\mathbf{Z}/p\mathbf{Z})^*$ is cyclic by 12.5, and its order $p-1$ is divisible by 4. As already recalled in the proof of 12.4, this means that there exists an element $\overline{x} \in \mathbf{F}_p^*$ of order 4. The square $\overline{x}^2$ is then the element of order 2 in $\mathbf{F}_p^*$, which is $-\overline{1}$. For an element $x \in \mathbf{Z}$ in the residue class $\overline{x} \in \mathbf{F}_p^*$, we now have $x^2 + 1 \equiv 0 \bmod p$, so

$$p \,|\, x^2 + 1 = (x+i)(x-i) \in \mathbf{Z}[i].$$

However, it is clear that p is not a divisor of $x+i$ or $x-i$ in $\mathbf{Z}[i]$. We see that p is *not* a prime element of $\mathbf{Z}[i]$, and by 12.10, p is then also not irreducible in $\mathbf{Z}[i]$. As we have already seen, there then exists an element $\pi = a + bi \in \mathbf{Z}[i]$ of norm $N(\pi) = \pi\overline{\pi} = a^2 + b^2 = p$. Since the norm of both π and $\overline{\pi}$ is equal to p, for a decomposition $\alpha\beta$ of π or $\overline{\pi}$, we have the equality $N(\alpha\beta) = N(\alpha)N(\beta) = p$, so $N(\alpha) = 1$ or $N(\beta) = 1$. We conclude that π and $\overline{\pi}$ are both irreducible in $\mathbf{Z}[i]$ and therefore prime. This also applies to the case $p = 2 = 1^2 + 1^2$, with $\pi = 1 + i$.

If $\pi = a + bi$ and $\overline{\pi} = a - bi$ are associated, then by 12.13, we have the equality $a + bi = u(a - bi)$ for a unit $u \in \mathbf{Z}[i]^*$. In the case $p \equiv 1 \bmod 4$, the four possible choices $u = \pm 1$ and $u = \pm i$ lead to the obviously false conclusions $b = 0$, $a = 0$, and $a = \pm b$. For $p = 2$ however, we can have $a = b$ for $\pi = 1 + i = i(1 - i)$: in this case $\pi = 1 + i$ and $\overline{\pi} = 1 - i$ are associate elements, and $2 = -i(1+i)^2$ is the unique prime number that is (up to units) equal to the square of the prime element π. □

Exercise 6. Prove: for $p \equiv 1 \bmod 4$ and $x \in \mathbf{Z}$ as above, $\mathrm{GCD}(p, x - i)$ is an element of norm p.

Fermat already wrote in a letter to Mersenne in 1640 that he could prove that every prime $p \equiv 1 \bmod 4$ can be written as the sum of two squares—a statement equivalent to 12.20.2. The first complete surviving proof was given by Euler in 1749.[2]

Exercise 7. Show that the notation $p = a^2 + b^2$ for a prime $p \equiv 1 \bmod 4$ is uniquely determined up to choosing signs for a and b and interchanging a and b.

More generally, a positive integer n can be written as a sum of two squares if and only if the exponent $\mathrm{ord}_p(n)$ of p in the factorization of n is *even* for every prime $p \equiv 3 \bmod 4$ (Exercise 54).

Theorem 12.20 gives us enough information to decompose an arbitrary element $x \in \mathbf{Z}[i]$. We illustrate this by factoring $x = 174 - 582\,i$ explicitly.

We first calculate $\mathrm{GCD}(174, 582) = 6$, for example as in 6.14 by applying the Euclidean algorithm, and write $174 - 582i = 6(29 - 97i)$. The decomposition of $6 = 2 \cdot 3$ is done by applying 12.20. For the factor $\alpha = 29 - 97i$, which has no rational prime divisors in $\mathbf{Z}[i]$, we first look at the norm

$$N(\alpha) = N(29 - 97i) = 29^2 + 97^2 = 841 + 9409 = 10250 = 2 \cdot 5^3 \cdot 41.$$

We deduce that α is the product of prime elements of norms 2, 5, and 41. Up to multiplication by powers of i, these prime elements are equal to $1 + i$, $2 \pm i$, and $5 \pm 4i$. To decide, for example, which of the two prime elements $5 \pm 4i$ of norm 41 is a divisor, we can calculate which of the quotients $\frac{29-97i}{5\pm 4i} \in \mathbf{Q}(i)$ is in $\mathbf{Z}[i]$.

Exercise 8. Carry out the calculation.

It is more instructive to view the divisibility of α by the prime elements $5 \pm 4i$ as whether or not α is contained in the principal ideals $(5 \pm 4i)$. Ideals are kernels of homomorphisms, and in the present case, $\mathbf{Z}[i]$ admits the following homomorphisms $\varphi_{1,2} : \mathbf{Z}[i] \to \mathbf{F}_{41}$. Corresponding to the two roots $\pm 9 \bmod 41$ of $-1 \bmod 41$, we have $\varphi_1 : a + bi \mapsto a + 9b \bmod 41$ and $\varphi_2 : a + bi \mapsto a - 9b \bmod 41$ with kernels $\ker \varphi_1 = (5 + 4i)$ and $\ker \varphi_2 = (5 - 4i)$. We have $\varphi_1(\alpha) = (29 + 9 \cdot -97 \bmod 41) = 17 \bmod 41$, so $5 + 4i$ is *not* a divisor of α. Hence $5 - 4i$ is the factor we are looking for. We indeed have $\varphi_2(\alpha) = (29 - 9 \cdot -97 \bmod 41) = 0 \bmod 41$.

Since α is not divisible by 5, only one of the prime elements $2 \pm i$ of norm 5 divides α; it has multiplicity 3. For the homomorphism $\psi_1 : a + bi \mapsto (a + 2b \bmod 5)$ with kernel $(2 - i)$, we have $\psi_1(\alpha) = (29 - 2 \cdot 97 \bmod 5) = 0 \bmod 5$, so $2 - i$ is the factor we want. Since there is only one prime element of norm 2 up to units, we obtain

$$\alpha = 29 - 97\,i = i^k \cdot (1 + i) \cdot (2 - i)^3 \cdot (5 - 4i).$$

To determine the necessary power i^k of i, we do not need to calculate the right-hand side by expanding it. If we apply the homomorphism $\psi_2 : a + bi \mapsto a - 2b \bmod 5$ corresponding to the prime element $2 + i$ that does *not* divide α, then we find that $\psi_2(\alpha) = (29 + 2 \cdot 97 \bmod 5) = 3 \bmod 5$ must equal $(-2)^k \cdot -1 \cdot (-1)^3 \cdot 3 \bmod 5$. It follows that $(-2)^k \equiv 1 \bmod 5$, so $k \equiv 0 \bmod 4$, and the "units contribution" i^k in the factorization above is 1.

Exercise 9. Verify this result by determining k again using $\varphi_1 : \mathbf{Z}[i] \to \mathbf{F}_{41}$ or a self-chosen homomorphism $\mathbf{Z}[i] \to \mathbf{F}_{13}$.

We now apply 12.19 to solving a *Diophantine equation*. By this we mean an algebraic equation with rational coefficients of which we want to find only the rational and

integer solutions, not all real or complex ones. Formulated more geometrically, this corresponds to determining the points with integer coordinates on an algebraic curve.

12.21. Theorem. *The only integer solution of* $X^2 + 1 = Y^3$ *is* $(X, Y) = (0, 1)$.

Proof. Let (x, y) be a solution of the equation, and write

$$x^2 + 1 = (x + i)(x - i) = y^3.$$

Let us prove that $x + i$ and $x - i$ are relatively prime in $\mathbf{Z}[i]$. We first note that x is *even*. After all, for x odd, y must be even, which gives a contradiction modulo 4: $x^2 + 1 \equiv 2 \bmod 4$ and $y^3 \equiv 0 \bmod 4$. A generator d of the ideal $(x + i, x - i) \subset \mathbf{Z}[i]$ now divides both $x + i$ and $x - i$, hence also $(x + i) - (x - i) = 2i$. Since $2i$ divides the even number x, the generator d divides both x and $x + i$, hence also the unit i. We find that d itself is a unit in $\mathbf{Z}[i]$, so $x + i$ and $x - i$ are relatively prime in $\mathbf{Z}[i]$.

Next, we prove that a product of two relatively prime elements in $\mathbf{Z}[i]$ can only be a cube if *each* of these two elements is a cube in $\mathbf{Z}[i]$. Namely, let π be an irreducible factor of $x + i$. Then π does not divide $x - i$, and the number of factors π in $x + i$ is equal to the number of factors π in $(x + i)(x - i) = y^3$, *hence* a multiple of 3. Looking at the decomposition of $x + i$, we see that $x + i$ is the product of a unit $u \in \mathbf{Z}[i]^*$ and a cube. Since $u^4 = 1$, $u = u^{-3}$ itself is also a cube. We conclude that $x + i$ can be written as a cube

$$x + i = (a + bi)^3 = a(a^2 - 3b^2) + (3a^2 - b^2)bi \qquad \text{with } a, b \in \mathbf{Z}.$$

If we compare the imaginary parts, then it follows easily from $(3a^2 - b^2)b = 1$ that $b = \pm 1$ and $3a^2 - 1 = \pm 1$. The only solution is $a = 0$ and $b = -1$, which gives the unique solution $(x, y) = (0, 1)$. □

▶ Prime Ideal Factorization

The example above shows that for finding the integer solutions of an equation with coefficients in $\mathbf{Z}$, it can be useful to work in a *larger* ring than $\mathbf{Z}$ itself, such as $\mathbf{Z}[i]$. It turns out that these larger rings are *not* always principal ideal domains. For example, the subring $\mathbf{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbf{Z}\}$ of $\mathbf{C}$ is an integral domain that somewhat resembles $\mathbf{Z}[i]$ but is *not* a principal ideal domain. In this ring,

$$21 = 3 \cdot 7 = (4 + \sqrt{-5})(4 - \sqrt{-5}) = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$$

has *three* totally different decompositions into irreducible elements (Exercise 61). Apparently, not every irreducible element of $\mathbf{Z}[\sqrt{-5}]$ is prime. Through work of the German Ernst Eduard Kummer (1810–1893) and other founders of *algebraic number theory*,[3] it became clear that these types of number rings do *not* always have unique factorization into prime elements as in 12.11 but *do* have the unique factorization into *prime ideals* as in 12.17. In the given example, there exist prime ideals

$$P_3 = (3, 1 + \sqrt{-5}), \quad Q_3 = (3, 1 - \sqrt{-5}), \quad P_7 = (7, 4 + \sqrt{-5}), \quad Q_7 = (7, 4 - \sqrt{-5})$$

that are *not* principal and for which the ideal (21) can be decomposed as

$$(21) = P_3 \cdot Q_3 \cdot P_7 \cdot Q_7.$$

Every product of two of the four "ideal factors" of (21) is a principal ideal, which "explains" the factorizations into elements mentioned above; see Exercises 61–63.

Exercise 10. Let I_1 and I_2 be ideals in a commutative ring R. Prove the *prime ideal property*

$$P \supset I_1 I_2 \quad \Longrightarrow \quad P \supset I_1 \text{ or } P \supset I_2$$

for prime ideals P in the sense of 12.14.

The word *ideal* introduced by Kummer, which is now a fundamental concept in algebra, was still perceived as rather mysterious around 1850. Kummer himself emphasized the analogy with chemistry, which in a somewhat similar way described the molecules of a substance in terms of constituents that do not occur in isolation, called "atoms."[4]

EXERCISES.

11. Show that every ideal in a product $R = R_1 \times R_2 \times \ldots \times R_n$ of principal ideal domains R_i is a principal ideal. Is R also a principal ideal domain?
12. Prove: the polynomial ring $K[X, Y]$ over a field K is not a principal ideal domain.
13. Is the ring $K[X, X^{-1}]$ of Laurent polynomials over a field K a principal ideal domain?
14. Is the ring $K[[X]]$ of power series over a field K a principal ideal domain?
15. Let R be a ring. Show that the rule $\deg(fg) = \deg(f) + \deg(g)$ holds for all $f, g \in R[X]$ if and only if R has no non-trivial zero divisors.
16. Prove that for non-negative integers a and b, the following statements are equivalent:
    (i) There exist f, $g \in (\mathbf{Z}/15\mathbf{Z})[X]$ with
    $$\deg(f) = a, \qquad \deg(g) = b, \qquad f \cdot g = X^{10}.$$
    (ii) We have $0 \leq a \leq 10$, $0 \leq b \leq 10$, and $a + b \geq 10$.
    [Hint: use the isomorphism $(\mathbf{Z}/15\mathbf{Z})[X] \cong (\mathbf{Z}/3\mathbf{Z})[X] \times (\mathbf{Z}/5\mathbf{Z})[X]$.]
17. For what pairs of non-negative integers (a, b) do there exist f, $g \in (\mathbf{Z}/8\mathbf{Z})[X]$ with
    $$\deg(f) = a, \qquad \deg(g) = b, \qquad f \cdot g = 1?$$
18. Let $\mathbf{H}$ be Hamilton's quaternion algebra. Prove that the polynomial $X^2 + 1 \in \mathbf{H}[X]$ has *infinitely* many different zeros in the division ring $\mathbf{H}$. Why does this not contradict 12.4? [Hint: square $xi + yj + zk$.]
19. Let R be a commutative ring and $f \in R[X]$ be a polynomial with $f \neq 0$. Suppose that $a_1, \ldots, a_n \in R$ are zeros of f with the property $a_i - a_j \in R^*$ for all i, j with $1 \leq i < j \leq n$. Prove: $n \leq \deg f$.
20. Let H be a finite subgroup of the group of units of an integral domain R, and consider the polynomial $f = \prod_{a \in H}(X - a) \in R[X]$. Prove: $f = X^{\#H} - 1$.
21. Let $\phi : R[X] \to \mathrm{Map}(R, R)$ be the map that sends a polynomial $f \in R[X]$ to the corresponding map $r \mapsto f(r)$ in the function ring $\mathrm{Map}(R, R)$. Prove:
    a. The map ϕ is a ring homomorphism if and only if R is commutative.
    b. The map ϕ is injective but not surjective if R is an infinite integral domain.
    c. The map ϕ is surjective but not injective if R is a finite integral domain.
22. The *derivative* of a polynomial $f = \sum_k a_k X^k$ with coefficients in a commutative ring R is the polynomial $f' = \sum_k k a_k X^{k-1}$. If f is of the form $f = (X - a)^2 q$ with $q \in R[X]$ and $a \in R$, then a is called a *double zero* of f.
    a. Prove the differentiation rules $(f + g)' = f' + g'$ and $(fg)' = f'g + fg'$ for $f, g \in R[X]$.
    b. Let $a \in R$ be a zero of f. Prove: a is a double zero of $f \iff f'(a) = 0$.
    c. Check which zeros of $f = X^2 - \overline{1}$ and $f = X^2$ in $\mathbf{Z}/8\mathbf{Z}$ are double.
23. Show that the polynomial $X^7 + 7X + 1 \in \mathbf{C}[X]$ has no double zeros in $\mathbf{C}$.

24. Let R be a commutative ring. We say that R is *connected* if the number of zeros of $X^2 - X$ in R is equal to 2.
    a. Prove: R is connected if and only if R is not the zero ring and there exist no rings $R_1 \neq \{0\}$ and $R_2 \neq \{0\}$ for which there is a ring isomorphism $R \cong R_1 \times R_2$.
    b. Suppose that R is connected and that I, J are ideals with $IJ = \{0\}$ and $I + J = R$. Prove: $\{I, J\} = \{\{0\}, R\}$.

25. Let R be a commutative ring and $f \in R[X]$. We call f *separable* if we have $R[X]f + R[X]f' = R[X]$, with f' the derivative of f.
    a. Suppose that $a \in R$ is a zero of f, and write $f = (X - a) \cdot g$ with $g \in R[X]$. Prove: $g(a) = f'(a)$, and if f is separable, then we have $g(a) \in R^*$.
    b. Suppose that f is separable, let a, $b \in R$ be zeros of f, and write $f = (X - a) \cdot g$ with $g \in R[X]$. Prove: the ideals $I = R \cdot (b - a)$ and $J = R \cdot g(b)$ satisfy $IJ = \{0\}$ and $I + J = R$.
    c. Suppose that R is connected and f is separable. Prove: $f \neq 0$, and the number of zeros of f in R is at most $\deg f$.

26. Let $x_0, x_1, x_2, \ldots, x_n$ be an $(n+1)$-tuple of *distinct* elements of a field K and $y_0, y_1, y_2, \ldots, y_n$ be an $(n+1)$-tuple of arbitrary elements of K. Prove that there is exactly one polynomial $f \in K[X]$ of degree $\deg(f) \leq n$ with $f(x_i) = y_i$ for $i = 0, 1, 2, \ldots, n$ and that it is given by the *Lagrange interpolation formula*:
$$f = \sum_{i=0}^{n} \left(\prod_{j=0, j\neq i}^{n} \frac{X - x_j}{x_i - x_j} \right) y_i.$$

27. Show that there is no polynomial $f \in (\mathbf{Z}/100\mathbf{Z})[X]$ with $f(\overline{1}) = \overline{1}$ and $f(\overline{11}) = \overline{17}$.

28. Let R be a commutative ring and $U \subset R^*$ be a finite subset. Prove that there exists an $f \in R[X]$ with $f(u) = u^{-1}$ for all $u \in U$.

29. Determine primitive roots modulo the primes 11, 31, 41, and 71.

30. Determine the six smallest primes $p > 2$ for which $5 \bmod p$ is a primitive root. What do you notice about the final digits of these primes?[5] *Are there infinitely many primes p with the stated property?[6]

31. Let F be a finite field. A *primitive root* of F is an element $a \in F^*$ with $F^* = \langle a \rangle$. Suppose that the product of all primitive roots of F is *not* equal to 1. Prove that F is isomorphic to $\mathbf{Z}/3\mathbf{Z}$.

*32. For a prime p, we denote by $s(p) \in \mathbf{F}_p$ the sum of all primitive roots of $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$. By calculating several values of $s(p)$ (use a calculator if necessary), find a conjectural formula for $s(p)$, and then prove that it is correct.

33. A commutative ring R is called *Noetherian* if there do not exist any infinite increasing chains
$$I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq I_4 \subsetneq \ldots$$
of ideals in R. Prove that R is Noetherian if and only if every ideal of R is generated by a finite set $S \subset R$.
[Amalie Emmy Noether (1882–1935) is one of the founders of modern algebra.]

34. Prove that every element $x \neq 0$ in a Noetherian integral domain R can be written as a product of a unit and finitely many irreducible elements.

35. On the additive group $R = \mathbf{Z} \times \mathbf{Z}/5\mathbf{Z}$, define multiplication by

$$(a_1, \overline{b}_1) \cdot (a_2, \overline{b}_2) = (a_1 a_2, \overline{a_1 b_2 + a_2 b_1}).$$

Show that this makes R into a commutative ring isomorphic to $\mathbf{Z}[X]/(5X, X^2)$. Prove that R^* is a cyclic group of order 10. Is R an integral domain?

36. Show that the elements $x = (0, \overline{1})$ and $y = (0, \overline{2})$ generate the same ideal in the ring R from the previous exercise but that there is no unit $u \in R^*$ with $y = ux$.

37. Let K be a field. Prove that every rational function $f \in K(X)^*$ can be written uniquely as $f = c \cdot \prod_p p^{n_p}$. Here, $c \in K^*$ is a constant, p runs through all monic irreducible polynomials in $K[X]$, and the exponents $n_p \in \mathbf{Z}$ are integers that are almost all 0.

38. A *Euclidean ring* is an integral domain R endowed with a function $g : R \setminus \{0\} \to \mathbf{Z}_{\geq 0}$ with the following property:

    $(*)$ for $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$ and r satisfies $r = 0$ or $g(r) < g(b)$.

    Prove that every Euclidean ring is a principal ideal domain.

39. Show that $\mathbf{Z}$ and the polynomial ring $K[X]$ over a field K are Euclidean rings and that in every Euclidean ring, an analog of the *Euclidean algorithm* given in 6.13 can be used to calculate GCDs.

40. Prove that the ring $\mathbf{Z}((X))$ is Euclidean.

41. Let K be a field. Prove that the rings $K((X))[Y]$ and $K[Y]((X))$ are both Euclidean, that the first ring is isomorphic to a subring of the second, and that the rings are not isomorphic.

42. Decompose $63 + 75i$ and $217 - 35i$ into factors in $\mathbf{Z}[i]$.

43. Calculate $\mathrm{GCD}(135 - 14i, 155 + 34i)$ both using the Euclidean algorithm and using explicit prime factorization in $\mathbf{Z}[i]$.

44. Show that $\mathbf{Q}(i) = \{a + bi : a, b \in \mathbf{Q}\}$ is the field of fractions of $\mathbf{Z}[i]$ and that there are isomorphisms $\mathbf{Q}(i) \cong \mathbf{Q}[X]/(X^2 + 1)$ and $\mathbf{Z}[i] \cong \mathbf{Z}[X]/(X^2 + 1)$.

45. Let $q \in \mathbf{Q}^*$. Prove: there exists an $\alpha \in \mathbf{Q}(i)$ with $\alpha^2 = q$ if and only if there exists an $r \in \mathbf{Q}$ with $q = r^2$ or $q = -r^2$.

46. Let $q \in \mathbf{Q}^*$. Prove: there exists an $\alpha \in \mathbf{Q}(i)$ with $\alpha^4 = q$ if and only if there exists an $r \in \mathbf{Q}$ with $q = r^4$ or $q = -4 \cdot r^4$.

47. Do there exist three different points P, Q, $R \in \mathbf{Q} \times \mathbf{Q}$ for which the line PQ makes a angle of 60° with the line RQ? Give such points, or prove that they do not exist.

48. Show that for a prime p with $p \equiv 3 \bmod 4$, $\mathbf{Z}[i]/p\mathbf{Z}[i]$ is a field with p^2 elements and that for $p \equiv 1 \bmod 4$, it is a product $\mathbf{F}_p \times \mathbf{F}_p$ of two fields. Which of the rings from Exercise 11.63 do we obtain for $p = 2$?

49. In this exercise, we prove that for every element $\alpha = a + bi \in \mathbf{Z}[i]$ different from 0, the norm $N(\alpha) = a^2 + b^2$ of α is equal to the number of elements of the ring $\mathbf{Z}[i]/\mathbf{Z}[i]\alpha$.

a. Prove: for every $n \in \mathbf{Z} \setminus \{0\}$, the ring $\mathbf{Z}[i]/n\mathbf{Z}[i]$ is finite of order n^2. Deduce that $\mathbf{Z}[i]/\mathbf{Z}[i]\alpha$ is finite for every $\alpha \in \mathbf{Z}[i]$, $\alpha \neq 0$.
b. Define $\mathcal{N} : \mathbf{Z}[i]\backslash\{0\} \to \mathbf{Z}_{>0}$ by $\mathcal{N}(\alpha) = \#(\mathbf{Z}[i]/\mathbf{Z}[i]\alpha)$. Prove: $\mathcal{N}(\alpha) = \mathcal{N}(\bar{\alpha})$ and $\mathcal{N}(\alpha \cdot \beta) = \mathcal{N}(\alpha) \cdot \mathcal{N}(\beta)$ for all α, $\beta \in \mathbf{Z}[i]\backslash\{0\}$.
c. Prove: $\mathcal{N}(\alpha) = N(\alpha)$ for all $\alpha \in \mathbf{Z}[i]\backslash\{0\}$.

50. Determine all integer solutions of the equation $X^2 + 4 = Y^3$.

51. Determine all integer solutions of the equation $X^2 + 1 = Y^5$.

52. A *Pythagorean triple* is a triple (a, b, c) of integers that satisfies the identity

$$a^2 + b^2 = c^2.$$

It is called *primitive* if a, b, and c have no common factors. Prove: for every primitive Pythagorean triple (a, b, c), there exist relatively prime integers m and n such that, after interchanging a and b if necessary, the following identities hold:

$$a = m^2 - n^2, \qquad b = 2mn, \qquad c = \pm(m^2 + n^2).$$

[Hint: write $(a + bi)(a - bi) = c^2$, and deduce $a + bi = (m + ni)^2$ as in 12.21.]

53. Deduce the description of Pythagorean triples from the previous exercise "inside $\mathbf{Z}$" by rewriting the equation for even b as $(\frac{b}{2})^2 = \frac{c+a}{2} \cdot \frac{c-a}{2}$.

54. Prove that for an integer $n > 0$ with prime factorization $\prod_p p^{e(p)}$, the number of pairs $(a, b) \in \mathbf{Z}^2$ with $n = a^2 + b^2$ is equal to

$$r(n) = \begin{cases} 0 & \text{if there is a prime } p \equiv 3 \bmod 4 \text{ with } e(p) \textit{ odd}, \\ 4 \prod_{p \equiv 1 \bmod 4} (e(p) + 1) & \text{otherwise.} \end{cases}$$

*How many pairs are there if we, moreover, require the inequality $a \geq b \geq 0$ to hold?

55. Define $r(n)$ as in the previous exercise. Prove: $\lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} r(n) = \pi$.

56. For the purpose of this exercise, we call an integer $N \in \mathbf{Z}_{>0}$ *normic of length* $2k$ if it has $2k$ decimal digits and the integers a and b described by the first and the second half of its decimal representation satisfy the peculiar identity

$$N = a \cdot 10^k + b = a^2 + b^2.$$

Examples of normic integers of length 4, 6 and 12 are given by

$$1233 = 12^2 + 33^2, \qquad 990100 = 990^2 + 100^2, \quad 123288328768 = 123288^2 + 328768^2.$$

Show that there exist normic integers of length $2k$ if and only if $10^{2k} + 1$ is *not* prime and that the number of such pairs is equal to $[r(10^{2k} + 1) - 8]/4$.
[Hint: look at $(2a - 10^k, 2b - 1)$.]

57. Determine all normic integers of length $2k \leq 20$.
[Hint: Sage can both decompose numbers and do arithmetic modulo n.]

58. Let $\rho = \frac{-1+\sqrt{-3}}{2} \in \mathbf{C}$ be a zero of the polynomial $X^2 + X + 1$. The *Eisenstein integers* are the complex numbers of the form $a + b\rho$ with $a, b \in \mathbf{Z}$. Prove:

a. $\mathbf{Z}[\rho] = \{a + b\rho : a, b \in \mathbf{Z}\}$ is a subring of $\mathbf{C}$, and the norm map $N : \mathbf{Z}[\rho] \to \mathbf{Z}$ given by $a + b\rho \mapsto |a + b\rho|^2 = a^2 - ab + b^2$ is multiplicative.
b. $\mathbf{Z}[\rho]^* = \langle -\rho \rangle$ is cyclic of order 6.
c. $\mathbf{Z}[\rho]$ is a principal ideal domain.

[Ferdinand Gotthold Max Eisenstein[7] (1823–1852) was a German number theorist.]

59. Prove that the primes $p \equiv 2 \bmod 3$ are irreducible in $\mathbf{Z}[\rho]$, that the primes $p \equiv 1 \bmod 3$ decompose as $p = \pi\overline{\pi}$, and that we have $3 = -(2\rho + 1)^2$. Deduce that every prime $p \not\equiv 2 \bmod 3$ can be written as $p = x^2 + 3y^2$ with $x, y \in \mathbf{Z}$ and that this representation is unique up to the signs of x and y.

60. Define $\mathbf{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbf{Z}\} \subset \mathbf{C}$.

a. Show that $\mathbf{Z}[\sqrt{-3}]$ is a subring of index 2 of $\mathbf{Z}[\rho]$.
b. Prove: every ideal $I \subset \mathbf{Z}[\sqrt{-3}]$ is of the form $x\mathbf{Z}[\sqrt{-3}]$ or $x\mathbf{Z}[\rho]$ with $x \in \mathbf{Z}[\sqrt{-3}]$.
c. Is $\mathbf{Z}[\sqrt{-3}]$ a principal ideal domain?

61. Show that $\mathbf{Z}[\sqrt{-5}]$ has group of units $\{\pm 1\}$ and that the elements 3, 7, $4 + \sqrt{-5}$, and $1 + 2\sqrt{-5}$ are irreducible in $\mathbf{Z}[\sqrt{-5}]$.
[Hint: use the norm $N : a + b\sqrt{-5} \mapsto a^2 + 5b^2$.]

62. Show that the map $\mathbf{Z}[\sqrt{-5}] \to \mathbf{Z}/7\mathbf{Z}$ given by $a + b\sqrt{-5} \mapsto (a + 3b \bmod 7)$ is a surjective homomorphism with kernel $P_7 = (7, 4 + \sqrt{-5})$ and that P_7 is not a principal ideal in $\mathbf{Z}[\sqrt{-5}]$.

63. Show that (21) has prime ideal factorization $P_3 \cdot Q_3 \cdot P_7 \cdot Q_7$ in $\mathbf{Z}[\sqrt{-5}]$. Show that none of these prime ideal factors is principal, but that the product of any two of them is.

64. Let $p \neq 5$ be a prime. Prove:

$$p = x^2 + 5y^2 \quad \text{for } x, y \in \mathbf{Z} \quad \Longrightarrow \quad p \equiv 1, 9 \bmod 20.$$

*Does the converse hold?

*65. Let p be an odd prime. Prove:

$$p = x^2 + 2y^2 \quad \text{for } x, y \in \mathbf{Z} \quad \Longleftrightarrow \; p \equiv 1, 3 \bmod 8.$$

[Hint for $\Longleftarrow$: first prove that $\mathbf{Z}[\sqrt{-2}]$ is a principal ideal domain; then show that $\overline{-2}$ is a square in $\mathbf{F}_p^*$ for $p \equiv 1, 3 \bmod 8$. Note that for $p \equiv 1 \bmod 8$ and $x \in \mathbf{F}_p^*$ of order 8, we have $(x + x^3)^2 = \overline{-2}$. For $p \equiv 3 \bmod 8$, such an x of order 8 only exists in $\mathbf{Z}[i]/p\mathbf{Z}[i]$, but we have $x + x^3 \in \mathbf{F}_p \subset \mathbf{Z}[i]/p\mathbf{Z}[i]$.]

66. Determine the decomposition of $5 + i$ and $239 + i$ in $\mathbf{Z}[i]$, and deduce the classic formula

$$\pi = 16 \arctan \frac{1}{5} - 4 \arctan \frac{1}{239} \qquad \text{with} \quad \arctan x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$$

that was used by John Machin (1680–1751) to calculate a hundred decimals of π in 1706.[8]

13 Polynomial Factorization

Polynomial rings in one or more variables with coefficients in integral domains such as $\mathbf{Z}$, $\mathbf{R}$, and $\mathbf{F}_p$ are ubiquitous in mathematics. In the simplest case of the polynomial ring $K[X]$ in one variable X over a field K, we have a principal ideal domain by 12.6, and the theory from §12 applies. This is no longer true when the coefficient ring is not a field or when the number of variables is greater than one. For example, in the ring $\mathbf{Z}[X]$, the ideal $(3, X)$ is not principal (Exercise 11.9), and neither is the ideal (X, Y) in the ring $K[X, Y]$. We will prove that for such rings, unique factorization is still possible in the sense of 12.12.

▶ Unique Factorization Domains

We first formally define the rings "with unique factorization" as those integral domains for which the statement of Theorem 12.12 holds. It will follow from 13.2 and 13.3 that these are not only the principal ideal domains.

13.1. Definition. *A unique factorization domain (UFD) or factorial ring is an integral domain R in which every element $x \neq 0$ can be written as a product*

$$x = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t$$

of a unit $u \in R^*$ and finitely many irreducible elements $p_i \in R$, uniquely up to the order and multiplication by units.

If in a unique factorization domain R, we choose a set $\mathcal{P}$ of irreducible elements with the property that every irreducible element of R is associated with exactly one element of $\mathcal{P}$, then every element $x \neq 0$ of R can be written uniquely as

$$x = u \cdot \prod_{p \in \mathcal{P}} p^{n_p}$$

with $u \in R^*$ and $n_p \in \mathbf{Z}_{\geq 0}$ equal to 0 for almost all $p \in \mathcal{P}$. As in the case $R = \mathbf{Z}$ known from 6.7, the exponent $n_p = \operatorname{ord}_p(x)$ is also called the *order* of x at p. We have the multiplicative property $\operatorname{ord}_p(xy) = \operatorname{ord}_p(x) + \operatorname{ord}_p(y)$. In particular, it follows that the irreducible elements of a unique factorization domain are prime:

$$\operatorname{ord}_p(xy) > 0 \quad \Longrightarrow \quad \operatorname{ord}_p(x) > 0 \ \text{ or } \ \operatorname{ord}_p(y) > 0$$

is an implication that describes the prime property from 12.9 for $p \in \mathcal{P}$.

Exercise 1. Show that ord_p can be extended to a homomorphism $\operatorname{ord}_p : K^* \to \mathbf{Z}$ on the group of units of the field of fractions K of R.

In unique factorization domains, we can define the *greatest common divisor* $\operatorname{GCD}(a, b)$ and the *least common multiple* $\operatorname{LCM}(a, b)$ of elements $a, b \in R \setminus \{0\}$ in terms of their decompositions:

$$\operatorname{GCD}(a, b) = \prod_{p \in \mathcal{P}} p^{\min\{\operatorname{ord}_p(a), \operatorname{ord}_p(b)\}},$$

$$\operatorname{LCM}(a, b) = \prod_{p \in \mathcal{P}} p^{\max\{\operatorname{ord}_p(a), \operatorname{ord}_p(b)\}}.$$

With the convention $\mathrm{ord}_p(0) = \infty$, we can also use this definition if a or b is equal to 0. When $\mathrm{GCD}(a, b) = 1$, the elements a and b are again called *relatively prime*.

Exercise 2. Give analogous definitions for $\mathrm{GCD}(a_1, a_2, \ldots, a_n)$ and $\mathrm{LCM}(a_1, a_2, \ldots, a_n)$.

The definitions of GCD and LCM given above depend on the *choice* of the set $\mathcal{P}$ of representatives of the irreducible elements. For other choices of $\mathcal{P}$, the GCD and LCM are multiplied by a unit. Consequently, in a unique factorization domain, as in the case of principal ideal domains, $g = \mathrm{GCD}(a, b)$ and $k = \mathrm{LCM}(a, b)$ are defined up to multiplication by a unit. For unique factorization domains that are principal ideal domains, the new definition is equivalent to that in 12.18 (Exercise 11). However, in arbitrary unique factorization domains, $(a) + (b)$ is not necessarily a principal ideal, and $d = \mathrm{GCD}(a, b)$ satisfies an inclusion $(d) \supset (a) + (b)$ that need not be an equality. In particular, for relatively prime elements a and b, we do not necessarily have that $1 \in R$ can be written as a linear combination of a and b. For example, the generators of the ideals $(3, X) \subset \mathbf{Z}[X]$ and $(X, Y) \subset K[X, Y]$ mentioned above are relatively prime, but the ideal they generate is *not* the entire ring.

▶ Polynomials over a Unique Factorization Domain

As mentioned, principal ideal domains are examples of unique factorization domains. The following theorem shows that there are many more examples.

13.2. Theorem. *Let R be a unique factorization domain. Then the polynomial ring $R[X]$ over R is also a unique factorization domain.*

By repeatedly applying 13.2, we see that for all $n \geq 1$, the polynomial ring $R[X_1, X_2, \ldots, X_n]$ in n variables over a unique factorization domain R is also a unique factorization domain. In particular, by taking R to be $\mathbf{Z}$ or a field, we have the following useful result.

13.3. Corollary. *Let $n \geq 1$ be an integer and K be a field. Then the polynomial rings $\mathbf{Z}[X_1, X_2, \ldots, X_n]$ and $K[X_1, X_2, \ldots, X_n]$ with coefficients in, respectively, $\mathbf{Z}$ and K are unique factorization domains.* □

The proof of 13.2 given further on uses the fact that $R[X]$ is a subring of $K[X]$, with K the field of fractions of R. We will use the decompositions in the principal ideal domain $K[X]$ to also obtain decompositions in $R[X]$.

So, let R be a unique factorization domain. The first step in the factorization of a non-zero polynomial $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ is "extracting" the factor $a = \mathrm{GCD}(a_0, a_1, \ldots, a_n) \in R$. We then obtain $f = a \cdot f_0 \in R[X]$ for a polynomial $f_0 \in R[X]$ for which no prime element $p \in \mathcal{P}$ divides all coefficients of f_0. Such a polynomial is called a *primitive polynomial* in $R[X]$. Observe that the representation $f = a \cdot f_0$ is unique up to multiplication by units of R.

13.4. Lemma. *The product of two primitive polynomials in $R[X]$ is primitive.*

Proof. Suppose that a prime element $p \in \mathcal{P}$ divides all coefficients of the product fg of two primitive polynomials in $R[X]$. Then in the quotient ring $(R/pR)[X]$, we have

the identity $\overline{f} \cdot \overline{g} = \overline{0}$. However, since p is a prime element, by 12.15 and 12.16, R/pR is an integral domain. Then $(R/pR)[X]$ is also an integral domain, so $\overline{f}$ or $\overline{g}$ is the zero element of $(R/pR)[X]$. But then all coefficients of f or g are divisible by p, in contradiction with the primitivity of f and g. □

Every non-zero polynomial $f \in K[X]$ can be written as $f = c \cdot f_0$ for a primitive polynomial $f_0 \in R[X]$ and a constant $c \in K^*$. After all, by multiplying f by the product b of all denominators of the coefficients of f, we have $bf \in R[X]$. If we now write $bf = a \cdot f_0$ with $a \in R$ and $f_0 \in R[X]$ primitive as above, then we have $f = c \cdot f_0$ for $c = ab^{-1} \in K^*$.

Exercise 3. Show that c and f_0 are uniquely determined up to multiplication by units of R.

Proof of 13.2. Let $f \in R[X]$ be a non-zero polynomial. We first decompose f in $K[X]$ as $f = a \cdot g_1 \cdot g_2 \cdot \ldots \cdot g_t$ with $a \in K^*$ and $g_i \in K[X]$ irreducible. By multiplying all g_i by an element $c_i \in K^*$ if necessary (and then adjusting a), we may assume that each g_i is a primitive polynomial in $R[X]$. This representation is, moreover, unique up to multiplication by units from R^*. Since $g_1 \cdot g_2 \cdot \ldots \cdot g_t \in R[X]$ is primitive by 13.4, we have $a \in R$: it is the GCD of the coefficients of f in R.

Since R is a unique factorization domain, we can decompose the element $a \in R$ as $a = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_s$, with $u \in R^*$ and all p_i irreducible in R. This gives the representation

$$(*) \qquad f = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_s \cdot g_1 \cdot g_2 \cdot \ldots \cdot g_t$$

for f in $R[X]$, which is unique up to the order and multiplication by units from $R^* = R[X]^*$.

We claim that $(*)$ gives the desired factorization of f in $R[X]$. We already have the uniqueness, so it suffices to show that the irreducible elements of $R[X]$ are exactly the prime elements of R and the primitive polynomials in $R[X]$ that are irreducible in $K[X]$. On the one hand, it is clear from the representation $(*)$ of elements of $R[X]$ that every irreducible element of $R[X]$ is either a prime element of R or a primitive polynomial that is irreducible in $K[X]$. On the other hand, it is clear that a representation of such an element as a product of two non-units in $R[X]$ contradicts the uniqueness of the representation given in $(*)$. □

Somewhat sloppily, we can conclude from the proof of 13.2 that decomposing a (primitive) polynomial f in $R[X]$ and in $K[X] = (Q(R))[X]$ amount to the same: every irreducible factor of f in $R[X]$ is also irreducible in $K[X]$. A classic consequence is the following lemma, which Gauss already proved for $R = \mathbf{Z}$.

13.5. Gauss's lemma. *Let R be a unique factorization domain with field of fractions K, and let $f \in R[X]$ be a monic polynomial. Suppose that there exist monic polynomials $g_1, g_2 \in K[X]$ such that we have $f = g_1 g_2$. Then g_1 and g_2 have coefficients in R.*

Proof. Pick $c_i \in K^*$ such that $c_i g_i$ is primitive in $R[X]$. Since c_i is the leading coefficient of $c_i g_i \in R[X]$, we have $c_i \in R$. The polynomial $c_1 c_2 \cdot f = (c_1 g_1) \cdot (c_2 g_2)$ is primitive by 13.4, justt like thhe monic polynomial $f \in R[X]$. Wec deduce that $c_1 c_2 \in R$ is a unit. It follows that both c_1 and c_2 are units of R, so g_1 and g_2 are polynomials in $R[X]$. □

▶ Decomposition in $\mathbf{Z}[X]$

If we apply Gauss's lemma to a linear factor $g_1 = X - q \in \mathbf{Q}[X]$ of a monic polynomial $f \in \mathbf{Z}[X]$, then it follows that every rational zero of a monic polynomial in $\mathbf{Z}[X]$ is integer. More generally, we have the following.

13.6. Lemma. *Let $f = \sum_{i=0}^{n} a_i X^i \in \mathbf{Z}[X]$ be a polynomial of degree $n \geq 1$ and $q \in \mathbf{Q}$ be a zero of f. If we write $q = \frac{b}{c}$ with $b, c \in \mathbf{Z}$ relatively prime, then we have $b | a_0$ and $c | a_n$.*

Proof. If $q = \frac{b}{c}$ is a zero of f, then $cX - b$ is a primitive irreducible polynomial in $\mathbf{Z}[X]$ that divides f in $\mathbf{Q}[X]$, so it occurs (up to sign) in the factorisation of f in $\mathbf{Z}[X]$ as a product of a constant and primitive irreducible polynomials. It follows that c divides the leading coefficient a_n of f and b divides the constant coefficient a_0. □

Exercise 4. What are the rational zeros of $f = 10X^5 + 23X^4 - 30X^3 - 20X^2 + 20X - 3$?

Lemma 13.6 gives a way to determine all rational zeros of a polynomial $f \in \mathbf{Z}[X]$. More generally, the question is how to determine the decomposition of a polynomial $f \in \mathbf{Z}[X]$ of degree $n > 0$ in finitely many steps. To check whether f has a factor $g \in \mathbf{Z}[X]$ of degree at most d, with $1 \leq d \leq n$, we can choose $d + 1$ different values $x_0, x_1, \ldots, x_d \in \mathbf{Z}$ that are not zeros of f. Since f has at most n zeros in $\mathbf{Z}$, this is not difficult; if we accidentally choose a zero, this can, moreover, be used directly to divide f by a linear factor. Now, if g is a divisor of f, then $g(x_i)$ is a divisor of $f(x_i)$ for $i = 0, 1, \ldots, d$. Each number $f(x_i)$ has only finitely many divisors in $\mathbf{Z}$, so we have finitely many possibilities for each of the numbers $g(x_i)$. To every $(d + 1)$-tuple of possible values for $(g(x_i))_{i=0}^{d}$ corresponds a *unique* polynomial $g \in \mathbf{Q}[X]$ of degree at most d that takes these values in the points x_i. *Lagrange's interpolation formula* from Exercise 12.26 gives a formula for it. This gives a finite list of possibilities for g, which includes all divisors of f of degree at most d. We can then use the method of 12.1 to find the divisors of $f \in \mathbf{Z}[X]$.

Exercise 5. Show that for $d \geq n/2$, we find all prime factors of f.

The method above quickly becomes very time-consuming, but it shows that the factorization problem in $\mathbf{Z}[X]$ is solvable in finitely many steps. (See Exercise 40 for another argument.) Because of the simplicity of the underlying idea, it is somewhat reminiscent of the method of *trial division* mentioned in §6 for factoring integers: by simply trying out divisors $d = 2, 3, \ldots$ up to $\sqrt{n}$, we can decompose every number $n > 1$ in $\mathbf{Z}$.

▶ Reduction Modulo Primes

Many factorization techniques in $\mathbf{Z}[X]$ use the reduction map $\mathbf{Z}[X] \to (\mathbf{Z}/n\mathbf{Z})[X]$, where n is a prime or a power of one.

13.7. Theorem. *Let p be a prime and $f = \sum_{i=0}^{n} a_i X^i \in \mathbf{Z}[X]$ be a primitive polynomial whose leading coefficient is not divisible by p. Then we have*

$$(f \bmod p) \text{ is irreducible in } \mathbf{F}_p[X] \quad \Longrightarrow \quad f \text{ is irreducible in } \mathbf{Z}[X].$$

Proof. If f is reducible in $\mathbf{Z}[X]$, then it follows from the primitivity of f that there exist non-constant polynomials $g, h \in \mathbf{Z}[X]$ with $f = g \cdot h$. By taking this identity modulo p, we obtain $\overline{f} = \overline{g} \cdot \overline{h} \in \mathbf{F}_p[X]$. Because of the assumption on the leading coefficient of f, the leading coefficients of g and h also are not divisible by p. It follows that $\overline{g}$ and $\overline{h}$ are not constant in $\mathbf{F}_p[X]$, so $\overline{f} = (f \bmod p)$ is reducible in $\mathbf{F}_p[X]$. Elementary logic shows that the proved implication is equivalent to the statement of the theorem. □

Exercise 6. Show that the condition on the leading coefficient of f in 13.7 cannot be left out.

Exercise 7. Generalize 13.7 to the case of a prime element p of a unique factorization domain R.

Suppose that we want to decompose the primitive polynomial $f = 143X^3 - 8X^2 + X + 105 \in \mathbf{Z}[X]$. Since f has degree 3, it is irreducible in $\mathbf{Z}[X]$ if it has no zeros in $\mathbf{Q}$. Testing all zeros allowed by 13.6 (64 in all!) is a bit of work. However, we note that $(f \bmod 2) = X^3 + X + 1 \in \mathbf{F}_2[X]$ has no zeros in $\mathbf{F}_2$ and therefore is irreducible; it follows that f is irreducible in $\mathbf{Z}[X]$.

Exercise 8. Is $7X^3 - 51X^2 + 5X + 70$ irreducible in $\mathbf{Z}[X]$?

For small primes p, the zeros of a polynomial in $\mathbf{F}_p[X]$ can be found by simply trying all elements of $\mathbf{F}_p$. Something similar is true for low-degree factors. For larger p, we use GCD calculations as in Exercise 22.

Even if $f \in \mathbf{Z}[X]$ is a polynomial for which $(f \bmod p)$ is reducible, the factorization of $(f \bmod p)$ in $\mathbf{F}_p[X]$ gives valuable information on the decomposition of f. For example, modulo the primes 2 and 3, the polynomial $f = X^4 + 13X^3 - 9X^2 - 2X + 65$ has the respective factorizations

$$\begin{aligned}(f \bmod 2) &= X^4 + X^3 + X^2 + 1 = (X \ \ + 1)(X^3 + X + 1) \in \mathbf{F}_2[X],\\ (f \bmod 3) &= X^4 + X^3 + X \ \ - 1 = (X^2 + 1)(X^2 + X - 1) \in \mathbf{F}_3[X].\end{aligned}$$

It follows from the second factorization that f has no linear factors in $\mathbf{Z}[X]$ and from the first that f has no irreducible quadratic factors in $\mathbf{Z}[X]$. We conclude that f is irreducible.

Sometimes, it is possible to prove the irreducibility of f in $\mathbf{Z}[X]$ using a variant of the *proof* of 13.7. As in Exercise 7, the result for polynomials holds over an arbitrary unique factorization domain R.

13.8. Eisenstein's citerion. *Let R be a unique factorization domain and $p \in R$ be a prime element. Let $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ be a primitive polynomial that satisfies*

$$p \nmid a_n, \qquad p | a_i \text{ for } i = 0, 1, \ldots, n-1, \qquad \text{and} \qquad p^2 \nmid a_0.$$

Then f is irreducible in $R[X]$.

Proof. If f is reducible in $R[X]$, then it again follows from the primitivity of f that there exist non-constant polynomials $g, h \in R[X]$ with $f = g \cdot h$. Modulo p, we obtain $\overline{g} \cdot \overline{h} = \overline{f} = \overline{a}_n X^n \in K[X]$, with K the field of fractions of the residue class ring R/pR. This gives $\overline{g} = \overline{c}_g X^k$ and $\overline{h} = \overline{c}_h X^{n-k}$ for the leading coefficients c_g and c_h of g and h in R and some $k \in \{0, 1, 2, \ldots, n\}$. By the assumption on the leading coefficient of f, the polynomials $\overline{g}$ and $\overline{h}$ have positive degree, so k is not equal to 0 or n. It follows that the constant coefficients of both g and h are divisible by p. The constant coefficient of f, which is a product of these two, is therefore divisible by p^2, in contradiction with the assumption. □

A polynomial that satisfies the conditions of the criterion is called an *Eisenstein polynomial at p* in $R[X]$.

13.9. Examples. 1. Let $n \geq 1$ be arbitrary. The polynomial $X^n - 2$ is Eisenstein at $p = 2$ for all $n \geq 1$ and therefore irreducible in $\mathbf{Z}[X]$. Likewise, for square-free $a \neq \pm 1$, $X^n - a$ is Eisenstein at every prime divisor of a and therefore irreducible in $\mathbf{Z}[X]$.

2. Let p be a prime. The *p-th cyclotomic polynomial* is the polynomial

$$\Phi_p(X) = \frac{X^p - 1}{X - 1} = X^{p-1} + X^{p-2} + X^{p-3} + \ldots + X + 1 \in \mathbf{Z}[X].$$

The complex zeros of Φ_p are the p-th roots of unity in $\mathbf{C}^*$ different from 1. To see that this polynomial is irreducible in $\mathbf{Z}[X]$, we apply an automorphism of the polynomial ring $\mathbf{Z}[X]$ induced by $X \mapsto X + 1$. Under this "shifting of polynomials," Φ_p is mapped to the polynomial

$$\Phi_p(X+1) = \frac{(X+1)^p - 1}{X} = X^{p-1} + pX^{p-2} + \binom{p}{2} X^{p-3} + \ldots + \binom{p}{p-2} X + p.$$

Since all binomial coefficients $\binom{p}{i}$ with $1 < i < p$ are divisible by p, this is an Eisenstein polynomial at p. Since $\Phi_p(X+1)$ is irreducible in $\mathbf{Z}[X]$, so is Φ_p; see Exercises 26 and 27.

3. Let $f \in \mathbf{Z}[X, Y]$ be the polynomial given by

$$f = X^3 + Y^3 + X^2Y + XY^2 + X^2 + Y^2 - Y.$$

View f as a cubic polynomial in Y with coefficients in $\mathbf{Z}[X]$, and write

$$f = Y^3 + (X+1)Y^2 + (X^2 - 1)Y + (X^3 + X^2).$$

This is a primitive polynomial in Y, and except for the leading coefficient, all coefficients are divisible by the prime element $X + 1 \in \mathbf{Z}[X]$. The constant coefficient $X^2(X+1)$ only has a single factor $X + 1$, so f is an Eisenstein polynomial in Y at $X + 1$. It follows that f is irreducible in $\mathbf{Z}[X, Y]$.

Exercise 9. Show that $X^3 + Y^3 + X^2Y + XY^2 + Y^2 - Y$ is irreducible in $\mathbf{Z}[X, Y]$.

▶ Numerical Methods

Modern computer algebra packages such as Maple, Mathematica, and Pari have standard routines for the decomposition of polynomials in one or more variables. This makes the modern mathematician less dependent on all sorts of "tricks" such as those in 13.9. On the other hand, such packages create a strong need for *efficient* methods for polynomial factorization, and this requires more than "smart programming." As for the factorization of numbers discussed in §6, practical methods rely on somewhat advanced mathematics; this makes what is known as *algorithmic algebra* an active field of research.[9]

For a polynomial $f \in \mathbf{Z}[X]$ in one variable, we first need an *upper bound* on the size of the coefficients of possible divisors $g \in \mathbf{Z}[X]$ of f. Such upper bounds are derived using the fundamental fact that a polynomial of degree n with complex coefficients has exactly n complex zeros. This theorem, known as the *fundamental theorem of algebra*, is proved in 26.3. The complex zeros of a divisor $g|f$ in $\mathbf{Z}[X]$ form a subset of these zeros, and by bounding the absolute values of the complex zeros of f (and observing that the leading coefficient of g divides that of f), we obtain an upper bound B on the absolute values of the coefficients of g (Exercise 40).

Then, we calculate the factorization of f modulo a *prime power* $p^k > 2B$. This is done by first factoring f modulo p—there exist reasonably fast methods for this—and then determining the factorization modulo increasingly high prime powers using a Newton-like iteration process. Admittedly, $\mathbf{Z}/p^k\mathbf{Z}$ is not a unique factorization domain, but it turns out that for k sufficiently large, this does not matter.[10] For every factor $(g \bmod p^k)$ obtained in this way, there is a unique polynomial $g \in \mathbf{Z}[X]$ with this reduction and coefficients no greater than B. We then test whether this is a factor of f. This algorithm, known as the Berlekamp algorithm, works well if $(g \bmod p^k)$ does not have too many factors. In the unfortunate case that it does, we can successfully resort to techniques developed in the 1980s that rely on methods for finding short vectors in lattices in $\mathbf{R}^n$. These techniques can also be applied to the factorization of polynomials in multiple variables.

The factorization of polynomials in $\mathbf{R}[X]$ or $\mathbf{C}[X]$ is part of the field of *numerical analysis*. Since real and complex numbers can only be represented with finite precision, unlike in the case of $\mathbf{Q}[X]$, zeros and factors can only be *approximated*. There are several methods, of which in particular *Newton iteration* is widely used. Again, numerous algorithmic refinements are possible.

EXERCISES.

10. Prove that in a unique factorization domain R, the element $\mathrm{LCM}(a, b)$ is a generator of $(a) \cap (b)$.

11. Let R be an integral domain. We call $d \in R$ a *greatest common divisor* of $a, b \in R$ if

    (i) $(a) + (b) \subset (d)$ and
    (ii) for all $x \in R$, we have $(a) + (b) \subset (x) \implies (d) \subset (x)$.

    a. Prove: a and b have a GCD if and only if the intersection of all principal ideals containing $(a) + (b)$ is a principal ideal.
    b. Prove: if R is a principal ideal or unique factorization domain, then this definition is equivalent to the previous definitions.

12. Show that the GCD of $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ in $\mathbf{Z}[\sqrt{-5}]$ is 1. Also show that 6 and $3 + 3\sqrt{-5}$ do *not* have a GCD in the sense of the previous exercise.

13. Determine the GCD of $X^4 + 2X$ and $X^2 + 5$ in $\mathbf{C}[X]$ and in $\mathbf{F}_3[X]$.

14. Determine the GCD of $X^{12} - 1$ and $X^4 + X$ in $\mathbf{C}[X]$, $\mathbf{R}[X]$, $\mathbf{F}_3[X]$, and $\mathbf{F}_2[X]$.

15. Prove: if R is a unique factorization domain, then the polynomial ring in countably many variables $\Omega = \bigcup_{n>0} R[X_1, X_2, \ldots, X_n]$ over R is also a unique factorization domain. Is Ω a principal ideal domain? Is Ω Noetherian?

16. Is the ring $R[X, X^{-1}]$ of Laurent polynomials over a unique factorization domain R also a unique factorization domain?

17. A *trigonometric polynomial* is a function $f : \mathbf{R} \to \mathbf{R}$ of the form

$$f(x) = a_0 + \textstyle\sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

    with $n \in \mathbf{Z}_{\geq 0}$ and $a_k, b_k \in \mathbf{R}$. If $(a_n, b_n) \neq (0, 0)$, then n is called the *degree* $\deg(f)$ of f.

    a. Show that the set $\mathcal{T}$ of trigonometric polynomials forms a subring of the ring of real-valued functions on $\mathbf{R}$.
    [Hint: write $f(x) = \sum_{k=0}^{n} (c_k e^{ikx} + \bar{c}_k e^{-ikx})$ with $c_k \in \mathbf{C}$.]
    b. Show that the degree on $\mathcal{T}$ satisfies $\deg(fg) = \deg(f) + \deg(g)$. Conclude that $\mathcal{T}$ is an integral domain.
    c. Show that the elements $\sin x$, $1 + \cos x$, and $1 - \cos x$ are irreducible in $\mathcal{T}$ but not prime. Conclude that $\mathcal{T}$ is not a unique factorization domain.
    [Hint: $\sin^2 x = 1 - \cos^2 x$.]

18. Show that the set of maps $f : \mathbf{C} \to \mathbf{C}$ of the form

$$f(z) = a_0 + \textstyle\sum_{k=1}^{n} (a_k \cos kz + b_k \sin kz)$$

    with $n \in \mathbf{Z}_{\geq 0}$ and $a_k, b_k \in \mathbf{C}$ forms a subring of the ring of complex-valued functions on $\mathbf{C}$ and that this ring is a principal ideal domain.

19. Let K be a field and $f \in K[X]$ be a polynomial of degree n. Show that there exist polynomials $f_0, f_1, \ldots, f_n \in K[X]$ with $f(X + Y) = \sum_{k=0}^{n} f_k Y^k$. Prove: $f_0 = f$, and f_1 is the *derivative* of f from Exercise 12.22.

20. Let $f \in \mathbf{F}_p[X]$ be a polynomial with derivative $f' = 0$. Prove: $f(X) = g(X^p)$ for some $g \in \mathbf{F}_p[X]$.

21. Show that the natural map $\mathbf{F}_p[X] \to \mathrm{Map}(\mathbf{F}_p, \mathbf{F}_p)$ that views a polynomial as a function on $\mathbf{F}_p$ induces a ring isomorphism $\mathbf{F}_p[X]/(X^p - X) \xrightarrow{\sim} \mathrm{Map}(\mathbf{F}_p, \mathbf{F}_p)$.

22. Let $f \in \mathbf{F}_p[X]$ be a non-zero polynomial and $S \subset \mathbf{F}_p$ be the set of zeros of f in $\mathbf{F}_p$. Prove: $\mathrm{GCD}(f, X^p - X) = \prod_{x \in S}(X - x)$.

23. Let $f \in \mathbf{Z}[X]$ be a monic polynomial with $f(4) = 17$. Prove that f has at most three rational zeros.

24. Let $f \in \mathbf{Z}[X]$ be an irreducible polynomial.
    a. Prove that f has no double zeros in $\mathbf{C}$.
    b. Prove that there are only finitely many primes p for which $f \bmod p$ has multiple prime factors in $\mathbf{F}_p[X]$. [A prime factor $q|f$ is said to be *multiple* if we have $q^2|f$.]

25. Define the *reciprocal polynomial* of a polynomial $f = \sum_{i=0}^n a_i X^i \in \mathbf{Q}[X]$ with $a_n a_0 \neq 0$ as $f^* = \sum_{i=0}^n a_i X^{n-i}$. Prove: f is irreducible $\iff f^*$ is irreducible.

26. Let K be a field and $\sigma : K[X] \xrightarrow{\sim} K[X]$ be an automorphism of the polynomial ring that is the identity on K. Prove: $\sigma(X) = aX + b$ for some $a \in K^*$ and $b \in K$. Conclude that the group $\mathrm{Aut}_K(K[X])$ of automorphisms of $K[X]$ that are the identity on K is isomorphic to the affine group $\mathrm{Aff}(K) \cong K \rtimes K^*$ over K from 8.14.1 and 8.14.4.

27. Let R be a unique factorization domain and $\sigma \in \mathrm{Aut}(R)$ be an automorphism. Prove: r is irreducible in $R \iff \sigma(r)$ is irreducible in R.

28. Decompose the following polynomials in $\mathbf{Z}[X]$ and in $\mathbf{Q}[X]$:

$$4X^2 + 8, \qquad 3X^4 + 6X + 6, \qquad X^4 - 7X^2 + 5X - 3, \qquad X^3 + 2X + 3.$$

29. Decompose the following polynomials in $\mathbf{Z}[X]$:

$$3X^{12} + 9X^4 + 7, \qquad X^4 + 3X^3 + 2X^2 + 8X + 6,$$
$$(X+1)^7 - X^7 - 1, \qquad X^4 + 2X^3 + 4X^2 + 8X + 16.$$

30. Decompose the following polynomials in $\mathbf{Z}[X]$:

$$X^{120} - 5X^{65} + 3X^{55} - 15, \qquad X^6 - X^2 + 20X - 100, \qquad X^4 - 4X^3 + 4X^2 - 25.$$

31. Decompose the following polynomials in $\mathbf{Q}[X, Y]$ and in $\mathbf{C}[X, Y]$:

$$Y^5 + X^2 - 2, \qquad X^{12} + Y^3 + Y, \qquad X^4 + 4Y^4, \qquad Y^3 - (X+2)Y^2 + Y + X(X+1).$$

32. Let p be a prime. Show that the p^k-th cyclotomic polynomial

$$\Phi_{p^k}(X) = \frac{X^{p^k} - 1}{X^{p^{k-1}} - 1} = X^{(p-1)p^{k-1}} + X^{(p-2)p^{k-1}} + X^{(p-3)p^{k-1}} + \ldots + X^{p^{k-1}} + 1 \in \mathbf{Z}[X]$$

    is irreducible in $\mathbf{Z}[X]$ for all $k \in \mathbf{Z}_{\geq 1}$.

33. Determine the factorization of $f = X^4 + 1$ in $\mathbf{F}_p[X]$ for all primes $p \leq 41$. For which $p \in \mathbf{Z}$ does f split into linear factors in $\mathbf{F}_p[X]$?

34. Show that X^4+1 is reducible in $\mathbf{F}_p[X]$ for all p but irreducible in $\mathbf{Z}[X]$.
[Hint: at least one of the elements $-1, 2, -2$ is a square in $\mathbf{F}_p^*$.]

35. Show that a polynomial $f = X^2 + aX + b \in \mathbf{Z}[X]$ is irreducible modulo a prime $p > 2$ if and only if $a^2 - 4b$ is not a square modulo p. *Suppose that f is reducible modulo all primes p. Is f reducible in $\mathbf{Z}[X]$?

36. Take $f = X^2 - X + 41 \in \mathbf{Z}[X]$. Determine the smallest positive integer x for which $f(x)$ is *not* a prime. Show that there exists a non-constant polynomial $f \in \mathbf{Q}[X]$ with the property that $f(x)$ is prime for all positive integers $x < 1000$. *Does such a polynomial also exist in $\mathbf{Z}[X]$?

37. Let $f \in \mathbf{Z}[X]$ be a polynomial with the property that $f(x)$ is a prime for *all* $x \in \mathbf{Z}$. Is f necessarily a constant polynomial?

*38. Let $f \in \mathbf{Z}[X]$ be a polynomial with the property that $f(x)$ is a square for all $x \in \mathbf{Z}$. Is f necessarily the square of a polynomial in $\mathbf{Z}[X]$?

39. Let $V = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ be the unit circle in $\mathbf{R}^2$ and $\phi : \mathbf{R}[X, Y] \to \mathrm{Cont}(V, \mathbf{R})$ be the map that views polynomials as continuous functions on V. Prove: ϕ is a homomorphism with kernel $(X^2 + Y^2 - 1)$. Is $\mathrm{im}[\phi]$ an integral domain? Is ϕ surjective?
[Hint: write $F \in \mathbf{R}[X, Y]$ as $f(X) + Yg(X) + (X^2 + Y^2 - 1)h(X, Y)$.]

40. Let $f \in \mathbf{Z}[X]$ be a monic polynomial of degree n, and suppose that the absolute values of the coefficients of f are bounded by A.
    a. Prove: for every zero $\alpha \in \mathbf{C}$ of f, we have $|\alpha| \leq nA$.
    b. Prove: every monic divisor $\sum_{j=0}^{d} b_j X^j$ of f in $\mathbf{C}[X]$ of degree d satisfies $|b_j| \leq \binom{d}{j} n^{d-j} A^{d-j}$.
    c. Deduce from (b) that the factorization of f in $\mathbf{Z}[X]$ can be found in finitely many steps.

14 Symmetric Polynomials

In this section, we consider a special type of polynomials in several variables, called *symmetric polynomials*. This is a very classic algebraic subject, which we will encounter later in great generality in Galois theory.

In 12.3, we saw that for an integral domain R, a non-zero polynomial $f \in R[X]$ has no more than $n = \deg(f)$ zeros in R. For $R = \mathbf{Z}$, the *fundamental theorem of algebra* 26.3 tells us that f has *exactly* $n = \deg(f)$ zeros, provided that we count them with multiplicity and are willing to consider zeros in a suitable ring larger than $\mathbf{Z}$ itself, for example $\mathbf{C}$. This is also true for an arbitrary integral domain, and in §21, we will construct the necessary "extension fields."

Even if the zeros of $f \in R[X]$ can only be found in a larger ring $R' \supset R$, it turns out that "symmetric expressions" in the zeros of f lie in R itself. We will see how to calculate them *without* ever venturing outside the ground ring R. The methods we give are used by all computer algebra packages.

▶ General Polynomial of Degree n

We define a "general" polynomial of degree n by working in the ring $R = \mathbf{Z}[T_1, T_2, \ldots, T_n]$ of polynomials in the n variables $T_1, T_2, \ldots, T_n$. We can also allow other rings than $\mathbf{Z}$ for the ring of coefficients, but for the sake of simplicity, we will not do so. The *(total) degree* of the monomial $T_1^{e_1} T_2^{e_2} \ldots T_n^{e_n}$ is equal to $e_1 + e_2 + e_3 + \ldots + e_n$, and we define the *degree* $\deg(f)$ of a non-zero polynomial $f \in R$ as the maximum of the degrees of the monomials in f. If all monomials in f are of the same degree d, then f is called *homogeneous* of degree d. An arbitrary polynomial $f \in R$ of degree d can be written as $f = f_0 + f_1 + f_2 + \ldots + f_d$ with f_k homogeneous of degree k by combining the monomials of equal degree.

The *general polynomial* F_n of degree n is the monic polynomial in $R[X]$ with zeros the variables $T_1, T_2, \ldots, T_n$:

$$F_n = (X - T_1)(X - T_2) \ldots (X - T_n) = X^n + \sum_{k=1}^{n} (-1)^k s_k X^{n-k} \in R[X].$$

The coefficients $s_k \in R$ are called the *elementary symmetric polynomials* in the zeros T_i of f, and the Frenchman François Viète (1540–1603) already knew that for $k = 1, 2, \ldots n$, the polynomial s_k is equal to

$$s_k = \sum_{1 \le i_1 < i_2 < \ldots i_k \le n} T_{i_1} T_{i_2} \ldots T_{i_k},$$

the sum of all products of exactly k different zeros of F. The polynomial $s_k \in R$ is homogeneous of degree k, and we have

$$\begin{aligned} s_1 &= T_1 + T_2 + \ldots + T_n, \\ s_2 &= T_1T_2 + T_1T_3 + \ldots + T_1T_n + T_2T_3 + T_2T_4 + \ldots + T_{n-1}T_n, \end{aligned}$$

and

$$s_n = T_1T_2T_3\dots T_n.$$

Note that, by definition, the general polynomial of degree n has coefficients in the integral domain $R_0 = \mathbf{Z}[s_1, s_2, \dots, s_n]$ and that its n zeros are the variables of the extension ring $R = \mathbf{Z}[T_1, T_2, \dots, T_n] \supset R_0$.

▶ Symmetric Polynomials

A polynomial in $R = \mathbf{Z}[T_1, T_2, \dots, T_n]$ (or, more generally, in $A[T_1, T_2, \dots, T_n]$ for a commutative ring A) is called *symmetric* (in the variables T_i) if it is invariant under all permutation of the variables T_i. Somewhat more formally, we can consider the natural action of the symmetric group S_n on R given by

$$(\sigma f)(T_1, T_2, \dots, T_n) = f(T_{\sigma(1)}, T_{\sigma(2)}, \dots, T_{\sigma(n)}) \qquad \text{for } f \in R, \sigma \in S_n.$$

The symmetric polynomials in R are then the polynomials in R that are invariant under the action of S_n. For every $\sigma \in S_n$, the map $f \mapsto \sigma f$ is an *automorphism* of R, so that we have an inclusion $S_n \subset \mathrm{Aut}(R)$. It follows easily that the symmetric polynomials form a *subring* $R_0 \subset R$.

Exercise 1. Verify this.

Since the kth elementary symmetric polynomial $s_k \in R$ is symmetric, we have the inclusion $\mathbf{Z}[s_1, s_2, \dots, s_n] \subset R_0$. The fundamental theorem for symmetric polynomials says that this inclusion is an equality: *every* symmetric polynomial is a polynomial in the elementary symmetric polynomials. This theme is elaborated further in Galois theory: polynomials with "many" symmetries are contained in "small" subrings.

14.1. Fundamental theorem. *Let $P \in R = \mathbf{Z}[T_1, T_2, \dots, T_n]$ be a symmetric polynomial. Then there is a unique way to write P as a polynomial in the elementary symmetric polynomials s_k.*

We give a constructive proof, that is, a proof that gives an *algorithm* to write a P as an element of $\mathbf{Z}[s_1, s_2, \dots, s_n]$.

Proof. We order the monomials in P *lexicographically*, as in a dictionary. In other words, monomials $T_1^{e_1}T_2^{e_2}\dots T_n^{e_n}$ with the highest exponent e_1 are in front, if two monomials have the same e_1, their order is determined by e_2, and so on.

Now, let $c \cdot T_1^{e_1}T_2^{e_2}\dots T_n^{e_n}$ with $c \in \mathbf{Z} \setminus \{0\}$ be the first term in P, lexicographically, and let $d = e_1 + e_2 + e_3 + \dots + e_n$ be its degree. Then we have $e_1 \geq e_2 \geq e_3 \geq \dots \geq e_n$. After all, if this is not the case, then through a suitable permutation of the T_i, we can transform this term into one that comes earlier lexicographically and that must also be in P because of the symmetry of P; this gives a contradiction. Now, form the monomial

$$\Sigma = s_1^{e_1-e_2}s_2^{e_2-e_3}s_3^{e_3-e_4}\dots s_{n-1}^{e_{n-1}-e_n}s_n^{e_n} \in R$$

of degree

$$\begin{aligned} &e_1 - e_2 + 2(e_2 - e_3) + 3(e_3 - e_4) + \ldots + (n-1)(e_{n-1} - e_n) + ne_n \\ &= e_1 + e_2 + e_3 + \ldots + e_n = d \leq \deg(P), \end{aligned}$$

with first term $T_1^{e_1} T_2^{e_2} \ldots T_n^{e_n}$ (lexicographically), and consider $P_1 = P - c\Sigma \in R$. Because $\deg(\Sigma) \leq \deg(P)$, we have $\deg(P_1) \leq \deg(P)$, and all monomials in P_1 come *later*, lexicographically, than $T_1^{e_1} T_2^{e_2} \ldots T_n^{e_n}$. Since only finitely many different monomials are possible when the degree is bounded, we see that by repeatedly subtracting an element of $\mathbf{Z}[s_1, s_2, \ldots, s_n]$, we can reduce the polynomial P to 0. In other words, P is itself contained in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$.

To prove that there are not two ways to write a polynomial in $\mathbf{Z}[T_1, \ldots, T_n]$ as a polynomial in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$, we must show that there is no non-zero polynomial $g \in \mathbf{Z}[X_1, X_2, \ldots, X_n]$ with $g(s_1, s_2, \ldots, s_n) = 0$. To do this, suppose that such a polynomial g does exist. Write every monomial in g in the form

$$cX_1^{e_1 - e_2} X_2^{e_2 - e_3} X_3^{e_3 - e_4} \ldots X_{n-1}^{e_{n-1} - e_n} X_n^{e_n},$$

and consider the monomial M in g for which the corresponding n-tuple $(e_1, e_2, \ldots, e_n)$ comes first lexicographically. When expanding $g(s_1, s_2, \ldots, s_n)$ as a polynomial in $\mathbf{Z}[T_1, T_2, \ldots, T_n]$, we see that M leads to a term $cT_1^{e_1} T_2^{e_2} \ldots T_n^{e_n}$ that does not disappear, so we have a contradiction. □

It follows from the uniqueness of representations in terms of the elementary symmetric polynomials s_k that $\mathbf{Z}[s_1, s_2, \ldots, s_n]$ can be viewed as a polynomial ring in the variables s_k: the elementary symmetric polynomials are *algebraically independent*.

We say that a monomial $s_1^{a_1} s_2^{a_2} \ldots s_n^{a_n}$ has *weight* $a_1 + 2a_2 + 3a_3 + \ldots + na_n$; more generally, the weight of $g \in \mathbf{Z}[s_1, s_2, \ldots, s_n]$ is the maximum of the weights of the monomials in g. If all monomials in g have the same weight d, then g is said to be *isobaric* of weight d. Note that the weight of $g \in \mathbf{Z}[s_1, s_2, \ldots, s_n]$ is nothing but the *degree* of g as an element of $R = \mathbf{Z}[T_1, T_2, \ldots, T_n]$. The proof of 14.1 shows the following.

14.2. Corollary. *A homogeneous symmetric polynomial in $\mathbf{Z}[T_1, T_2, \ldots, T_n]$ of degree d can be written in a unique way as an isobaric polynomial of weight d in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$.* □

An arbitrary symmetric polynomial P can be written as a sum of homogeneous polynomials P_k of degree k; the polynomials P_k are symmetric because the action of S_n on $\mathbf{Z}[T_1, T_2, \ldots, T_n]$ leaves the degree invariant. By 14.2, the polynomial P_k can be written as an isobaric polynomial of weight k in the elementary symmetric polynomials s_k.

▶ Calculations with Symmetric Polynomials

There is a shortened notation for *symmetric polynomials* P, defined as follows: for every S_n-orbit in P, we write a single representative preceded by the symbol $\sum_n$ to

indicate that we take the sum over the monomials in the S_n-orbit of the representative. In this notation, the kth elementary symmetric polynomial $s_k \in R$ is equal to

$$s_k = \sum_n T_1 T_2 T_3 \dots T_k.$$

More generally, $\sum_n f$ with $f \in \mathbf{Z}[T_1, T_2, \dots, T_n]$ is the notation for the sum of the polynomials in the S_n-orbit of f. Examples:

$$\sum_3 T_1^2 T_2 = T_1^2 T_2 + T_1^2 T_3 + T_2^2 T_3 + T_1 T_2^2 + T_1 T_3^2 + T_2 T_3^2$$

$$\sum_4 T_1 T_2 T_3 = T_1 T_2 T_3 + T_1 T_2 T_4 + T_1 T_3 T_4 + T_2 T_3 T_4$$

$$\sum_4 T_1 T_2 = T_1 T_2 + T_1 T_3 + T_1 T_4 + T_2 T_3 + T_2 T_4 + T_3 T_4$$

If we want to use the method from the proof of 14.1 to write a given symmetric polynomial as a polynomial in the s_k, then the short notation is often useful. After all, if a monomial $r T_1^{e_1} T_2^{e_2} \dots T_n^{e_n}$ occurs in a symmetric polynomial $f \in \mathbf{Z}[T_1, T_2, \dots, T_n]$, then $\sum_n r T_1^{e_1} T_2^{e_2} \dots T_n^{e_n}$ also occurs in that polynomial.

14.3. Examples. 1. Take $n \geq 2$ and $P = T_1^2 + T_2^2 + \dots + T_n^2 = \sum_n T_1^2$. Then T_1^2 is lexicographically the highest term in P, so we form

$$s_1^2 = (T_1 + T_2 + \dots + T_n)^2 = \sum_n T_1^2 + 2 \sum_n T_1 T_2$$

and calculate $P_1 = P - s_1^2 = -2 \sum_n T_1 T_2 = -2 s_2$. In this case, we are done after one step, and we find $P = s_1^2 - 2 s_2$. Note that P is homogeneous of degree 2 and $s_1^2 - 2 s_2$ is isobaric of weight 2.

2. Now, take $n \geq 3$ and $P = T_1^3 + T_2^3 + \dots + T_n^3 = \sum_n T_1^3$. Then T_1^3 is lexicographically the highest term in P, so we form

$$s_1^3 = (T_1 + T_2 + \dots + T_n)^3 = \sum_n T_1^3 + 3 \sum_n T_1^2 T_2 + 6 \sum_n T_1 T_2 T_3.$$

The coefficients 3 and 6 indicate how often a term occurs when we expand s_1^3. We find $P_1 = P - s_1^3 = -3 \sum_n T_1^2 T_2 - 6 \sum_n T_1 T_2 T_3$. The lexicographically highest term in P_1 is $-3 T_1^2 T_2$, so we subtract -3 times

$$s_1 s_2 = \sum_n T_1 \cdot \sum_n T_1 T_2 = \sum_n T_1^2 T_2 + 3 \sum_n T_1 T_2 T_3$$

and obtain

$$P_2 = P_1 + 3 s_1 s_2 = P - s_1^3 + 3 s_1 s_2 = 3 \sum_n T_1 T_2 T_3 = 3 s_3.$$

Conclusion: $P = T_1^3 + T_2^3 + \dots + T_n^3 = s_1^3 - 3 s_1 s_2 + 3 s_3$. Again, note that P is homogeneous of degree 3 and $s_1^3 - 3 s_1 s_2 + 3 s_3$ is isobaric of weight 3.

Exercise 2. What happens in the cases $n < 2$ (in Example 1) and $n < 3$ (in Example 2)?

See Exercise 24 for the representation of the *sum of powers* $\sigma_k = T_1^k + T_2^k + \dots + T_n^k$ in terms of the elementary symmetric polynomials using *Newton's identities*.

▶ Discriminant

A common symmetric polynomial in $\mathbf{Z}[T_1, T_2, \ldots, T_n]$ is the *discriminant*

$$\Delta_n = \prod_{1 \le i < j \le n} (T_i - T_j)^2 \tag{14.4}$$

of the general polynomial F_n of degree n. The polynomial Δ_n is homogeneous of degree $n(n-1)$, so by 14.2, it is a universal isobaric expression of weight $n(n-1)$ in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$. In other words, the discriminant of the general polynomial F_n of degree n is a polynomial in the coefficients $(-1)^{n-k}s_k$ of F_n.

After the uninteresting case $\Delta_1 = 1$, we have

$$\Delta_2 = (T_1 - T_2)^2 = (T_1 + T_2)^2 - 4T_1T_2 = s_1^2 - 4s_2,$$

a result that is also written as $\Delta(X^2 + AX + B) = A^2 - 4B$. Using the method of 14.1, we can express Δ_n in the elementary symmetric polynomials, and with some effort, we obtain

$$\Delta_3 = s_1^2s_2^2 - 4s_2^3 - 4s_1^3s_3 - 27s_3^2 + 18s_1s_2s_3$$
$$\Delta_4 = \frac{1}{27}\left(4(s_2^2 - 3s_1s_3 + 12s_4)^3 - (2s_2^3 - 72s_2s_4 + 27s_1^2s_4 - 9s_1s_2s_3 + 27s_3^2)^2\right),$$

formulas that are too unpleasant to remember. For large values of n, the formulas for the Δ_n are, moreover, too unpleasant to write down.

Exercise 3. Verify that Δ_4 is in $\mathbf{Z}[s_1, s_2, s_3, s_4]$.

Next, let A be an arbitrary integral domain and $f \in A[X]$ a monic polynomial of degree n. Then by 12.3, the polynomial f has at most n zeros in A; in 21.13, we will prove that the number of zeros of f in a sufficiently large integral domain $A' \supset A$ is *exactly* n is the sense that

$$f = \prod_{i=1}^{n}(X - \alpha_i)$$

holds with $\alpha_i \in A'$. If A is a subring of $\mathbf{C}$ such as $\mathbf{Z}$, $\mathbf{Q}$, or $\mathbf{R}$, then by the fundamental theorem of algebra 26.3, we can always take $A' = \mathbf{C}$. The discriminant of f is now defined as

$$\Delta(f) = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2. \tag{14.5}$$

Since A' is an integral domain, we have $\Delta(f) = 0$ if and only if f has a double zero in A'.

The homomorphism $\mathbf{Z}[T_1, T_2, \ldots, T_n] \to A'$ that sends T_i to α_i maps Δ_n onto $\Delta(f)$. Since the images of the elementary symmetric polynomials s_k, which are sent to (plus or minus) the coefficients of f, are contained in A, we see that $\Delta(f)$ is also an element of A. In other words, the discriminant of a polynomial $f \in A[X]$ is an element of A, and it is given by a universal polynomial in the coefficients of f.

14.6. Example. The general formula for the discriminant of a cubic polynomial is difficult to remember, but the well-known formula

$$\Delta(X^3 + pX + q) = -4p^3 - 27q^2 \tag{14.7}$$

that arises by substituting $(s_1, s_2, s_3) = (0, p, -q)$ in the general expression for Δ_3 is both easy to remember and easy to deduce. After all, by taking $A = \mathbf{Z}[p, q]$, we know that the discriminant is a universal polynomial in p and q. Since there are only two monomials in $\mathbf{Z}[s_2, s_3]$ of weight $3(3-1) = 6$, namely s_2^3 and s_3^2, there in fact exist constants $c_1, c_2 \in \mathbf{Z}$ such that we have $\Delta(X^3 + pX + q) = c_1 p^3 + c_2 q^2$. We can easily calculate c_1 and c_2 by choosing a few suitable values for p and q. Namely, we have $c_1 = -\Delta(X^3 - X) = -4$ and $c_2 = \Delta(X^3 - 1) = -27$.

Exercise 4. Verify this.

▶ Resultant

Usually, discriminants of polynomials of higher degree in $A[X]$ are not calculated using the general formula for Δ_n; instead, the *resultant* is used. To avoid restrictions on division, if necessary, we replace the integral domain A with its field of fractions; from now on, we assume that $A = K$ is a *field*. For polynomials

$$f = a\prod_{i=1}^{n}(X - \alpha_i) \qquad \text{and} \qquad g = b\prod_{j=1}^{m}(X - \beta_j)$$

in $K[X]$ of degree n and m, respectively, the resultant $R(f, g)$ is defined by

$$R(f, g) = a^m b^n \prod_{i=1}^{n}\prod_{j=1}^{m}(\alpha_i - \beta_j). \tag{14.8}$$

The following properties are immediate from this definition:

(R1) $R(f, g) = (-1)^{mn} R(g, f)$.
(R2) $R(f, g) = a^m \prod_{i=1}^{n} g(\alpha_i)$.
(R3) If $g_1 \in K[X]$ is of degree m_1 with $g \equiv g_1 \bmod (f)$, then we have $R(f, g) = a^{m-m_1} R(f, g_1)$.

Using these properties and division with rest in $K[X]$, we can calculate resultants *without* ever needing any explicit knowledge of the zeros α_i and β_j that occur in the definition. Using (R1), if necessary, we may assume that we have $\deg(g) \geq \deg(f)$. We then use (R3) to replace g with the remainder $g_1 \in A[X]$ of the division of g by f. We can lower the degree of the polynomials by repeating these steps, and as soon as f has degree 0 or 1 and we know the zeros $\alpha_i \in A$, property (R2) gives the value of the resultant.

For a monic polynomial $f = \prod_{i=1}^{n}(X - \alpha_i)$, the derivative in a zero α_i is equal to

$$f'(\alpha_i) = (\alpha_i - \alpha_1)(\alpha_i - \alpha_2)\ldots(\alpha_i - \alpha_{i-1})(\alpha_i - \alpha_{i+1})\ldots(\alpha_i - \alpha_n).$$

If we take the product of these expressions for $i = 1, 2, \ldots, n$, then after counting the number of factors -1, we see that

$$\Delta(f) = (-1)^{n(n-1)/2} R(f, f'). \tag{14.9}$$

This is how we can use the resultant to calculate discriminants of polynomials in $K[X]$.

14.10. Example. 1. Take $K = \mathbf{Q}(p, q)$ and $f = X^3 + pX + q$. Then we have

$$\Delta(X^3 + pX + q) = -R(X^3 + pX + q, 3X^2 + p).$$

If we apply (R1) and then (R3) for $f = 3X^2+p$, $g = X^3+pX+q$, and $g_1 = g-(X/3)f = (2p/3)X + q$, then this becomes $-3^2 \cdot R(3X^2 + p, (2p/3)X + q)$. If we apply (R1) once more and recall that g_1 has a single zero $\alpha = -3q/(2p)$, then property (R2) gives

$$\Delta(X^3 + pX + q) = -3^2 \cdot \left(\frac{2p}{3}\right)^2 \left[3\left(\frac{-3q}{2p}\right)^2 + p\right] = -4p^3 - 27q^2,$$

in accordance with (14.7).

2. Take $K = \mathbf{Q}$ and $f = X^5 + X + 1$. Then we have

$$\begin{aligned}\Delta(X^5 + X + 1) &= R(X^5 + X + 1, 5X^4 + 1) = R(5X^4 + 1, X^5 + X + 1) \\ &= 5^4 \cdot R(5X^4 + 1, \frac{4}{5}X + 1) \\ &= 5^4 \cdot \left(\frac{4}{5}\right)^4 \cdot \left[5 \cdot \left(\frac{-5}{4}\right)^4 + 1\right] = 5^5 + 4^4 = 3381.\end{aligned}$$

Since f is irreducible in $\mathbf{Z}[X]$, the complex zeros of f do not lie in $\mathbf{Z}$ or $\mathbf{Q}$, and it is therefore not easy to give them "explicitly." However, this is not at all necessary to calculate the discriminant.

EXERCISES.

5. Suppose that $f \in \mathbf{Z}[T_1, T_2, \ldots, T_n]$ is invariant under every exchange $T_i \leftrightarrow T_j$ of two variables. Is f necessarily symmetric?
6. Let α_1, α_2, and α_3 be the zeros of $X^3 - X - 1$ in $\mathbf{C}$, and set $p_k = \alpha_1^k + \alpha_2^k + \alpha_3^k$ for all $k \in \mathbf{Z}$. Prove: the sequence $\{p_k\}_{k\in\mathbf{Z}}$ consists of integers that satisfy $p_{-1} = -1$, $p_0 = 3$, $p_1 = 0$, and the recurrence relation $p_k = p_{k-2} + p_{k-3}$ for $k \in \mathbf{Z}$.
7. Show: for $f \in \mathbf{Z}[T_1, T_2, \ldots, T_n]$ homogeneous of degree d and $r \in \mathbf{Z}$, we have
$$f(rT_1, rT_2, \ldots, rT_n) = r^d f(T_1, T_2, \ldots, T_n).$$
Conversely, is f homogeneous of degree d if this identity holds for all $r \in \mathbf{Z}$?
8. Give an example of a ring A and a non-homogeneous polynomial $f \in A[T_1, T_2, \ldots, T_n]$ that satisfies $f(rT_1, rT_2, \ldots, rT_n) = r^d f(T_1, T_2, \ldots, T_n)$ for all $r \in A$ for some $d \geq 1$.
9. Show that there are exactly $\binom{n+d-1}{n-1}$ distinct monomials $T_1^{e_1} T_2^{e_2} \ldots T_n^{e_n}$ of degree d.
10. Express the symmetric polynomials $\sum_n T_1^2 T_2$ and $\sum_n T_1^3 T_2$ in the elementary symmetric polynomials.
11. Show that the orbit of the polynomial $B = T_1T_2 + T_3T_4 \in \mathbf{Z}[T_1, T_2, T_3, T_4]$ under the action of S_4 consists of three elements B, B', and B'', and determine the cubic polynomial in $\mathbf{Z}[s_1, s_2, s_3, s_4][X]$ with these three zeros.
12. Show that the discriminant of the cubic polynomial from the previous exercise is equal to the discriminant Δ_4 of the general polynomial of degree 4, and use this to determine $\Delta_4 \in \mathbf{Z}[s_1, s_2, s_3, s_4]$.
13. Show that the resultant of the polynomials $f = \sum_{i=0}^n a_i X^i$ and $g = \sum_{j=0}^m b_j X^j$ of degree n and m in $K[X]$ is equal to the $(m+n) \times (m+n)$ determinant
$$\underbrace{\begin{vmatrix}
a_n & a_{n-1} & \cdots & a_0 & & \\
 & a_n & a_{n-1} & \cdots & a_0 & \\
 & & & \ddots & & \\
 & & a_n & a_{n-1} & \cdots & a_0 \\
b_m & b_{m-1} & \cdots & b_0 & & \\
 & b_m & b_{m-1} & \cdots & b_0 & \\
 & & & \ddots & & \\
 & & b_m & b_{m-1} & \cdots & b_0
\end{vmatrix}}_{m+n}$$
(with m rows of the a's and n rows of the b's) in terms of the coefficients of f and g. (Take all coefficients left blank equal to 0.) [Hint: show that this determinant also has properties (R1) and (R3).]
14. Prove: for $n \in \mathbf{Z}_{>0}$, we have $\Delta(X^n + a) = (-1)^{\frac{1}{2}n(n-1)} n^n a^{n-1}$.
15. Calculate the discriminant of the polynomial $X^4 + pX + q \in \mathbf{Q}(p, q)[X]$.
16. For every $n > 1$, determine an expression for the discriminant of the polynomial $X^n + pX + q \in \mathbf{Q}(p, q)[X]$.

17. Let $f \in \mathbf{Z}[X]$ be a monic polynomial. Prove that the following are equivalent:
    a. $\Delta(f) \neq 0$.
    b. The polynomial f has no double zeros in $\mathbf{C}$.
    c. The decomposition of f in $\mathbf{Q}[X]$ has no multiple prime factors.
    d. The polynomial f and its derivative f' are relatively prime in $\mathbf{Q}[X]$.
    e. The polynomials $f \bmod p$ and $f' \bmod p$ are relatively prime in $\mathbf{F}_p[X]$ for almost all prime numbers p.
18. Determine the "exceptional primes" in part (e) of the previous exercise for $f = X^3 + X + 1$ and for the polynomial $X^7 + 7X + 1$ from Exercise 12.23.
19. Let $f \in \mathbf{Q}[X]$ be a monic polynomial with $n = \deg(f)$ distinct complex zeros. Prove: the *sign* of $\Delta(f)$ is equal to $(-1)^s$, where $2s$ is the number of non-real zeros of f.
20. Prove: $X^3 + pX + q \in \mathbf{R}[X]$ has three (counted with multiplicity) real zeros $\iff$ $4p^3 + 27q^2 \leq 0$.
21. Express the sum of powers $\sum_n T_1^4$ in the elementary symmetric polynomials. Does the value of n matter?
22. A rational function $f \in \mathbf{Q}(T_1, T_2, \ldots, T_n)$ is called symmetric if it is invariant under all permutations of the variables T_i. Prove that every symmetric rational function is a rational function in the elementary symmetric functions.
23. Write $\sum_n T_1^{-1}$ and $\sum_n T_1^{-2}$ as rational functions in $\mathbf{Q}(s_1, s_2, \ldots, s_n)$.
24. *(Newton's identities)* Show that the sums of powers $\sigma_k = \sum_n T_1^k$ satisfy

$$\sigma_k - s_1\sigma_{k-1} + s_2\sigma_{k-2} - \ldots + (-1)^{k-1}s_{k-1}\sigma_1 + (-1)^k k s_k = 0 \quad \text{for } 1 \leq k \leq n$$

    and that we can use this to, inductively, write the sums of powers σ_k for $1 \leq k \leq n$ as polynomials in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$. Also show that for $k > n$, the sum of powers σ_k can be written as a polynomial in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$ from the relation

$$\sigma_k - s_1\sigma_{k-1} + s_2\sigma_{k-2} - \ldots + (-1)^n s_n\sigma_{k-n} = 0.$$

    [Hint: determine the logarithmic derivative $f'/f \in R[[X]]$ of $f = \prod_{i=1}^n (1 - T_i X) \in R[X]$.]
25. Can the method in the previous exercise also be used to write the sums of powers σ_k for $k < 0$ as elements of $\mathbf{Q}(s_1, s_2, \ldots, s_n)$?
26. Show that in terms of the sums of powers σ_k, the discriminant Δ_n can be written as

$$\Delta_n = \det\big(\sigma_{i+j-2}\big)_{i,j=1}^n.$$

    Use this relation to calculate $\Delta_3 \in \mathbf{Z}[s_1, s_2, s_3]$.
    [Hint: start with the Vandermonde-determinant $\det(T_i^{j-1})_{i,j=1}^n$.]
27. Show that the discriminant $\Delta_n \in \mathbf{Z}[s_1, s_2, \ldots, s_n]$ of the general polynomial of degree n is an irreducible polynomial in $\mathbf{Z}[s_1, s_2, \ldots, s_n]$. Is Δ_n also irreducible in the ring $\mathbf{C}[s_1, s_2, \ldots, s_n]$?

15 The Geometry of Commutative Rings

In *algebraic geometry*, where one studies solution sets of systems of polynomial equations, the polynomial rings $R = K[X_1, X_2, \ldots, X_n]$ over a field K from Chapter 13 play a central role. Every polynomial in R gives rise to a function $K^n \to K$, and for a k-tuple of elements $f_1, f_2, \ldots, f_k \in R$, we are interested in their set of common zeros, the zero locus

$$Z(f_1, f_2, \ldots, f_k) = \{x \in K^n : f_1(x) = f_2(x) = \ldots = f_k(x) = 0\}.$$

Sets of this form are called *algebraic sets* in K^n. Note that the points $x \in K^n$ in this definition are n-tuples $x = (x_1, x_2, \ldots, x_n)$ with coordinates in K. If $I = (f_1, f_2, \ldots, f_k) \subset R$ is the R-ideal generated by the polynomials f_i, then we have

$$Z(f_1, f_2, \ldots, f_k) = Z(I) = \{x \in K^n : \quad f(x) = 0 \quad \text{for all } f \in I\},$$

since every element of the form $g_1 f_1 + g_2 f_2 + \ldots + g_n f_n \in I$ with $g_i \in R$ assumes the value zero for $x \in Z(f_1, f_2, \ldots, f_k)$. The zero locus $Z(I)$ of the ideal $I \subset R$ therefore only depends on I, and not on the particular set of generators that we choose. Since polynomials in R that differ by an element in I assume the same value on $Z(I)$, we can view the quotient ring R/I as the *ring of polynomial functions* on the algebraic set $Z(I)$.

Exercise 1. Show that finite intersections of algebraic sets are algebraic sets.

15.1. Examples. Take $K = \mathbf{R}$ and pick $f_1 = X^2 + Y^2 - 1$ and $f_2 = XY - 1$ in $R = \mathbf{R}[X, Y]$. Then $Z(f_1)$ is the *unit circle* in the plane, and $R/(f_1)$ is the ring of polynomial functions on the unit circle. Similarly, $R/(f_2)$ is the ring of polynomial functions on the "unit hyperbola." Ring-theoretically, these are integral domains that differ from $\mathbf{R}[X, Y]$ and from each other in various ways: see Exercises 18–20.

Exercise 2. Let K be an *infinite* field. Show that distinct polynomials in $R = K[X_1, X_2, \ldots, X_n]$ cannot give rise to the same function $K^n \to K$. Find $I \subset R$ and two distinct elements in R/I that assume the same values on $Z(I)$.

Apart from the map

$$\begin{aligned} \{\text{ideals in } K[X_1, X_2, \ldots, X_n]\} &\longrightarrow \{\text{algebraic sets in } K^n\} \\ I &\longmapsto Z(I) \end{aligned}$$

that we just defined, there is also a map in the opposite direction that sends $V \subset K^n$ to the ideal

$$I(V) = \{f \in K[X_1, X_2, \ldots, X_n] : \quad f(x) = 0 \quad \text{for all } x \in V\} \subset R.$$

These two fundamental maps are used in algebraic geometry to set up an extensive dictionary between geometry and algebra in which *geometric* properties of the set $Z(I)$ are related to *algebraic* properties of the ring R/I. The dictionary has its simplest form when K is an *algebraically closed* field; in this case, every non-constant univariate polynomial in $K[X]$ can be written as a product of linear factors in $K[X]$. The field $K = \mathbf{C}$ of complex numbers is such a field, by the *fundamental theorem of algebra* 26.3.

Exercise 3. Prove the implications $I_1 \subset I_2 \Rightarrow Z(I_1) \supset Z(I_2)$ and $V_1 \subset V_2 \Rightarrow I(V_1) \supset I(V_2)$.

In this chapter, we will only take a little sip from the ocean of results that is provided by algebraic geometry. In the case of the algebraically closed base field $K = \mathbf{C}$, we will use low-dimensional examples to discuss the relation between the points of the set $Z(I)$ and the *maximal ideals* of the ring R/I and to define the *dimension* of arbitrary commutative rings. This will lead us to the modern insight that *every* commutative ring can be viewed as the "function ring" of a corresponding space, the *spectrum* $\operatorname{Spec}(R)$ of the ring.

▶ The Affine Plane

This section should actually be called *The Complex Affine Plane*, but since we will assume everywhere in this section that $\mathbf{C}$ is our base field, we will omit the adjective "complex" for most of the objects that we discuss.

The polynomial ring in n variables over $\mathbf{C}$ is the ring of polynomial functions on the *n-dimensional affine space* $\mathbf{A}^n(\mathbf{C}) = \mathbf{C}^n$ *over* $\mathbf{C}$. The adjective "affine" distinguishes this space from the *n-dimensional projective space* $\mathbf{P}^n(\mathbf{C})$ *over* $\mathbf{C}$, which we will not consider; that space can be obtained from $\mathbf{C}^n$ through "completion."[11]

The function ring of the *affine line* $\mathbf{A}^1(\mathbf{C}) = \mathbf{C}$ is the polynomial ring $\mathbf{C}[X]$ in one variable, which is a principal ideal domain by 12.6. Every ideal $I \subset \mathbf{C}[X]$ is of the form $I = (f)$, and the zero set $Z(I)$ is the set of zeros of f. For the trivial ideals $I = 0$ and $I = \mathbf{C}[X]$, we have $f = 0$ and $f = 1$, and the algebraic set $Z(I)$ is equal to $\mathbf{C}$ and empty, respectively. In the other cases, because we took $K = \mathbf{C}$, the ideal I is the product of a finite number of prime ideals of the form $(X - a) \subset \mathbf{C}[X]$ with $a \in \mathbf{C}$. Every prime ideal $(X - a)$ that divides f contributes a point a to the set $Z(I)$. In the "irreducible case" where $I = (X - a)$ is itself prime, we have $Z(I) = \{a\}$, and the quotient map $\mathbf{C}[X] \to \mathbf{C}[X]/I \cong \mathbf{C}$ is the evaluation map from 11.16 that sends a polynomial $f \in \mathbf{C}[X]$ to its value in a.

15.2. Definition. *An irreducible algebraic set* or *affine algebraic variety over* $\mathbf{C}$ *is a non-empty algebraic set* $Z \subset \mathbf{A}^n(\mathbf{C})$ *that cannot be written as a union* $Z = Z_1 \cup Z_2$ *of algebraic sets* $Z_i \subsetneq Z$.

In this section, for short, we call a complex affine algebraic variety simply a variety. Varieties in $\mathbf{A}^1(\mathbf{C})$ are not very exciting.

Exercise 4. Prove that a variety in $\mathbf{A}^1(\mathbf{C})$ is either a point or $\mathbf{A}^1(\mathbf{C})$ itself.

The case of varieties in the complex affine plane $\mathbf{A}^2(\mathbf{C})$, also called *plane varieties*, is less trivial. Here, we are dealing with the polynomial ring $\mathbf{C}[X, Y]$ in two variables. This is not a principal ideal domain, but by 13.3, it is a unique factorization domain.

Every point $(a, b) \in \mathbf{A}^2(\mathbf{C})$ is a variety; the corresponding evaluation map $f \mapsto f(a, b)$ is (Exercise 11.35) a surjection $\mathbf{C}[X, Y] \to \mathbf{C}$ with kernel $(X - a, Y - b)$, and the ring $\mathbf{C}[X, Y]/(X - a, Y - b)$ of polynomial functions in (a, b) is isomorphic to $\mathbf{C}$.

Of much greater interest are the zero set $Z(f)$ of non-constant polynomials $f \in \mathbf{C}[X, Y]$, which are called *plane algebraic curves* and will simply be referred to as curves

in this section. If f is irreducible, then we speak of an *irreducible curve*. Every non-constant polynomial in the unique factorization domain $\mathbf{C}[X,Y]$ can be written as a product $\prod_{i=1}^{t} p_i^{e(i)}$ of pairwise non-associated irreducible polynomials p_i, so every curve is a finite union $\bigcup_{i=1}^{t} Z(p_i)$ of irreducible curves.

Exercise 5. Show that a plane curve in $\mathbf{A}^2(\mathbf{C})$ has infinitely many points and that $\mathbf{A}^2(\mathbf{C})$ is not a finite union of plane curves.

We will prove that irreducible curves are varieties. First, we show that distinct irreducible curves intersect in only finitely many points.

15.3. Lemma. *Let f and g be relatively prime polynomials in $\mathbf{C}[X,Y]$. Then $Z(f,g)$ is a finite set.*

Proof. If we view f and g as polynomials in Y with coefficients in $\mathbf{C}[X]$, then they are relatively prime in the polynomial ring $(\mathbf{C}[X])[Y]$ in the variable Y. By the remark made before 13.5, f and g are then also relatively prime as polynomials in the principal ideal domain $\mathbf{C}(X)[Y]$ of polynomials in Y over the field $\mathbf{C}(X)$ of rational functions in X. In $\mathbf{C}(X)[Y]$, we have $r_1 f + r_2 g = 1$ for $r_i \in \mathbf{C}(X)[Y]$. By multiplying by a common multiple of the denominators of the coefficients of r_1 and r_2, we see that the ideal $(f,g) \subset \mathbf{C}[X,Y]$ contains a non-zero polynomial $h \in \mathbf{C}[X]$. We now have $h(x) = 0$ for every point $(x,y) \in Z(f,g)$, which shows that there are only finitely many possibilities for x. By symmetry, there are also only finitely many possibilities for y, so $Z(f,g)$ is finite. □

15.4. Theorem. *Every algebraic set in $\mathbf{C}^2$ is a finite union of varieties; these varieties varietyare the points, the irreducible curves, and the plane $\mathbf{C}^2$ itself. The map $I \mapsto Z(I)$ induces a bijection between the prime ideals of $\mathbf{C}[X,Y]$ and the irreducible algebraic sets in $\mathbf{A}^2(\mathbf{C})$.*

Proof. Let $Z(I)$ be a plane algebraic set. For $I = (0)$, we have $Z(I) = \mathbf{C}^2$. For $I \neq (0)$, there exists an $f \neq 0$ in I, and we have $Z(I) \subset Z(f)$. If we write the curve $Z(f)$ as a finite union $\bigcup_{i=1}^{n} Z(p_i)$ of irreducible curves, then we find

$$Z(I) = \bigcup_{i=1}^{n} \left(Z(I) \cap Z(p_i) \right).$$

There are two possibilities for each of the algebraic sets $Z(I) \cap Z(p_i)$. If all elements of I are divisible by p_i, then $Z(I) \cap Z(p_i) = Z(p_i)$ is an irreducible curve. If this is not the case, then I contains an element that is relatively prime to p_i and $Z(I) \cap Z(p_i)$ is a finite set by 15.3. We conclude that every plane algebraic set is a finite union of sets each equal to $\mathbf{C}^2$, to an irreducible curve, or to a point.

Points in the plane are clearly irreducible algebraic sets, and the same holds for the plane $\mathbf{C}^2$ itself (Exercise 5). To see that an irreducible curve $Z(p)$ associated with an irreducible polynomial p is also an irreducible algebraic set, we note that every polynomial that is identically zero on $Z(p)$ is divisible by p by 15.3 (and Exercise 5). Moreover, the irreducible element p is a prime element in the unique factorization

domain $\mathbf{C}[X,Y]$. If we have $Z(p) = Z(I_1) \cup Z(I_2)$, then all elements of $I_1 I_2$ are divisible by p. It follows that I_1 and I_2 cannot both contain an element that is not divisible by p—after all, the product of elements of I_1 and of I_2 is divisible by the prime element p—and we find that $Z(I_1)$ or $Z(I_2)$ is equal to $Z(p)$, as desired. We have proved the first two statements of 15.4.

The plane $\mathbf{C}^2$ is the zero set of the zero ideal (0), which is prime in the integral domain $\mathbf{C}[X,Y]$. An irreducible curve is by definition the zero set of a prime ideal (p) generated by an irreducible polynomial p. A point (a,b) is the zero set of the ideal $(X-a, Y-b)$, which, as the kernel of the evaluation map $\mathbf{C}[X,Y] \to \mathbf{C}$ in the point (a,b), is also prime.

Conversely, let $P \subset \mathbf{C}[X,Y]$ be a prime ideal. If $P \neq (0)$, then P contains a polynomial $f \neq 0$, which is not constant because $P \neq \mathbf{C}[X,Y]$. If we write f as a product of irreducible elements, then it follows from the definition of a prime ideal 12.14 that P contains an irreducible polynomial p. If p is a generator of P, then P is associated with the irreducible curve $Z(p)$. If p is not a generator of P, then P contains an element that is relatively prime to p, and as in the proof of 15.3, it follows that P contains non-constant polynomials in both X and Y. If we decompose these in linear factors over $\mathbf{C}$, then, again by the definition of a prime ideal, it follows that P contains an ideal of the form $(X-a, Y-b)$. Since $\mathbf{C}[X,Y]/(X-a,Y-b) \cong \mathbf{C}$ does not contain any non-trivial ideals, P is not strictly greater than $(X-a,Y-b)$ (Exercise 11.51), so $P = (X-a, Y-b)$ corresponds to the point (a,b). □

Exercise 6. Prove that every plane algebraic set can be written *uniquely* as a finite union of plane varieties that do not contain one another.

The algebraic theorem 15.4 expresses the geometric fact that there are three types of plane varieties: points, curves, and the plane itself. The fundamental theorems from algebraic geometry show that the analog of 15.4 for arbitrary dimension $n \geq 1$ holds: every algebraic set in $\mathbf{A}^n(\mathbf{C})$ is a finite union of algebraic varieties, and every algebraic variety $V \subset \mathbf{A}^n(\mathbf{C})$ is the zero set of a corresponding *prime ideal* $P = I(V) \subset R = \mathbf{C}[X_1, X_2, \ldots, X_n]$. The proofs of these theorems require more ring theory than we cover here.[12]

For a variety V, the associated ring $R/I(V)$ of polynomial functions on V is an integral domain, the *coordinate ring* R_V of V. All "geometric" properties of V, such as the dimension of V or the "smoothness" of the points of V, can be formulated in terms of the coordinate ring R_V.

Exercise 7. Show that there is a natural correspondence between the points of the "unit circle" $V = Z(X^2+Y^2-1) \subset \mathbf{A}^2(\mathbf{C})$ and the prime ideals $P \neq 0$ of $R_V = \mathbf{C}[X,Y]/(X^2+Y^2-1)$.

Algebro-geometric characterizations can often be generalized to other base fields than $\mathbf{C}$, and in 15.4, one can take an arbitrarily algebraically closed field for $\mathbf{C}$. The geometry over non-algebraically closed fields such as $\mathbf{Q}$ or $\mathbf{F}_p$, also called *arithmetic algebraic geometry*,[13] is often much more complicated. This geometry forms an interface between number theory and geometry that currently draws great interest. Unlike in classical geometry, it is much more difficult here to derive proofs from "figures."

▶ Dimension

We give an example of an algebraic characterization of a geometric concept using the notion of *dimension*. Intuitively, it is clear that a variety in the affine plane, which by 15.4 is the entire plane, an irreducible plane curve, or a point in the plane, has dimension 2, 1, and 0, respectively. Since the dimension always increases for inclusions of varieties of the type point–curve–plane, it is natural to define the dimension of a variety V "geometrically" in terms of maximal lengths of sequences of "nested subvarieties" of V. In other words, $V \subset \mathbf{A}^n(\mathbf{C})$ has dimension $\dim(V) = d$ if there exists a strictly ascending sequence $V_0 \subsetneq V_1 \subsetneq \ldots \subsetneq V_d = V$ of varieties $V_i \subset V$ of *length* d but no sequence of length greater than d.

If we interpret the above sequence in terms of inclusions of prime ideals in the ring $R = \mathbf{C}[X_1, X_2, \ldots, X_n]$, then we see that $\dim(V)$ is the maximal length of a descending sequence $P_0 \supsetneq P_1 \supsetneq \ldots \supsetneq P_d = I(V)$ of prime ideals $P_i \subset R$ that contain $I(V)$. If, using the correspondence of Exercise 11.51, we view the prime ideals P_i as prime ideals in the coordinate ring $R_V = R/I(V)$ of V, then the dimension of V is nothing but the maximal length of a *chain* of prime ideals in R_V. A chain of ideals of length d in a ring is, by definition, a collection $\{I_k\}_{k=0}^d$ of ideals for which strict inclusions $I_0 \subsetneq I_1 \subsetneq \ldots \subsetneq I_d$ hold.

15.5. Definition. *The dimension* $\dim(R)$ *of a commutative ring* R *is the supremum of the lengths of the chains of prime ideals in* R.

With this definition, the dimension of an affine variety is equal to the dimension of its coordinate ring R_V. The dimension in 15.5, also called the *Krull dimension* of R after the German algebraist Wolfgang Krull (1899–1971), is, however, *much* more general than our "geometric" dimension, which is only meaningful in the context of varieties over algebraically closed fields. Indeed, *every* commutative ring that has a prime ideal is assigned a dimension by 15.5. We will prove in 15.10 that every commutative ring $R \neq 0$ has a prime ideal.

15.6. Examples. A *zero-dimensional ring* is a ring in which there are no inclusions between prime ideals. An example is the coordinate ring $\mathbf{C}$ of a point in $\mathbf{A}^n(\mathbf{C})$ or, more generally, an arbitrary field K. After all, in a field, (0) is the only prime ideal.

Every *finite* ring R, such as $\mathbf{Z}/n\mathbf{Z}$, is zero-dimensional. Indeed, for a prime ideal P in a finite ring R, the quotient R/P is a finite integral domain and therefore a field by Exercise 11.18. This means that there are no inclusions between prime ideals in R, and we find that $\dim(R) = 0$.

For $R = \mathbf{Z}$, the prime ideals are the zero ideal (0) and the ideals $p\mathbf{Z}$ for the prime numbers p. There is a chain $(0) \subset p\mathbf{Z}$ of length 1, and since distinct primes do not divide each other, we have $\dim(\mathbf{Z}) = 1$. More generally, the proof of 12.10 shows that every prime ideal $(p) \neq (0)$ in a principal ideal domain R is a "maximal" ideal: for $x \notin (p)$ arbitrary, we have $(p, x) = R$. By 12.11, a principal ideal domain R that is not a field contains a prime ideal $(p) \neq (0)$; we find that $\dim(R) = 1$.

Exercise 8. Prove that $\dim(\mathbf{Z}[\sqrt{-5}]) = 1$.

It should not come as a surprise that the coordinate ring $R = \mathbf{C}[X_1, X_2, \ldots, X_n]$ of the affine n-dimensional space $\mathbf{A}^n(\mathbf{C})$ has dimension n. We have already seen this for $n \leq 2$. For $n \geq 3$, it is not difficult to see that the chain $\{P_k\}_{k=0}^n$, with $P_0 = (0)$ and $P_k = (X_1, X_2, \ldots, X_k)$ for $k \geq 1$ the ideal generated by the first $k \leq n$ variables, is a chain of prime ideals of length n. We do not prove here that there do not exist any longer chains.[14]

▶ Maximal Ideals

The largest ideals in the chains of prime ideals of maximal length in a ring are the so-called *maximal ideals* of the ring: ideals that cannot be enlarged without immediately obtaining the entire ring. In geometry, such ideals are related to the "points" of varieties.

15.7. Definition. *An ideal $I \subset R$ in a commutative ring R is called maximal if it is not equal to R and there do not exist any ideals $J \subset R$ with $I \subsetneq J \subsetneq R$.*

Since the ideals J that satisfy the inclusions $I \subset J \subset R$ correspond to the ideals of the quotient ring R/I, 15.7 says that I is maximal in R if and only if R/I is not the zero ring and has only trivial ideals. In a commutative ring, every element $x \neq 0$ that is not a unit generates a non-trivial ideal (x), so fields are the only commutative rings $R \neq 0$ without non-trivial ideals. This gives the following analog of (12.15):

$$I \subset R \text{ is a maximal ideal } \iff R/I \text{ is a field.} \tag{15.8}$$

In particular, the zero ideal $(0) \subset R$ is maximal if and only if R is a field. Since every field is an integral domain, (12.15) and (15.8) give the following implication.

15.9. Lemma. *Every maximal ideal in a commutative ring R is prime.* □

Every point $a = (a_1, a_2, \ldots, a_n)$ of an affine variety $V \subset \mathbf{A}^n(\mathbf{C})$ gives rise to an evaluation map $R_V \to \mathbf{C}$ with kernel (Exercise 11.36) the prime ideal

$$M_a = (X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n) \subset R_V.$$

The isomorphism theorem gives $R_V/M_a \cong \mathbf{C}$, so by (15.8), M_a is maximal in R_V. For $n \leq 2$, it follows from 15.4 that *all* maximal ideals are of the form M_a for a point $a \in V$. This is also true for arbitrary n, but the proof is more difficult. This is one of the formulations of the so-called *Hilbert Nullstellensatz* from algebraic geometry. We do not give the proof in this syllabus.[15]

For an arbitrary commutative ring $R \neq 0$, it is not a priori obvious that R has a maximal ideal. In theory, they seem very simple to construct. One starts with the zero ideal (0) and checks whether it is maximal. If this is not the case, then by 15.7, there exists an ideal I_1 with $(0) \subsetneq I_1 \subsetneq R$. If I_1 is not yet maximal, then we take an ideal I_2 with $I_1 \subsetneq I_2 \subsetneq R$ and check whether it is maximal. As long as the new ideal is not maximal, we can enlarge it, and we continue doing so "until this is no longer possible without obtaining the entire ring." The resulting ideal is then maximal. The following theorem therefore sounds very plausible.

15.10. Theorem. *Every commutative ring $R \neq 0$ has a maximal ideal.*

To prove Theorem 15.10, we must show that the process $I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \ldots$ we just described of "repeatedly enlarging the ideal" ultimately leads to a maximal ideal. In *Noetherian* rings, which by definition do not have infinite ascending chains of ideals (Exercise 12.33), it is clear that the process we described yields a maximal ideal after finitely many steps. In arbitrary rings, however, an uncountable number of steps may be required, and in such situations, it is not always possible to specify the obtained ideal "explicitly" (Exercises 40–41). Dissatisfaction with this situation played an important role in the emergence of *intuitionism* in mathematics.[16]

▶ Zorn's Lemma

All proofs in 15.10 use the *axiom of choice* from set theory. We use this axiom, which cannot be deduced from the "basic axioms" of set theory, in the following form:[17]

15.11. Zorn's lemma. *Let X be a partially ordered set in which every chain has an upper bound in X. Then X contains a maximal element.*

First, we give a brief explanation for those unfamiliar with this lemma. A *partial ordering* on a set X is, by definition, a relation $\leq$ on X with the following properties:

(P1) For $x \in X$, we have $x \leq x$.
(P2) If $x \leq y$ and $y \leq z$, then we also have $x \leq z$.
(P3) If $x \leq y$ and $y \leq x$, then we have $x = y$.

A standard example of a partially ordered set is the power set $\mathcal{P}(A)$ of a set A, the collection of all subsets of A, with the *inclusion relation* as partial ordering. Partial orderings on the set $\mathbf{Z}_{\geq 0}$ of natural numbers are the "usual" ordering $\leq$ that we know from $\mathbf{R}$, but also the *divisibility relation* $x|y$. For an abstract partial ordering on X, we generally say "x less than or equal to y" for $x \leq y$.

A *chain* in a partially ordered set is a *totally ordered* subset $K \subset X$. This means that for any two elements $x, y \in K$, we have either $x \leq y$ or $y \leq x$. For every n-tuple of elements in a chain, after possibly renumbering, we have

$$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_{n-1} \leq x_n.$$

For the ordering of the subgroups in a group by inclusion, we already encountered this in Exercise 2.29.

An *upper bound* in X for a subset $Y \subset X$ is an element $x \in X$ that satisfies $y \leq x$ for all $y \in Y$. Every element $x \in X$ is an upper bound for the empty subset $\emptyset \subset X$, which, in the absence of any requirements, is also a chain in X, the *empty chain*. A *maximal element* in X is an element $x \in X$ such that the only element $z \in X$ that satisfies $x \leq z$ is the element x itself: there are no elements in X that are "truly greater" than x.

Exercise 9. What are the maximal elements of $\mathcal{P}(A)$ and $\mathbf{Z}_{\geq 0}$ with respect to the partial orderings mentioned above?

Proof of 15.10. The existence of a maximal ideal in a commutative ring $R \neq 0$ is a direct application of 15.11 to the collection of ideals

$$X = \{I \subset R : I \text{ is an ideal different from } R\}$$

with the inclusion relation as partial ordering. The zero ideal $(0) \in X$ is an upper bound for the empty chain in X. For a non-empty chain $K = \{I_k\}_k$ in X, the union $I = \bigcup_k I_k$ is an ideal in R (verify this!). Since $1 \in R$ is not contained in any ideal I_k, we have $1 \notin I$ and $I \neq R$, so I is an upper bound for K that lies in X. By 15.11, there exists a maximal element in X, and that is a maximal ideal of R. □

15.12. Corollary. *Every ideal $I \subsetneq R$ is contained in a maximal ideal of R. The union of all maximal ideals of R is equal to*

$$\bigcup_{M \subset R \text{ maximal}} M = R \setminus R^*.$$

Proof. The maximal ideals M/I of the quotient ring $R/I \neq 0$ correspond to the maximal ideals $M \subset R$ that contain I, and by 15.10, such an ideal exists. If we take the principal ideal (x) generated by a non-unit $x \in R$ as I, then it follows that every non-unit of R is contained in a maximal ideal of R. Since units are not contained in any maximal ideals of R, it follows that $\bigcup_M M = R \setminus R^*$. □

For a principal ideal domain R, 15.12 boils down to the simple statement that every element that is not a unit is divisible by a prime element.

15.13. Example. Let $R = \mathrm{Map}(X, \mathbf{R})$ be the ring of real-valued functions on a finite set X. For every point $x \in X$, the kernel of the evaluation map $f \mapsto f(x)$ in the point x is equal to the ideal $M_x = \{f \in R : f(x) = 0\}$ of R. The isomorphism theorem gives an isomorphism $R/M_x \cong \mathbf{R}$, so by 15.8, M_x is maximal. The complement of $\bigcup_{x \in X} M_x$ in R consists of the functions on X without zeros, which are exactly the units of R.

To prove that the maximal ideals of R are exactly the ideals M_x, it suffices to show that every ideal $I \subsetneq R$ is contained in an ideal M_x. If I is not contained in any ideal of the form M_x, then for every element $x \in X$, there exists a function $f_x \in I$ with $f_x(x) \neq 0$. The function $f = \sum_{x \in X} f_x^2 \in I$ is now strictly positive on X, hence a unit in R. It follows that $I = R$.

Exercise 10. Use the isomorphism $R \cong \prod_{x \in X} \mathbf{R}$ and Exercise 11.47 to prove that the maximal ideals of R are the ideals M_x.

For *every* ring R of functions on a set X with values in a field, if R contains all constant functions, then, as in 15.13, the points of X give rise to maximal ideals via point evaluations. In many (but not all) cases, this is a bijection, so that one can "reconstruct" X from the set of maximal ideals of R. See Exercise 29 for an example.

▶ Nilradical

Zorn's lemma is useful in all sorts of other situations where maximal sets with a given property need to be constructed. We apply it to describe the *nilradical* $\mathrm{nil}(R)$ of a

commutative ring R. This is by definition the subset of R consisting of the elements x for which a positive power x^k is equal to zero. Such elements in R are called *nilpotent*. The zero element of R is clearly nilpotent, and if R is an integral domain, it is the only element with this property. For the ring $R = \mathbf{Z}/p^2\mathbf{Z}$, however, we have $\mathrm{nil}(R) = p\mathbf{Z}/p^2\mathbf{Z}$.

15.14. Theorem. *The nilradical* $\mathrm{nil}(R)$ *is an ideal of* R*, and it is equal to*

$$\mathrm{nil}(R) = \bigcap_{P \subset R \text{ prime}} P.$$

Proof. If $x \in \mathrm{nil}(R)$ is a nilpotent element and $P \subset R$ is a prime ideal, then we have $x^k = 0 \in P$ for a suitable k, which implies that $x \in P$ because P is prime. So $\mathrm{nil}(R)$ is contained in the intersection of all prime ideals.

Now let $f \in R$ be an element that is not nilpotent. We want to show that there is a prime ideal P that does not contain f. This time, we take the collection of ideals

$$X = \{I \subset R : I \text{ is an ideal that does not contain any powers of } f\}$$

with the inclusion relation as partial ordering on X. Since f is not nilpotent, we have $(0) \in X$, so the empty chain has an upper bound in X. For a non-empty chain, we obtain an upper bound by taking the union of the chain. If we apply 15.11 again, then it follows that X contains a maximal element P. This is an ideal of R that does not contain any powers of f, and we claim that P is a *prime ideal* of R.

Suppose that $x, y \in R$ have product $xy \in P$ but that P contains neither element. Then by the maximality of P, the ideals $P + (x)$ and $P + (y)$, which are each strictly larger than P, each do contain a power of f. The product ideal $(P + (x)) \cdot (P + (y))$ then also contains a power of f. Since

$$(P + (x)) \cdot (P + (y)) \subset P + (xy) \subset P,$$

P then also contains a power of f, so we have a contradiction. We conclude that P is a prime ideal that does not contain f, and this proves that $f \notin \bigcap_{P \subset R \text{ prime}} P$, as desired. In particular, the just-proved identity shows that $\mathrm{nil}(R)$, as an intersection of ideals, is an ideal of R. □

Exercise 11. Prove that $\mathrm{nil}(R)$ is an ideal of R *without* using the description from 15.14.

▶ Spectrum of a Ring

For coordinate rings and many other function rings from geometry, the elements of the ring can be viewed as functions on the set of maximal ideals of the ring. In non-geometric situations, such a description is often also possible: for every maximal ideal $p\mathbf{Z}$, an element $x \in \mathbf{Z}$ has "function value" $x \bmod p$ in the residue field $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ corresponding to the "point" $p\mathbf{Z}$ of $\mathbf{Z}$. It is moreover clear that an element $x \in \mathbf{Z}$ is uniquely determined by its function values $x \bmod p$.

Exercise 12. Give an example of a commutative ring R and distinct elements $x, y \in R$ such that $x \equiv y \bmod M$ for all maximal ideals $M \subset R$.

The idea of viewing elements of a ring R as functions on the set of maximal ideals of R proves to be very evocative. It is even better to consider the *spectrum* $\mathrm{Spec}(R)$ of R, the set consisting of *all* prime ideals of R. This has the nice property that for every homomorphism $f : R_1 \to R_2$ of commutative rings, there is an induced map

$$f^* : \mathrm{Spec}(R_2) \longrightarrow \mathrm{Spec}(R_1), \quad P_2 \longmapsto f^{-1}[P_2]. \tag{15.15}$$

Indeed, the kernel $f^{-1}[P_2]$ of the composed map $R_1 \xrightarrow{f} R_2 \to R_2/P_2$ is a prime ideal by 12.15 and the simple fact that a subring of an integral domain is also an integral domain.

Exercise 13. Show that the inverse image of a maximal ideal is not always maximal.

In the early 1960s, the French mathematician Grothendieck[18] showed how interpreting rings as function spaces on spectra can be used to formulate the entire theory of commutative rings in algebro-geometric terms. The method in geometry, dating back to Bernhard Riemann (1826–1866), of constructing varieties by gluing together "local pieces" (*charts*) leads to the definition of *schemes* when applied to spectra. Conversely, one can set up all of algebraic geometry in terms of schemes. The resulting flexibility makes it possible to develop geometry not only over $\mathbf{C}$ or $\mathbf{R}$, but over arbitrary base fields or even commutative base rings. This approach has proven to be very productive in arithmetic algebraic geometry.

The algebraic abstraction inherent in the theory has long given Grothendieck's schemes an unattainable aura. However, because of the theory's successful applications over the past forty years, its underlying geometric ideas have become part of the foundations of algebra.

▶ Topology of Spectra

The spectrum of a ring is not only a *set* of prime ideals; it has a natural structure of *topological space*, although of a somewhat different type than the familiar *metric* topological spaces. The maps f^* from (15.15) induced by ring homomorphisms are *continuous* for the natural topology on spectra, the Zariski topology named after the geometer Oscar Zariski (1899–1986) (Exercise 49).

Recall that a collection $\mathcal{U}$ of subsets of a set X is called a *topology* on X if $\emptyset$ and X are contained in $\mathcal{U}$ and $\mathcal{U}$ is closed under taking *arbitrary* unions and intersections of *finitely* many elements of $\mathcal{U}$. The elements of $\mathcal{U}$ are called the *open* subsets of X. The complement of an open subset of X is called a *closed* subset of X.

The Zariski topology is the topology in the following theorem.

15.16. Theorem. *The set* $\mathrm{Spec}(R)$ *of prime ideals of a commutative ring* R *admits a topology in which the closed sets are the sets of the form*

$$Z(I) = \{P \text{ prime} : P \supset I\} \subset \mathrm{Spec}(R),$$

with I *an ideal of* R*. The topological space* $\mathrm{Spec}(R)$ *is compact.*

Proof. The subsets $Z(I) \subset \mathrm{Spec}(R)$ form the closed sets of a topology if their complements satisfy the axioms for a topology on $\mathrm{Spec}(R)$. The sets $Z(R) = \emptyset$ and $Z(0) = \mathrm{Spec}(R)$ are both open and closed.

The identity $Z(I_1) \cup Z(I_2) = Z(I_1 I_2)$, which easily follows from the prime ideal property, shows that finite unions of closed sets are closed, and therefore finite intersections of open sets are open.

An arbitrary intersection $\bigcap_\alpha Z(I_\alpha)$ of closed sets $Z(I_\alpha)$ is closed, since it is equal to $Z(I)$, with I the ideal *generated* by the ideals I_α. This shows that unions of open sets are open, so the Zariski topology is indeed a topology. It also follows that the closed sets in $\mathrm{Spec}(R)$ have the "finite-intersection property," which is equivalent to the compactness of $\mathrm{Spec}(R)$. Indeed, if $\bigcap_\alpha Z(I_\alpha) = Z(I)$ is empty, then I is not contained in any prime ideal, so by 15.12, it is equal to R. If we write $1 \in I$ as an R-linear combination of elements from the ideals I_α, then we only need finitely many ideals I_α, and the finite intersection of the corresponding sets $Z(I_\alpha)$ is then empty. □

Exercise 14. Prove that $Z(I_1) \cup Z(I_2) = Z(I_1 \cap I_2) = Z(I_1 I_2)$.

15.17. Examples. 1. The spectrum of the zero ring is the empty set. The spectrum $\mathrm{Spec}(K) = \{(0)\}$ of a field K is a singleton.

2. The spectrum $\mathrm{Spec}(R)$ of a principal ideal domain R consists of the ideals (p) generated by the irreducible elements of R, which are prime by 12.10, and the zero ideal (0). In particular, we have

$$\mathrm{Spec}(\mathbf{Z}) = \{(0)\} \cup \{p\mathbf{Z} : p \text{ a prime}\},$$
$$\mathrm{Spec}(\mathbf{C}[X]) = \{(0)\} \cup \{(X - \alpha) : \alpha \in \mathbf{C}\}.$$

For $I = (x) \subset R$, the closed set $Z(I) = Z(x)$ consists of the prime ideals (p) that divide (x). For $x \neq 0$, this is a *finite* set; for $x = 0$, it is $\mathrm{Spec}(R)$ itself. The zero ideal is not contained in any closed set different from $\mathrm{Spec}(R)$. We conclude that the open sets $U \neq \emptyset$ in $\mathrm{Spec}(R)$ are the sets that contain $\{(0)\}$ and have finite complement. This topology has the property that for any two non-empty open sets U_1, U_2, the intersection $U_1 \cap U_2$ is non-empty. Note that this phenomenon does not occur in the classical (metric) topological spaces, which are always Haussdorff spaces.

3. It follows from 15.4 that $\mathrm{Spec}(\mathbf{C}[X, Y])$ consists of three types of prime ideals: the zero ideal, the prime ideals (p) corresponding to the irreducible curves in $\mathbf{A}^2(\mathbf{C})$, and the maximal ideals $(X - a, Y - b)$ corresponding to the points $(a, b) \in \mathbf{C}^2$. For $I \neq (0)$, the closed set $Z(I) \subset \mathrm{Spec}(\mathbf{C}[X, Y])$ consists of the prime ideals corresponding to the points and the irreducible curves contained in the *algebraic set* $Z(I) \subset \mathbf{A}^2(\mathbf{C})$. We see that plane algebraic sets and closed subsets of $\mathrm{Spec}(\mathbf{C}[X, Y])$ are essentially the same: the topological space $\mathrm{Spec}(\mathbf{C}[X, Y])$ is an "algebraic model" for the affine plane. The corresponding *Zariski topology* on the affine plane is *not* the well-known metric topology on $\mathbf{C}^2$ (Exercise 45).

Every closed set $Z(I)$ that contains a point $P \in \mathrm{Spec}(R)$ also contains all prime ideals of R that contain P. It follows that the *closure* of $\{P\}$ is equal to $Z(P)$. In particular, the "closed points" of $\mathrm{Spec}(R)$ are exactly the *maximal ideals* of R. In

integral domains R, the zero ideal (0) is a point whose closure is the entire space $\mathrm{Spec}(R)$. It is also called the *generic point* of $\mathrm{Spec}(R)$.

EXERCISES.

All rings R in the following exercises are commutative.

15. Show that the zero set $Z(f_1, f_2)$ in Example 15.1 is the empty set but that $R/(f_1, f_2)$ is not the zero ring. [Hint: evaluate $f \in R$ in a suitable point of $\mathbf{A}^2(\mathbf{C})$.]
16. Show that a finite field is not algebraically closed.
17. Determine which of the following subsets of $\mathbf{C}^2$ are algebraic sets:
    a. $\{(t^2, t^3) : t \in \mathbf{C}\}$,
    b. $\{(t, \sin t) : t \in \mathbf{C}\}$,
    c. $\{(\cos t, \sin t) : t \in \mathbf{C}\}$,
    d. $\{(e^t, \sin t) : t \in \mathbf{C}\}$,
    e. $\{(e^t + e^{-t}, e^t - e^{-t}) : t \in \mathbf{C}\}$,
    f. $\{(e^{2t}, e^{3t}) : t \in \mathbf{C}\}$.
18. Let $R = \mathbf{R}[X, Y]/(X^2 + Y^2 - 1)$ be the ring of real polynomial functions on the unit circle.
    a. Show that $\mathbf{R}^*$ is the group of units of R.
    [Hint: use a norm function $f(X) + Yg(X) \mapsto f(X)^2 - (1 - X^2)g(X)^2$, as we did for the ring $\mathbf{Z}[i]$ in 12.19.]
    b. Show that $M = (X - 1, Y) \subset R$ is the kernel of the point evaluation in $(1, 0)$.
    c. Show that M is not a principal ideal.
    [Hint: what can the norm be of an element that divides both $X - 1$ and Y?]
    d. Show that $X - 1$ and $Y - 1$ are irreducible elements in R that are not prime. Conclude: R is not a unique factorization domain.
    [Hint: we have $(X + Y - 1)^2 = 2(X - 1)(Y - 1) \in R$—draw a picture!]
19. Show that the ring of trigonometric polynomials from Exercise 13.17 is isomorphic to the ring of polynomial functions on the unit circle from the previous exercise.
20. Let $R = \mathbf{R}[X, Y]/(XY - 1)$ be the ring of polynomial functions on the "unit hyperbola."
    a. Show that R is isomorphic to the ring $\mathbf{R}[T, T^{-1}] \subset \mathbf{R}(T)$ of Laurent polynomials.
    b. Prove that R is a principal ideal domain with group of units
    $$R^* = \{c\overline{X}^i c \in \mathbf{R}^*, i \in \mathbf{Z}\} \cong \mathbf{R}^* \times \langle X \rangle$$
    with $\overline{X} = X \bmod (XY - 1)$.
21. Prove that the rings $R = \mathbf{C}[X, Y]/(X^2 + Y^2 - 1)$ and $\mathbf{C}[U, V]/(UV - 1)$ are isomorphic, and determine a generator of the ideal $(X - 1, Y) \subset R$.
22. Are the rings $K[X, Y]/(X^2 + Y^2 - 1)$ and $K[U, V]/(UV - 1)$ isomorphic for K equal to $\mathbf{R}$, $\mathbf{F}_5$, and $\mathbf{F}_3$?
23. Show that the ring Ω from Exercise 13.15 has infinite Krull dimension.
24. Let $P \subset \mathbf{Z}[X]$ be a prime ideal.

a. Prove that $P \cap \mathbf{Z} = p\mathbf{Z}$ with $p = 0$ or p a prime.
b. Prove that the prime ideals $P \subset \mathbf{Z}[X]$ with $P \cap \mathbf{Z} = 0$ are (0) and the principal ideals $P = (f)$, with $f \in \mathbf{Z}[X]$ non-constant and irreducible.
c. Prove that the prime ideals $P \subset \mathbf{Z}[X]$ such that $P \cap \mathbf{Z} = p\mathbf{Z}$ with p prime are the ideals of the form $P = (p, f_p)$, with $f_p \in \mathbf{Z}[X]$ a polynomial that is irreducible modulo p or equal to 0.
d. Conclude that $\mathbf{Z}[X]$ is a 2-dimensional ring.

25. For each of the following ideals in $\mathbf{Z}[X]$, determine whether it is prime or maximal:

$$(X - 7, 3), \qquad (X^2 - 7), \qquad (X^2 - 7, 3), \qquad (X^2 - 7, 5).$$

Answer the same question for $\mathbf{Q}[X]$ instead of $\mathbf{Z}[X]$.

26. A commutative ring R with exactly one maximal ideal is called a *local ring*. Prove that R is local if and only if $R \setminus R^*$ is an ideal in R.

27. Show that the ring $R_p \subset \mathbf{Q}$ from Exercise 11.17 is a local ring. More generally, prove that for a prime element p in a principal ideal domain R, the ring $R_p = \{\frac{a}{b} : \mathrm{ord}_p(b) = 0\}$ is a subring of $K = Q(R)$ and that it is a local ring.

28. A *discrete valuation* on a field K is a surjective group homomorphism $v : K^* \to \mathbf{Z}$ that satisfies $v(x + y) \geq \min\{v(x), v(y)\}$ for $y \neq -x$. The corresponding *discrete valuation ring* is defined as $R_v = \{0\} \cup \{x \in K^* : v(x) \geq 0\}$.

a. Show that for every prime number p, the function $\mathrm{ord}_p : \mathbf{Q}^* \to \mathbf{Z}$ from §6 is a discrete valuation on $\mathbf{Q}$ and that the subring $R_p \subset \mathbf{Q}$ from the previous exercise is the corresponding discrete valuation ring.
b. Show that a discrete valuation ring R_v is a local ring and that every element $\pi \in R_v$ with $v(\pi) = 1$ generates the maximal ideal of R_v.
c. Prove that if R_v and π are as in (b), then every element $x \in K^*$ can be uniquely written as $x = u\pi^k$ with $u \in R_v^*$ and $k \in \mathbf{Z}$.

29. Let $R = C([0, 1])$ be the ring of continuous real-valued functions on $[0, 1]$. Prove that the maximal ideals of R are the kernels M_x of the point evaluations.
[Hint: use compactness.]

30. Determine all prime ideals of $\mathbf{Z}/300\mathbf{Z}$, as well as its nilradical.

31. Show that the following are equivalent:

a. $R/\mathrm{nil}(R)$ is a field.
b. $R \neq 0$, and every element of R is nilpotent or a unit.
c. R has exactly one prime ideal.

32. The *Jacobson radical* $J(R)$ of R is the set of elements $x \in R$ for which $1 + xR = \{1 + xr : r \in R\} \subset R^*$ holds. Prove that $J(R)$ is an ideal of R and that it is equal to the intersection of all maximal ideals $M \subset R$. (Convention: an empty intersection of ideals is equal to the entire ring.)

33. A ring R is called *reduced* if $\mathrm{nil}(R) = 0$. Show that $R/\mathrm{nil}(R)$ is a reduced ring and that the map $P \mapsto P/\mathrm{nil}(R)$ maps the primes of R bijectively onto those of $R/\mathrm{nil}(R)$.

34. Show that R is reduced if and only if R can be embedded in a product of fields.

35. Prove that for every ideal $I \subset R$, the radical $\sqrt{I} = \{x \in R : x^k \in I \text{ for some } k \geq 1\}$ of I is an ideal of R and that it is the intersection of all prime ideals that contain I. Is $R/\sqrt{I}$ a reduced ring?

36. Prove that a polynomial $f \in R[X]$ is nilpotent if and only if all coefficients of f are nilpotent in R.

37. Show that for $x \in \mathrm{nil}(R)$, the element $1 + x$ is a unit and that, more generally, the sum of a unit and a nilpotent element is a unit.

38. Show that $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ is a unit in $R[X]$ if and only if a_0 is a unit in R and the coefficients a_i with $i > 0$ are nilpotent.
[Hint: reduce modulo primes of R.]

39. Show that $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ is a zero divisor in $R[X]$ if and only if there exists an element $a \in R \setminus \{0\}$ with $af = 0$.
[Hint: suppose $gf = 0$ with $g = \sum_{j=0}^{m} b_j X^j$ of minimal degree; use $a_n f g = 0$ to obtain, by induction, $a_{n-k} g = 0$ for $k = 0, 1, \ldots, n$.]

40. Call an $\mathbf{F}_2$-valued function $f \in R = \mathrm{Map}(\mathbf{Z}, \mathbf{F}_2)$ on $\mathbf{Z}$ *small* if it assumes the value $1 \in \mathbf{F}_2$ in only finitely many points of $\mathbf{Z}$. Prove that there exists a ring homomorphism $\phi : R \to \mathbf{F}_2$ that assumes the value zero on all small functions.

41. Show that there is a way to divide the subsets of $\mathbf{Z}$ into "small" and "large" sets in such a way that the following hold:
    1. Finite sets are small.
    2. Subsets of small sets are small.
    3. An intersection of two large sets is large.
    4. A union of two small sets is small.
    5. The set of prime numbers is large.

*42. Let X be a set, $\mathcal{F}$ a collection of subsets of X, and $\mathcal{P}(X) \cong \mathrm{Map}(X, \mathbf{F}_2)$ the Boolean ring from Exercise 11.31. We call $\mathcal{F}$ a *filter* on X if the following hold:

$$\begin{aligned} &X \in \mathcal{F}, \qquad \emptyset \notin \mathcal{F}, \\ &A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}, \\ &A \in \mathcal{F} \ \&\ A \subset B \implies B \in \mathcal{F}. \end{aligned}$$

If we moreover have $(A \cup B \in \mathcal{F} \implies A \in \mathcal{F} \text{ or } B \in \mathcal{F})$, then $\mathcal{F}$ is called an *ultrafilter* on X. Prove the following statements:

a. $\mathcal{F}$ is a filter $\Leftrightarrow \{A \subset X : (X \setminus A) \in \mathcal{F}\}$ is an ideal $\neq \mathcal{P}(X)$ of $\mathcal{P}(X)$.
b. $\mathcal{F}$ is an ultrafilter $\Leftrightarrow \{A \subset X : (X \setminus A) \in \mathcal{F}\}$ is a maximal ideal of $\mathcal{P}(X)$.
c. For $x \in X$, $\mathcal{F}_x = \{A \subset X : x \in A\}$ is an ultrafilter on X.
d. For infinite X, there exist ultrafilters on X that are *not* of the form $\mathcal{F}_x$.
[The ultrafilters in (d) are called *free ultrafilters*[19] on X.]

*43. Let K_n be a field for $n \geq 0$ and $R = \prod_{n \geq 0} K_n$ the product ring. For $x = (x_n)_n \in R$, we write $V(x) = \{n \geq 0 : x_n = 0\}$. Show that the maximal ideals of R correspond bijectively to the ultrafilters on $\mathbf{Z}_{\geq 0}$ under the map $I \mapsto \{V(x) : x \in I\}$.

44. Let K be a field. Show that the algebraic sets in the affine space $\mathbf{A}^n(K)$ are the closed sets of a *topology* on K^n, the *Zariski topology* on K^n. [You may assume that $K[X_1, X_2, \ldots, X_n]$ is Noetherian.]

45. The *Zariski closure* of $V \subset \mathbf{C}^n$ is the closure of V in the Zariski topology.
    a. Determine the Zariski closure of $\mathbf{Z} \subset \mathbf{C}$.
    b. Determine the Zariski closure of $\{(x, y) : |x| \leq 1 \text{ and } |y| < 1\} \subset \mathbf{C}^2$.
    c. Show that the classical (metric) topology on $\mathbf{C}^n$ is (strictly) finer than the Zariski topology on $\mathbf{C}^n$.

46. A topological space X is called *irreducible* if it is not empty and not the union $X = Z_1 \cup Z_2$ of closed sets $Z_i \subsetneq X$.
    a. Prove that the intersection of two non-empty open subsets of an irreducible topological space is non-empty.
    b. Prove that the spectrum of an integral domain is irreducible.

47. A topological space $X \neq \emptyset$ is called *connected* if $\emptyset$ and X itself are the only subsets of X that are both open and closed. Prove that for a reduced ring $R \neq 0$, the following are equivalent:
    1. $\mathrm{Spec}(R)$ is not connected.
    2. There exist rings $R_1, R_2 \neq 0$ and an isomorphism $R \cong R_1 \times R_2$.
    3. R contains an idempotent element different from 0 or 1.

    * Is the condition that R is reduced necessary?

48. For $f \in R$, define the set $D(f) = \{P \in \mathrm{Spec}(R) : f \notin P\} \subset \mathrm{Spec}(R)$. Prove the following statements:
    a. $D(f)$ is open in $\mathrm{Spec}(R)$, and we have $D(f) \cap D(g) = D(fg)$.
    b. $D(f) = \emptyset \Leftrightarrow f$ is nilpotent.
    c. $D(f) = \mathrm{Spec}(R) \Leftrightarrow f \in R^*$.
    d. $U \subset \mathrm{Spec}(R)$ is open $\Leftrightarrow U$ is a union of sets of the form $D(f)$.

    [The sets $D(f)$ form a *basis* for the Zariski topology on $\mathrm{Spec}(R)$.]

49. Let $f : A \to B$ be a homomorphism of commutative rings and $f^* : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$ the induced map on the spectra. Prove the following statements:
    a. f^* is a continuous map.
    b. If f is surjective with kernel I, then $f^* : \mathrm{Spec}(B) \to \mathrm{im}\, f^* = Z(I)$ is a homeomorphism.
    c. For the natural map $f : A \to B = A/\mathrm{nil}(A)$, the map f^* is a homeomorphism.

50. Let $f : \mathbf{C}[X] \to \mathbf{C}[X, Y]$ be the natural inclusion map. Describe the induced map on the spectra and the fibers of this map. What is the corresponding "geometric picture"?

51. Answer the same questions for the map $f : \mathbf{C}[X] \to \mathbf{C}[X]$ given by $f(X) \mapsto f(X^2)$.

52. Let $f : \mathbf{Z} \to \mathbf{Z}[i]$ be the natural inclusion map. Describe the induced map on the spectra and the fibers of this map.

16 Modules

It often happens that elements of an abelian group can be "multiplied" by elements of a ring. Examples are the multiplication in §9 of elements of an arbitrary abelian group by integers in $\mathbf{Z}$, or the multiplications in linear algebra of the vectors in an n-dimensional K-vector space by the matrices in the ring $\mathrm{Mat}_n(K)$. As these multiplications respect the group operation, they are *endomorphisms* of the abelian group. In this situation, we are dealing with *modules* over a ring.

The formal definition is reminiscent of the definition of group actions that we gave in 5.1. Instead of the *group* of permutations on a set, we now use the *ring* of endomorphisms of the abelian group, defined in 11.8.

16.1. Definition. *Let R be a ring. An R-module is an abelian group M equipped with a ring homomorphism $\phi : R \to \mathrm{End}(M)$.*

It is customary to use additive notation for the group M, and to write rm for $\phi(r)(m)$. With these conventions, the requirement "$\phi(r) \in \mathrm{End}(M)$" and the conditions in 11.11 for ϕ to be a homomorphism can be spelled out as identities that need to be satisfied for all elements $r, s \in R$ and $m, n \in M$:

(M1) $r(m+n) = rm + rn$,
(M2) $(r+s)m = rm + sm$,
(M3) $(rs)m = r(sm)$,
(M4) $1 \cdot m = m$.

Definition 16.1 describes what is actually a *left module* M over R. Just as for group actions, where every left action of a group G on a set X gives rise to a corresponding right action on X (Exercise 5.20), every left module over R can be viewed as a right module over the *opposite ring* R^{opp} from Exercise 11.66, which may be identified with R itself when R is commutative.

Exercise 1. A *right module* over a ring R is an abelian group M equipped with a ring homomorphism $\phi : R^{\mathrm{opp}} \to \mathrm{End}(M)$. Write $\phi(r)(m)$ as mr and give the identities corresponding to (M1)–(M4).

It is rarely a source of confusion that the zero elements of R and M are both denoted by 0, say in the identity $0 \cdot m = 0$, valid for all $m \in M$, and that we write $M = 0$ for the *zero module*. Identities like $(-r)m = -(rm) = r(-m)$ follow easily from the identities (M1)–(M4) above, and show that we can unambiguously write $-rm$ for $r \in R$ and $m \in M$.

Exercise 2. Derive these identities from 16.1.

16.2. Examples. For every abelian group M, there is a unique ring homomorphism $\phi : \mathbf{Z} \to \mathrm{End}(M)$, so abelian groups and $\mathbf{Z}$-modules are the same objects. We already used the "$\mathbf{Z}$-module notation" km for $k \in \mathbf{Z}$ and $m \in M$ in the chapters on group theory.

If we take $R = \mathrm{End}(M)$ and ϕ the identity in 16.1, we see that every abelian group M is naturally a module over its endomorphism ring $\mathrm{End}(M)$. Note that $\mathrm{End}(M)$ is often a non-commutative ring.

If $f : R_1 \to R_2$ is a ring homomorphism, then every R_2-module becomes an R_1-module "via f." The R_1-module structure is obtained from the composition $R_1 \xrightarrow{f} R_2 \xrightarrow{\phi} \mathrm{End}(M)$. This is called *restriction of scalars*; it applies in particular to the inclusion map of a subring $R_1 \subset R_2$. For the unique map $R_1 = \mathbf{Z} \to R_2$, we see once more that every module is automatically a $\mathbf{Z}$-module, i.e., an abelian group.

Just like groups and rings, modules come with "corresponding" maps, the module homomorphisms.

16.3. Definition. *A homomorphism $M \to N$ of R-modules is a group homomorphism $f : M \to N$ that satisfies $f(rm) = rf(m)$ for all $r \in R$ and $m \in M$.*

A homomorphism of R-modules is also referred to as an *R-homomorphism* or an *R-linear map*. In the case where R is a field, these are the well-known linear maps from linear algebra. For $R = \mathbf{Z}$, we simply have group homomorphisms.

It is well known (Exercise 4.41) that the set $\mathrm{Hom}(M, N)$ of homomorphisms $M \to N$ of abelian groups itself is an abelian group, where the addition is defined by $(f_1 + f_2)(m) = f_1(m) + f_2(m)$. The subset $\mathrm{Hom}_R(M, N)$ of R-homomorphisms $M \to N$ is actually a *subgroup* of $\mathrm{Hom}(M, N)$.

Exercise 3. Show that for *commutative* R, the group $\mathrm{Hom}_R(M, N)$ becomes an R-module when we define $(rf)(m) = rf(m)$ for $r \in R$ and $f \in \mathrm{Hom}_R(M, N)$.

In the special case $M = N$, the set $\mathrm{Hom}_R(M, M) = \mathrm{End}_R(M)$ of R-endomorphisms $M \to M$ is a *subring* of the endomorphism ring $\mathrm{End}(M) = \mathrm{End}_{\mathbf{Z}}(M)$ defined in 11.8.

▶ Examples

The "Cayley map" $R \to \mathrm{End}(R^+) = \mathrm{End}_{\mathbf{Z}}(R)$ from 11.12 shows that every ring R is a module over itself under left multiplication. More generally, every left ideal $I \subset R$ is an R-module, so the notion of R-module may be viewed as a generalization of the notion of R-ideal.

If K is a field, then a K-module V is nothing but a *K-vector space*. The K-linear endomorphisms of V form a subring $\mathrm{End}_K(V) \subset \mathrm{End}(V)$, over which V is a module by restriction of scalars. If V has a finite basis $e_1, e_2, \ldots, e_n$ over K, then we can represent the elements of V as elements of the n-dimensional K-vector space K^n and identify $\mathrm{End}_K(V)$ with the matrix ring $\mathrm{Mat}_n(K)$. The module structure of V over $\mathrm{End}_K(V)$ is then the familiar multiplication of vectors in K^n by matrices in $\mathrm{Mat}_n(K)$. Note that $\mathrm{End}_K(V)$ is itself also a K-vector space.

Exercise 4. Is the product ring R^n also a module over the matrix ring $\mathrm{Mat}_n(R)$?

If R is a ring, then a module over the polynomial ring $R[X]$ is an R-module M endowed with an R-linear endomorphism $f \in \mathrm{End}_R(M)$. Indeed, the structure map $\phi : R[X] \to \mathrm{End}(M)$ is determined by the restriction $\phi|_R$ and the image $f = \phi(X) \in \mathrm{End}(M)$ of X by ϕ. The restriction $\phi|_R$ makes M into an R-module, and the commutativity relation $Xr = rX$ in $R[X]$ implies that $f = \phi(X)$ respects multiplication by $r \in R$: $f(rm) = rf(m)$ for $m \in M$. In particular, we see that for a field K, a module over

$K[X]$ is a K-vector space V endowed with a K-linear map $\phi(X) : V \to V$. This perspective is enlightening in linear algebra.

Exercise 5. Give a similar description of a module over the ring $K[X, X^{-1}]$ of Laurent polynomials over a field K.

For a ring R and a group G, a module over the group ring $R[G]$ is an R-module M endowed with an action of G on M as a group of R-linear automorphisms. Indeed, since every $g \in G$ is a *unit* in $R[G]$, the structure map $\phi : R[G] \to \mathrm{End}(M)$ is determined by the restriction $\phi|_R$, which makes M into an R-module, and the induced homomorphism $G \subset R[G]^* \to \mathrm{End}(M)^* = \mathrm{Aut}(M)$ of groups of units. Once again, the commutativity relation $gr = rg$ for $r \in R$ and $g \in G$ shows that $\phi(g)$ lies in the group $\mathrm{Aut}_R(M) = (\mathrm{End}_R(M))^*$ of R-linear automorphisms. If M is an arbitrary abelian group and G is a subgroup of $\mathrm{Aut}(M)$, then M has a natural $\mathbf{Z}[G]$-module structure. Such modules are common in Galois theory.

If G is an arbitrary group and K a field, then a $K[G]$-module is a K-vector space on which G acts as a group of K-linear automorphisms. In group theory, we saw that a group G can be successfully studied through its *actions* on various sets (both obvious and not). The same holds for the *linear* actions of G on well-chosen K-vector spaces, also known as the K-linear *representation*s of G.

If $f : R_1 \to R_2$ is a ring homomorphism, then R_2 is not only a module over itself under left multiplication, but also an R_1-module through restriction of scalars. The multiplication of $r_1 \in R_1$ by $r_2 \in R_2$ is defined by $r_1 r_2 = f(r_1)r_2$. The case where R_1 is commutative and $f[R_1]$ is contained in the center of R_2 is very common and has a special name.

16.4. Definition. *A (central) algebra over a commutative ring R is a ring A endowed with a ring homomorphism $R \to A$ that maps R into the center $Z(A)$.*

Since every ring A admits a unique homomorphism $\mathbf{Z} \to A$ with image in the center of A, every ring is a $\mathbf{Z}$-algebra.

The field $\mathbf{C}$ of complex numbers and the polynomial ring $\mathbf{R}[X]$ are commutative $\mathbf{R}$-algebras. Hamilton's *quaternion algebra* $\mathbf{H} = \mathbf{R} + \mathbf{R}i + \mathbf{R}j + \mathbf{R}k$ is a non-commutative $\mathbf{R}$-algebra. It is not an algebra over the subring $\mathbf{R} + \mathbf{R}i \cong \mathbf{C}$ of $\mathbf{H}$.

For a commutative ring R, the polynomial ring $R[X]$ and the group ring $R[G]$ are algebras over R. The ring $R[X]$ is commutative; for $R \neq 0$, the ring $R[G]$ is only commutative if G is abelian.

For every n-dimensional vector space V over a field K, the ring $\mathrm{End}_K(V)$ of K-linear endomorphisms of V is a K-algebra. For $n > 1$, it is not commutative.

▶ Standard Constructions

Many notions already familiar to us from the theory of (abelian) groups and from linear algebra can be easily generalized to the case of modules over rings other than $\mathbf{Z}$ or a field.

If M is an R-module and $M' \subset M$ is a subgroup that is mapped onto itself by multiplication with R, then M' is called a *submodule* of M. For R equal to $\mathbf{Z}$, a field K,

or the group ring $K[G]$ over K, we obtain the definitions of subgroup, subspace, and subrepresentation, respectively.

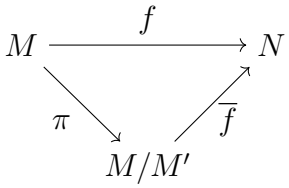
For an R-module homomorphism $f : M \to N$, the kernel $\ker f \subset M$ and the image $\operatorname{im} f = f[M] \subset N$ of f are submodules of M and N, respectively. Since f is, in particular, a group homomorphism, by 4.4, f is injective if and only if $\ker f$ is the zero module. If $f : M \to N$ is both injective and surjective, then f has a two-sided inverse $f^{-1} \in \operatorname{Hom}_R(N, M)$ and f is an R-module isomorphism. The isomorphism relation on R-modules, denoted by $M \cong_R N$, is an equivalence relation, as it is for groups. Note that two R-modules that are isomorphic as abelian groups need not be isomorphic as R-modules: the action of R on the two modules may differ.

Exercise 6. Give an example of two non-isomorphic modules over a ring R that are isomorphic as abelian groups.

If M is an R-module and $M' \subset M$ a submodule, then the quotient group M/M' (which always exists since M is abelian!) has a natural R-module structure given by

$$r(m + M') = rm + M' \in M/M'.$$

The homomorphism and isomorphism theorems from group theory can be immediately generalized to R-modules. If $f : M \to N$ is a homomorphism of R-modules and $M' \subset M$ is a submodule contained in $\ker f$, then the *homomorphism theorem* for modules says that f factors through the natural map $\pi : M \to M/M'$ to the quotient module M/M', as $f = \overline{f} \circ \pi$:



For $M' = \ker f$, there is the *isomorphism theorem* for modules: $\overline{f}$ gives an isomorphism

$$\overline{f} : M/\ker f \xrightarrow{\sim} f[M] \subset N.$$

The proofs of these can be copied directly from group theory; see 4.10 and 8.4. Where necessary, note that the subgroups and homomorphisms that appear there are now submodules and module homomorphisms.

Exercise 7. State and prove module analogs of the homomorphism theorems for groups 8.1 and 8.2.

Let S be a subset of an R-module M. The subset

$$RS = \{\textstyle\sum_{k=1}^{n} r_k s_k : r_k \in R, s_k \in S, n \in \mathbf{Z}_{\geq 0}\} \subset M$$

is the smallest submodule of M that contains S; it is called the submodule *generated by* S. If we have $RS = M$, then we say that M is generated by S, and S is called a set of generators of M. The module M is called *finitely generated* if there exists

a finite subset $S \subset M$ with $RS = M$. A module generated by a single element is called *cyclic*. For $\mathbf{Z}$-modules, we recover the well-known definitions for abelian groups. A module that is finitely generated or cyclic *as a module* is not necessarily so as an abelian group. For example, every ring is finitely generated as a module over itself, and is even cyclic with generator 1, but there are many rings whose underlying abelian group is not finitely generated.

Exercise 8. Show that $\mathbf{R}^n$ is a cyclic $\mathrm{Mat}_n(\mathbf{R})$-module.

A subset $S \subset M$ of an R-module is called *linearly independent* over R if for distinct $s_1, s_2, \ldots, s_n \in S$, we can only have $r_1 s_1 + r_2 s_2 + \ldots + r_n s_n = 0$ for $r_i \in R$ if $r_i = 0$ for all i. An R-module M generated by a linearly independent subset $S \subset M$ is called *free* with basis S. Every element $m \in M$ then has a *unique* representation as a *finite* sum $m = \sum_{s \in S} r_s s$.

16.5. Example. In the polynomial ring $R[X]$, the set $\{X^k\}_{k=0}^{\infty}$ of powers of X forms an R-basis. For $R[X_1, X_2, \ldots, X_n]$, we have a basis consisting of the "monic monomials" $X_1^{e_1} X_2^{e_2} \ldots X_n^{e_n}$ with $e_i \in \mathbf{Z}_{\geq 0}$. In the power series ring $R[[X]]$, which is also an R-module, the set $\{X^k\}_{k=0}^{\infty}$ is linearly independent but does not form a basis.

Given a set S, we construct the free R-module $R^{(S)}$ with basis S as the set of finite formal sums

$$R^{(S)} = \{\textstyle\sum_{s \in S} r_s s : r_s \in R \text{ and } r_s \neq 0 \text{ for only finitely many } s\}$$

with componentwise addition and R-multiplication. This means that, by definition, we have the equalities

$$\textstyle\sum_{s \in S} r_s s + \sum_{s \in S} r'_s s = \sum_{s \in S} (r_s + r'_s) s \qquad \text{and} \qquad r \sum_{s \in S} r_s s = \sum_{s \in S} (r r_s) s.$$

There is a natural inclusion $S \to R^{(S)}$ given by $s \mapsto 1 \cdot s$. For $R = \mathbf{Z}$, we recognize in the above the free abelian group with basis S; for $R = K$ a field, we obtain the K-vector space with basis S.

Given a family $\{M_i\}_{i \in I}$ of R-modules, their *direct product* $\prod_{i \in I} M_i$ is defined as the product group

$$\textstyle\prod_{i \in I} M_i = \{(m_i)_{i \in I} : m_i \in M_i\}$$

with componentwise R-multiplication: $r \cdot (m_i)_{i \in I} = (r m_i)_{i \in I}$. For every $i_0 \in I$, the projection $p_{i_0} : \prod_{i \in I} M_i \to M_{i_0}$ is a surjective R-homomorphism.

The *direct sum* $\bigoplus_{i \in I} M_i$ of the family $\{M_i\}_{i \in I}$ is the submodule of $\prod_{i \in I} M_i$ consisting of the elements $(m_i)_{i \in I}$ for which $m_i \neq 0$ for only finitely many $i \in I$. For every $i_0 \in I$, there is an injective R-homomorphism $\varepsilon_{i_0} : M_{i_0} \to \bigoplus_{i \in I} M_i$ that assumes the value zero on all coordinates other than i_0. We can view $\bigoplus_{i \in I} M_i$ as the submodule of $\prod_{i \in I} M_i$ generated by all images $\varepsilon_i[M_i]$. Note that for *finite* families, the direct sum and the direct product coincide. On the other hand, if we view the rings $R[X]$ and $R[[X]]$ in Example 16.5 as R-modules, then $R[X]$ is a direct *sum* of countably infinitely many modules R, while $R[[X]]$ is a direct *product* of countably infinitely many modules R.

Exercise 9. Show that for a commutative ring $R \neq 0$, $R[X]$ and $R[[X]]$ are not isomorphic as R-modules. (The general case is more difficult: see Exercises 57 and 57.)

As in the case of abelian groups, R-modules can often be written as sums or products of "simpler" modules. The notion of an exact sequence is particularly useful here as well for finding such a "decomposition" into smaller building blocks.

A sequence $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ of modules and module homomorphisms is called *exact* if it is exact as a sequence of abelian groups, that is, if we have $\mathrm{im}(f) = \ker(g)$. By the isomorphism theorem, a short exact sequence

$$0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$$

of R-modules induces an R-module isomorphism $\overline{g} : M_2/M_1 \xrightarrow{\sim} M_3$. The short exact sequence is called *split* if there exists an R-isomorphism $\phi : M_2 \xrightarrow{\sim} M_1 \oplus M_3$ such that $\phi \circ f$ is the embedding on the first coordinate and g can be obtained from ϕ by composing ϕ with the projection onto the second coordinate. This definition is a direct generalization of 9.2, and as in 9.3, we can show that the sequence splits exactly when f or g admits a *section*.

Exercise 10. Show that g admits a section if M_3 is a free R-module.

▶ Modules over Principal Ideal Domains

Almost everything we proved in §9 about the structure of $\mathbf{Z}$-modules can be easily generalized to the case of modules over a *principal ideal domain* R. We illustrate this by giving the analog of 9.11: the structure theorem for finitely generated modules over a principal ideal domain R.

First, we generalize several notions for $\mathbf{Z}$-modules to R-modules. For an R-module M, every element $m \in M$ has an *annihilator*

$$\mathrm{Ann}_R(m) = \{r \in R : rm = 0\}.$$

This is a left ideal of R, and $m \in M$ is called a *torsion element* if its annihilator $\mathrm{Ann}_R(m) \subset R$ is not the zero ideal. For an integral domain R, the torsion elements form the *torsion submodule* $T(M) \subset M$ of M. A module M over an integral domain R is called *torsion-free* if $T(M)$ is the zero module, and it is called a *torsion module* over R if we have $T(M) = M$.

Exercise 11. Show that if R is not an integral domain, then the sum of two torsion elements is not always torsion.

We now further assume that M is a module over a *principal ideal domain* R. The *order* of a torsion element $m \in M$ is then defined as a generator of the annihilator $\mathrm{Ann}_R(m)$ of m. Like any generator of an ideal, this is determined up to multiplication with units in R. The ideal

$$\mathrm{Ann}_R(M) = \bigcap_{m \in M} \mathrm{Ann}_R(m)$$

is the *annihilator* of M in R. If $\mathrm{Ann}_R(M) \neq 0$, then M is a torsion module and we call a generator of $\mathrm{Ann}_R(M)$ the *exponent* of M. By 12.17, in this case, $\mathrm{Ann}_R(M)$ is a product of prime ideals $(p) \neq 0$ in R.

Exercise 12. Verify that for $R = \mathbf{Z}$, we recover the known definitions of order and exponent for abelian groups.

For a prime ideal $(p) \subset R$ different from 0, we define the *p-power torsion module* $M(p) \subset M$ as the subgroup of M consisting of the elements whose order is a power of p. By the commutativity of R, this is a submodule of M.

16.6. Theorem. *Let R be a principal ideal domain and M a finitely generated module over R. Then M is a direct sum of finitely many cyclic R-modules. There exist an $r \in \mathbf{Z}_{\geq 0}$, the free rank of R, and an isomorphism*

$$M \cong_R T(M) \oplus R^r.$$

The torsion submodule $T(M)$ of M is the direct sum of finitely many p-power torsion modules $M(p)$, one for every prime ideal (p) that divides the exponent of $T(M)$. For every prime ideal (p), there is an R-isomorphism

$$M(p) \xrightarrow{\sim} R/(p^{k_1}) \oplus R/(p^{k_2}) \oplus \ldots \oplus R/(p^{k_m})$$

for uniquely determined integers $k_1 \geq k_2 \geq \ldots \geq k_m > 0$.

Proof. The proof is a direct generalization of the case $R = \mathbf{Z}$. First, as in 9.7, we prove that every submodule of R^n is free of rank $k \leq n$, and as in 9.9, we deduce from this that a finitely generated torsion-free R-module is a free R-module. For our finitely generated R-module M, $M/T(M)$ is finitely generated and torsion-free, hence isomorphic to R^r for some $r \geq 0$. The exact sequence $0 \to T(M) \to M \to M/T(M) \to 0$ splits (Exercise 10); this gives the first part of the theorem.

Since $T(M)$ is finitely generated and torsion, the exponent of $T(M)$ is non-zero, so by 12.17, it can be written as a finite product $E = \prod_{(p)} p^{e_p}$ for prime ideals (p). Take a prime element $p|E$, and write $E = p^{e_p}E'$. Since p^{e_p} and E' are relatively prime, we can write $xp^{e_p} + yE' = 1$ in R, and every element $m \in M$ can then be written as $m = xp^{e_p}m + yE'm$. The first term is annihilated by E', the second by p^{e_p}. This gives a decomposition $T(M) = M' \oplus M(p)$ of R-modules, and, by induction, we find an isomorphism $T(M) \cong_R \bigoplus_{p|E} M(p)$.

Finally, for the p-power torsion module $M(p)$, we essentially copy the proof of 9.11, where the induction is now carried out on the smallest number of generators of $M(p)$. If $m \in M(p)$ is an element in a minimal set of generators that is of maximal order $p^{k_1} = p^{e_p}$ in $M(p)$, then $M(p) = Rm \oplus M'$ holds again by constructing a suitable decomposition, and by induction this gives a representation of the desired form. For the uniqueness of the k_i and of the free rank r, we refer to Exercises 32 and 55. □

▶ Linear Algebra

To conclude this section, we show how the theory of modules can be applied in linear algebra. We take the *field* K as the base ring. As already noted, a K-module is nothing but a K-vector space, and a K-module homomorphism is a K-linear map. Unlike modules over arbitrary rings, modules over fields are always free. However, in

the infinite-dimensional case, the existence of bases for vector spaces is not obvious: just try to give a basis of $\mathbf{R}$ as a $\mathbf{Q}$-vector space!

16.7. Theorem. *Let K be a field and V a K-vector space. Then V is free over K. Given a set of generators $T \subset V$ of V and a linearly independent set $S \subset T$, there exists a K-basis B of V with $S \subset B \subset T$.*

Proof. The first statement follows from the second by taking $S = \emptyset$ and $T = V$.

For the second statement, we apply Zorn's lemma 15.11 to the collection $\mathcal{C}$ of linearly independent subsets of T containing S. The collection $\mathcal{C}$ is not empty since it contains S, and under the partial ordering given by inclusion, every chain in $\mathcal{C}$ has an upper bound that consists of the union of its elements. By Zorn's lemma, there is a maximal element $B \in \mathcal{C}$. We must show that V is generated by B. Since V is generated by T, it suffices to show that every element $t \in T$ is contained in the submodule $K \cdot B$ generated by B. For $t \in B$, this is clear, so take $t \in T \setminus B$. Then by the maximality of B, $B \cup \{t\}$ is not linearly independent, so there is a non-trivial relation $at + \sum_{x \in B} a_x x = 0$. Since B is linearly independent, we have $a \neq 0$, and since K is a field, we find that $t = -\sum_{x \in B} a^{-1} a_x x \in K \cdot B$, as desired. □

Theorem 16.8 shows that we can view a basis of V as a maximal linearly independent set in V or as a minimal set of generators of V. The cardinality of a K-basis of V is called the *dimension* $\dim(V) = \dim_K(V)$ of V over K. We will show that it does not depend on the choice of the basis.

16.8. Theorem. *Any two bases of a vector space V have the same cardinality.*

Proof. First, suppose that V has a basis B of finite cardinality n. We prove by induction on n that every other basis B' of V has *at most* n elements; by the symmetry in B and B', this suffices. For $n = 0$, we have $V = 0$, and there is nothing to prove. For $n \geq 1$, B' has an element $x \neq 0$, and applying 16.7 with $S = \{x\}$ and $T = \{x\} \cup B$ gives a subset $B_0 \subset B$ with $x \notin B_0$ for which $\{x\} \cup B_0$ is a basis of V. We have $B_0 \neq B$ by the maximality of B, so B_0 has at most $n - 1$ elements. Since B_0 and $B' \setminus \{x\}$ both give a basis of the quotient space V/Kx, the induction hypothesis says that $B' \setminus \{x\}$ has at most $n - 1$ elements, so B' has at most n elements, and we are done.

Now, suppose that V does not have a basis of finite cardinality. Given two bases B and B' of V, for every $x \in B$, we choose a finite subset $C_x \subset B'$ such that $x \in K \cdot C_x$. For example, we can take C_x to be the set of elements $x' \in B'$ that occur in the representation of x in the basis B'. We have $\bigcup_{x \in B} C_x = B'$ because $\bigcup_{x \in B} C_x$ generates V. Since every C_x is finite and B is infinite, the cardinality of $B' = \bigcup_{x \in B} C_x$ is not greater than that of B. By symmetry, B and B' now have the same cardinality. □

Since K-vector spaces are free, *every* short exact sequence $0 \to U \to V \to W \to 0$ of K-vector spaces splits (Exercise 10), and we have $V \cong U \oplus W$. In particular, for such a sequence, we have the dimension relation $\dim(V) = \dim(U) + \dim(W)$. If we apply this to the short exact sequence $0 \to \ker f \to V \xrightarrow{f} f[V] \to 0$ associated with

a K-linear map $f : V \to W$, then we obtain the well-known *dimension theorem* from linear algebra:

$$\dim \ker f + \dim f[V] = \dim(V).$$

Linear maps between finite-dimensional vector spaces have a *matrix representation* with respect to chosen bases of the vector spaces. If V and W are vector spaces with bases $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$, then we represent the linear map $A : V \to W$ given by $Ax_j = \sum_{i=1}^{m} a_{ij} y_i$ for $j = 1, 2, \ldots, n$ as an $m \times n$ matrix

$$A = (a_{ij})_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,n}} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

From this, we deduce that the set $\mathrm{Hom}_K(V, W)$ of K-linear maps $V \to W$, which itself has a natural K-vector space structure, has dimension mn. In terms of matrices, the composition of linear maps between vector spaces leads to the well-known *matrix multiplication* of $m \times n$ and $n \times r$ matrices ("row times column") that the reader may write out using multi-index summation signs.

▶ Normal Forms for Matrices

Many explicit calculations in linear algebra use the matrix representation of linear maps, but "linear map" and "matrix" are not synonyms: a linear map can only be described by a matrix after a *choice* of bases. For an n-dimensional K-vector space V, the choice of a basis leads to an identification $\mathrm{End}_K(V) \xrightarrow{\sim} \mathrm{Mat}_n(K)$ of the endomorphism ring of V with the ring of $n \times n$ matrices with entries in K. Given an endomorphism $A \in \mathrm{End}_K(V)$, we often try to choose a basis of V over which A assumes a simple form. For example, if V has a basis consisting of *eigenvectors* of A, then with respect to this, A is represented by a *diagonal matrix*. Such matrices are particularly efficient in calculations.

To write an endomorphism $A \in \mathrm{End}_K(V)$ in diagonal form, we need to write the vector space V as a direct sum of 1-dimensional subspaces, each of which is mapped onto itself by A. This is not always possible, but more generally, we can try to write V as a sum of subspaces that are mapped onto themselves by A. A *decomposition* $V = V_1 \oplus V_2$ as a sum of two A-invariant subspaces of dimension k and $n - k$ leads to a matrix representation of A as a "block matrix"

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

where A_i is a matrix that describes the action of A on the subspace V_i.

Decomposing V into A-invariant subspaces of smaller dimension means viewing V as a module over the polynomial ring $K[T]$ by letting T act as the transformation A, and writing V as a direct sum of $K[T]$-submodules Since $K[T]$ is a principal ideal domain, 16.6 applies immediately. The resulting normal form for matrices is called the

rational canonical form; see Exercise 42. For algebraically closed base fields K, such as $K = \mathbf{C}$, this form is particularly simple.

16.9. Jordan normal form of matrices. *Let A be an endomorphism of a finite-dimensional vector space V over an algebraically closed field K. Then there exists a decomposition $V = V_1 \oplus V_2 \oplus \ldots \oplus V_d$ of V as a sum of A-invariant subspaces, and for each subspace V_i, there is a basis over which A has matrix representation*

$$A|_{V_i} = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_i & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \lambda_i & 1 \\ 0 & 0 & \ldots & 0 & 0 & \lambda_i \end{pmatrix}$$

with entries $\lambda_i \in K$ on the diagonal, entries 1 on the superdiagonal, and entries 0 elsewhere.

Proof. Consider V as a module over $R = K[T]$ by letting T act as A. Then V is finitely generated over R, and even over K. Since R is infinite-dimensional as a vector space over K, every element $v \in V$ is an R-torsion element, and V is an R-torsion module. Since K is algebraically closed, every prime element in $R = K[T]$ is, up to a unit, equal to $T - \lambda$ for some $\lambda \in K$. Theorem 16.6 gives us a decomposition $V = \bigoplus_{i=1}^{d} V_i$ of V as a sum of cyclic R-modules $V_i = R/(T - \lambda_i)^{n_i}$. If we choose the powers $(T - \lambda_i)^k$ for $k = n_i - 1, n_i - 2, \ldots, 2, 1, 0$ as a basis of V_i, then multiplication by $T - \lambda_i$ acts on this basis by "shifting":

$$0 \leftarrow (T - \lambda_i)^{n_i - 1} \leftarrow (T - \lambda_i)^{n_i - 2} \leftarrow (T - \lambda_i)^{n_i - 3} \leftarrow \ldots \leftarrow T - \lambda_i \leftarrow 1.$$

This means that with respect to this basis, the matrix of $A - \lambda_i$ on V_i has entries 1 op the superdiagonal and entries 0 elsewhere. It immediately follows that A has the desired matrix representation on V_i. □

With respect to the basis of V consisting of the union of the bases specified in 16.9 for each V_i, the matrix A has the so-called Jordan normal form: a block matrix consisting of the "*Jordan blocks*" from the theorem along the diagonal. Since the Jordan blocks are nothing but a "translation" of the structure of V as a $K[T]$-module in the sense of 16.6, this form is uniquely determined up to the order of the blocks.

The values $\lambda_i \in K$ in 16.9 associated with the subspaces V_i are the *eigenvalues* of A, and in our representation, the first basis vector of V_i is an eigenvector with eigenvalue λ_i. Note that Jordan blocks can have *the same* eigenvalue λ_i for distinct components V_i and that the *number* of Jordan blocks with eigenvalue λ is equal to the dimension of the eigenspace $\ker[A - \lambda]$ associated with λ. For an eigenvalue λ of A, we call the subspace $V_\lambda = \bigoplus_{\lambda_i = \lambda} V_i$ of V the *generalized eigenspace* associated with λ. We have

$$V_\lambda = \ker[(A - \lambda)^n] \qquad \text{with} \quad n = \dim_K V.$$

In fact, it suffices to let the exponent n be the maximum of the dimensions n_i of the "Jordan subspaces" $V_i \subset V_\lambda$. The matrix A is *diagonalizable* over K if and only if all subspaces V_i are 1-dimensional. In this case, $V_\lambda = \ker[A - \lambda]$ is the "usual" eigenspace associated with λ. We can read off all known invariants of a linear map from the Jordan normal form of the corresponding matrix. For example, for the *characteristic polynomial* $P_A = \det(T \cdot I - A) \in K[T]$ of the map A in 16.9, we find the expression

$$P_A = \prod_{i=1}^{d} (T - \lambda_i)^{n_i} \in K[T].$$

Here, n_i is the dimension of the subspace $V_i \subset V$. By analogy with the case of finite abelian groups, we could call the polynomial P_A the "order" of the $K[T]$-module V. Every element of V is annihilated by this "order," just as, for finite abelian groups, all group elements are annihilated by the order of the group. The classical formulation of this result in linear algebra, which holds for arbitrary base fields (Exercise 45), is as follows.

16.10. Theorem (Cayley–Hamilton). *Let $P_A \in K[T]$ be the characteristic polynomial of $A \in \mathrm{End}_K(V)$. Then we have $P_A(A) = 0 \in \mathrm{End}_K(V)$.* □

As the example $A = \mathrm{id}_V$ for $\dim_K(V) > 1$ already shows, the characteristic polynomial P_A is not in general the smallest polynomial $f \in K[T]$ with $f(A) = 0$. We define the *minimal polynomial* $f_A \in K[T]$ of A as the monic polynomial of the smallest possible degree with this property. It is the analog of the *exponent* of a finite abelian group. See Exercise 46 for some properties of f_A.

EXERCISES.

13. Let K be a field. Show that a module over the polynomial ring $K[X_1, X_2, \ldots, X_n]$ is "the same" as a K-vector space V endowed with n commuting K-linear maps $V \to V$.

14. Let M be a module over a commutative ring R, and view $\mathrm{Hom}_R(R, M)$ as an R-module, as in Exercise 3. Prove that the map $\mathrm{Hom}_R(R, M) \to M$ given by $f \mapsto f(1)$ is an isomorphism of R-modules.

15. Let I be a left ideal of a ring R and M an R-module. Show that
$$IM = \left\{\textstyle\sum_{k=1}^{n} i_k m_k : i_k \in I, m_k \in M, n \in \mathbf{Z}_{\geq 0}\right\}$$
is a submodule of M and that when I is two-sided, the quotient module M/IM has a natural R/I-module structure.

16. An R-module M is called *simple* if it has exactly two submodules, 0 and M.
    a. Prove that a simple R-module is cyclic and isomorphic to R/I for a left ideal I of R that is maximal under the inclusion of left ideals.
    b. Prove that every R-homomorphism $f \neq 0$ between simple R-modules is an isomorphism and that the endomorphism ring $\mathrm{End}_R(M) = \mathrm{Hom}_R(M, M)$ of a simple R-module M is a division ring.
    c. What do items (a) and (b) amount to for $R = \mathbf{Z}$?

17. Let K be a field and V a finite-dimensional K-vector space. Determine when V is a simple R-module for $R = K$ and for $R = \mathrm{End}_K(V)$. In each case, determine the endomorphism ring $\mathrm{End}_R(V)$ for V simple.

18. Let M be an R-module and $f \in \mathrm{End}_R(M)$ a *projection*, i.e., an R-homomorphism $f : M \to M$ with $f \circ f = f$. Prove that there is an isomorphism $M \cong_R \ker f \oplus \mathrm{im}\, f$.

19. Let R be a ring and M an abelian group endowed with an R-multiplication $R \times M \to M$ that satisfies conditions (M1)–(M3) given after 16.1 ("a non-unitary module").
    a. Show that there exist a subgroup $M_1 \subset M$ that is an R-module and a group isomorphism $M \cong M_1 \oplus M_0$ such that R-multiplication on $M_1 \oplus M_0$ is given by $r(m_1, m_0) = (rm_1, 0)$.
    b. Show that an R-homomorphism $M = M_1 \oplus M_0 \to N = N_1 \oplus N_0$ of such objects (the definition is clear) is of the form $(m_1, m_0) \mapsto (f_1(m_1), f_0(m_0))$, with $f_1 \in \mathrm{Hom}_R(M_1, N_1)$ and $f_0 \in \mathrm{Hom}_{\mathbf{Z}}(M_0, N_0)$.

20. Show that for every set S, the free R-module with basis S is isomorphic to the direct sum $\bigoplus_{s \in S} R$.

21. Let R be a ring and S a set. Show that the set R^S of functions $f : S \to R$ has a natural R-module structure and that there is an R-module isomorphism $R^S \cong_R \prod_{s \in S} R$.

22. Let R be a commutative ring. Show that there exist R-isomorphisms
$$\mathrm{Hom}_R(M_1 \oplus M_2, N) \cong_R \mathrm{Hom}_R(M_1, N) \oplus \mathrm{Hom}_R(M_2, N),$$
$$\mathrm{Hom}_R(M, N_1 \oplus N_2) \cong_R \mathrm{Hom}_R(M, N_1) \oplus \mathrm{Hom}_R(M, N_2)$$
and, more generally,
$$\mathrm{Hom}(\textstyle\bigoplus_{i \in I} M_i, N) \xrightarrow{\sim} \prod_{i \in I} \mathrm{Hom}(M_i, N),$$
$$\mathrm{Hom}(M, \textstyle\prod_{i \in I} N_i) \xrightarrow{\sim} \prod_{i \in I} \mathrm{Hom}(M, N_i).$$

23. Suppose given an exact sequence $0 \to A \to B \to C \to 0$ of R-modules.
    a. Prove that B is finitely generated if A and C are.
    b. Show that the converse of (a) does not hold in general.
    c. Does the converse of (a) hold for a principal ideal domain R?

24. Let R be an integral domain. Show that for an exact sequence $0 \to A \to B \to C$ of modules over R, the induced sequence $0 \to T(A) \to T(B) \to T(C)$ of torsion submodules is exact. Is $T(B) \to T(C)$ necessarily surjective if $B \to C$ is?

25. Show that for short exact sequences $0 \to A \to B \xrightarrow{f} C \to 0$ and $0 \to C \xrightarrow{g} D \to E \to 0$ of R-modules, the induced sequence $0 \to A \to B \xrightarrow{gf} D \to E \to 0$ is exact. Conclude that every long exact sequence

$$0 \to A_1 \to A_2 \to A_3 \to \ldots \to A_{k-1} \to A_k \to 0$$

    of length k can be obtained from $k-2$ short exact sequences.

26. State and prove the analog of 9.12 for modules over a principal ideal domain.

27. For a module M over a commutative ring R, we define the *dual module* as $M^* = \mathrm{Hom}_R(M, R)$. Show that for an exact sequence $M_1 \to M_2 \to M_3 \to 0$ of R-modules, there naturally exists an induced sequence $0 \to M_3^* \to M_2^* \to M_1^*$ and that this is exact. Is $M_2^* \to M_1^*$ surjective if $M_1 \to M_2$ is injective?

28. Determine M^* for $M = \mathbf{Z}/m\mathbf{Z}$ as a module over $R = \mathbf{Z}$ and over $R = \mathbf{Z}/m\mathbf{Z}$. Also determine the dual of an arbitrary finitely generated module over $\mathbf{Z}$.

29. Show that the dual of a free module $R^{(S)}$ with basis S is isomorphic to the module R^S of R-valued functions on S.

30. Show that for every module M over a commutative ring R, there is a natural R-homomorphism $M \to M^{**}$ from M to its double dual $M^{**} = (M^*)^*$. Describe M^{**} for $R = \mathbf{Z}$ and M finitely generated.

31. An R-module M is called *reflexive* if the natural map $M \to M^{**}$ is an isomorphism. Show that finite-dimensional vector spaces are reflexive and that a $\mathbf{Q}$-vector space of infinite dimension is *not* reflexive. *Is the base field $\mathbf{Q}$ essential for this statement?

32. Show that for a commutative ring $R \neq 0$, the free R-modules $R^{(X)}$ and $R^{(Y)}$ are isomorphic if and only if X and Y have the same cardinality.
    [Hint: use that R has a maximal ideal to reduce to 16.8.]

33. Let R be a ring.
    a. Show that there exists an A-module $M \neq 0$ with $M \oplus M \cong_A M$.
    b. Take $R = \mathrm{End}_A(M)$, with M an A-module as in (a). Prove that for non-empty finite sets X and Y, the free modules $R^{(X)}$ and $R^{(Y)}$ are isomorphic.

34. A *hyperplane* in a vector space V is the kernel of a non-zero element $f \in V^*$. Show that every subspace $V' \subset V$ is the intersection of the hyperplanes that contain V'.

35. Let B be a basis of the vector space V and $y \in V$ a non-zero element. Show that there exists an element $x \in B$ such that $\{y\} \cup (B \setminus \{x\})$ is a basis of V. Deduce the finite-dimensional case of 16.8 from this *Steinitz exchange principle*.

36. Suppose given a descending sequence of sets $V \supset T_1 \supset T_2 \supset T_3 \supset \dots$ in V, and suppose that every set T_i generates the vector space V. Is $\bigcap_{i=1}^{\infty} T_i$ necessarily a generating set for V?

37. Let R be a principal ideal domain. Prove that every submodule of a free R-module is free.
[Hint: Let N be free with basis B and $M \subset N$. Now look at the subsets $C \subset B$ such that $M \cap R \cdot C$ is free and apply Zorn's lemma.]

38. Prove that for an exact sequence of finite-dimensional K-vector spaces

$$0 \to V_1 \to V_2 \to V_3 \to \dots \to V_{n-1} \to V_n \to 0,$$

we have the relation $\sum_{i=1}^{n} (-1)^i \dim_K(V_i) = 0$.

39. Determine the Jordan normal form of the complex matrix

$$A = \begin{pmatrix} 5 & 0 & -2 \\ 13 & 3 & 7 \\ 3 & 0 & 0 \end{pmatrix}.$$

40. Determine the Jordan normal form of the matrix

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

for $K = \mathbf{C}$ and for $K \supset \mathbf{F}_3$ algebraically closed.

41. Determine the Jordan normal form of the matrix in $\mathrm{Mat}_n(K)$ whose entries are all equal to 1. Does the characteristic of K matter?

The vector spaces V in all subsequent exercises are assumed to be *finite-dimensional*.

42. (*Rational canonical form*) Prove that given an $A \in \mathrm{End}_K(V)$, there exists a decomposition $V = \bigoplus_i V_i$ of V as a sum of A-invariant subspaces such that for a suitable basis of V_i, A acts as

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & a_1 \\ 1 & 0 & 0 & \dots & 0 & a_2 \\ 0 & 1 & 0 & \dots & 0 & \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & a_{n-1} \\ 0 & 0 & 0 & \dots & 1 & a_n \end{pmatrix}.$$

43. For a complex vector space V and an $A \in \mathrm{End}_{\mathbf{C}}(V)$, we define the *exponential map* $\exp(A) : V \to V$ by the identity $\exp(A) \cdot v = \sum_{k=0}^{\infty} \frac{1}{k!} A^k v$. Show that this leads to a well-defined endomorphism $\exp(A) \in \mathrm{End}_{\mathbf{C}}(V)$, and calculate $\exp(A)$ for the matrices in Exercises 39 and 40.

44. Let V be a complex vector space and $A \in \mathrm{End}_{\mathbf{C}}(V)$ an endomorphism such that $A^m = I$ for some $m \in \mathbf{Z}_{>0}$. Prove that A is diagonalizable. *Is this also true if we replace $\mathbf{C}$ with an arbitrary algebraically closed field?

45. Show that Theorem 16.10 holds for arbitrary base fields K.

46. For $A \in \mathrm{End}_K(V)$, recall the definition of the minimal polynomial $f_A \in K[T]$ given after 16.10. We view V as a $K[T]$-module by letting T act as A. Prove the following statements:
    1. f_A is a well-defined element of $K[T]$.
    2. V is a $K[T]$-torsion module with exponent f_A.
    3. f_A is a divisor of the characteristic polynomial P_A.
    4. P_A divides a power of f_A.

47. Suppose that the characteristic polynomial P_A and the minimal polynomial f_A of an $A \in \mathrm{End}_K(V)$, with K an algebraically closed field of characteristic different from 2, satisfy
$$P_A = f_A \cdot (T^2+1)^2,$$
$$f_A^3 = P_A \cdot (T-1)^4.$$
    Determine the Jordan normal form of A.

48. Prove that an endomorphism $A \in \mathrm{End}_K(V)$ is diagonalizable (over $\overline{K}$) if and only if its minimal polynomial has no multiple zeros in $\overline{K}$.

49. Let the vector space V be a $K[T]$-module through $A \in \mathrm{End}_K(V)$. Prove that the following statements are equivalent:
    1. $f_A = P_A$.
    2. V is a cyclic $K[T]$-module.
    3. There exists a basis $\{e_1, e_2, \ldots, e_n\}$ of V on which A acts by shifting: $e_1 \mapsto e_2 \mapsto e_3 \mapsto \ldots \mapsto e_{n-1} \mapsto e_n$.

50. Let $A, B \in \mathrm{End}_K(V)$ be diagonalizable endomorphisms, and suppose that A and B commute. Prove that there is a basis of V with respect to which A and B are both diagonal matrices.

51. Let $f \in K[T]$ be a polynomial and $A \in \mathrm{End}_K(V)$ an endomorphism with characteristic polynomial $P_A = \prod_{i=1}^n (T-\lambda_i) \in K[T]$. Prove that $B = f(A) \in \mathrm{End}_K(V)$ has characteristic polynomial $P_B = \prod_{i=1}^n (T - f(\lambda_i))$.

52. Let $A, B \in \mathrm{End}_K(V)$ be diagonalizable endomorphisms. Are $A+B$ and AB necessarily diagonalizable?

53. Give a proof or a counterexample for each of the following statements concerning complex matrices:
    1. If A^2 is diagonalizable, then A is diagonalizable.
    2. If $A^2 = A$, then A is diagonalizable.
    3. If A^5 is diagonalizable and A is invertible, then A is diagonalizable.

54. Let $A \in \mathrm{End}_{\mathbf{R}}(V)$ be a matrix that is diagonalizable over $\mathbf{C}$. Prove that V can be written as a sum $\bigoplus_i V_i$ of A-invariant subspaces of dimension $\dim V_i \leq 2$, where the 2-dimensional subspaces V_i have a basis over which A has matrix representation $\begin{pmatrix} \lambda_i \cos\phi & -\lambda_i \sin\phi \\ \lambda_i \sin\phi & \lambda_i \cos\phi \end{pmatrix}$ for some $\lambda_i \in \mathbf{R}^*$ and $\phi \in (0, \pi)$.

55. Show that for the p-power torsion module $M(p)$ in Theorem 16.6, the *number* of k_i greater than $e \in \mathbf{Z}_{\geq 0}$ is equal to the dimension of $p^e M(p)/p^{e+1}M(p)$ as a vector space

over $R/(p)$. Conclude that the k_i are uniquely determined by the isomorphism type of the R-module $M(p)$.

56. Determine all isomorphism types of R-modules with at most 16 elements for $R = \mathbf{Z}$ and for $R = \mathbf{Z}[i]$.

57. Determine all isomorphism types of $\mathbf{F}_2[X]$-modules M with $\dim_{\mathbf{F}_2}(M) \leq 3$. For each type, give the matrix by which X acts on M (with respect to a basis you choose).

*58. Let R be a ring. Prove that there is a right R-module $M \neq 0$ such that M is isomorphic to $\bigoplus_{n=1}^{\infty} M$ and that for every such M, the ring $A = \mathrm{End}_R M$ has the property that A and $\prod_{n=1}^{\infty} A$ are isomorphic as left A-modules.

*59. (G. M. Bergman). Let R be a ring.

a. Suppose that M_1, M_2, ... are R-modules, none of which are finitely generated. Prove that $M = \prod_{n=1}^{\infty} M_n$ is not countably generated, i.e., M has no countable subset that generates M as an R-module.
[Hint: given x_1, x_2, ... $\in M$, use a diagonal method to construct an element of M that does not lie in $Rx_1 + \ldots + Rx_i$ for any i.]

b. Prove that if the R-module $\prod_{n=1}^{\infty} R$ is countably generated, then it is even finitely generated.

c. Prove that for $R \neq 0$, the R-modules $\bigoplus_{n=1}^{\infty} R$ and $\prod_{n=1}^{\infty} R$ are not isomorphic.

60. Let R be a ring, M a cyclic R-module, and $N \subset M$ a sub-R-module.

a. Prove that M/N is a cyclic R-module.

b. Is N automatically cyclic? Give a proof or a counterexample.

c. As in (b), but now with the assumption that R is a principal ideal domain.

Literature

1. Many algebra books treat not only groups, but also rings and fields. This applies in particular to the books of Artin, Shafarevich, Lang, Gallian and Van der Waerden that we mentioned already in the Algebra I notes.

2. The representation of primes by quadratic forms such as $x^2 + y^2$ is the starting point of a very accessible book of Cox that can be used as an introduction to Algebraic Number Theory.

- D. A. Cox, *Primes of the form $x^2 + ny^2$*, Wiley, 1989. Second edition 2013.

3. Besides the book of Cox we just mentioned, there are several books providing more or less elementary introductions to Algebraic Number Theory. The book of Ireland en Rosen continues where our Chapter 12 stops. Kummer's theory of ideals is treated in Stewart en Tall.

- K. Ireland, M. Rosen, *A classical introduction to modern number theory*, Springer GTM 84, 1982. Second edition 1990.
- I. Stewart, D. O. Tall, *Algebraic number theory and Fermat's last theorem*, Chapman & Hall, 1979, fourth edition 2015.

4. The first of the two volumes of Kummer's collected papers contain his work in number theory. For those interested in the history of mathematical ideas, the long 1847 paper on ideal prime factors has an 'Anmerkung' (pp. 243–245) discussing analogies with concepts in chemistry. It is up to the modern reader to decide whether these are more than *Spiele des Witzes*.

- E. E. Kummer, *Collected papers*, edited by A. Weil, Springer, 1975.

5. There is an unexpected relation between $5 \bmod p$ en $p \bmod 5$: the first is a square in $(\mathbf{Z}/p\mathbf{Z})^*$ if and only if the second is a square in $(\mathbf{Z}/5\mathbf{Z})^*$. This is a special case of the quadratic reciprocity law, discovered by Euler and proved by Gauss at age 19. There are proofs by 'clever counting' such as the one found in the book by Hardy and Wright. Chapter 26 in the Algebra III course notes provides a more conceptual proof based on Gauss sums.

- G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1938.

  There are various corrected reprints.

6. Somewhat surprisingly, it is still unproved that $5 \bmod p$ is a primitive root modulo infinitely primes p. A famous conjecture of the German Emil Artin (1898–1962) from 1927 claims that this is the case, and provides an explicit density for the set of such p. Under assumption of another famous open conjecture, the Generalized Riemann Hypothesis on the location of the complex zeros of certain zeta-functions, it was proved after 40 years by the English mathematician Christopher Hooley.

- M. Ram Murty, *Artin's conjecture for primitive roots*, Math. Intelligencer 10, no. 4, 59–67 (1988).

7. André Weil's review of the collected works of Eisenstein is a fairy tale that is quite well known.

- A. Weil, *Review of 'Mathematische Werke' by Gotthold Eisenstein*, Bulletin of the AMS **VI**.82, pp. 658–663 (1976). It occurs as pp. 398–403 in part III of Weil's Œuvres scientifiques, Springer, 1979.

8. There is a wealth of literature dealing with closed formulas for π and ways to approximate it. Historical material related to π of a diverse nature has been collected in the following reference.

- L. Berggren, J. Borwein, P. Borwein, *Pi, a source book*, Springer, 1997, 3rd edition 2004.

9. For algorithms used in practice to factor polynomials over various fields, there is the book by Henri Cohen, founding father of the freely available GP-Pari computer algebra program.

Upper bounds on the coefficients of a polynomial $g = \sum_{j=0}^{d} b_j X^j$ of degree d dividing a polynomial $f = \sum_{i=0}^{n} a_i X^i$ of degree n in $\mathbf{Z}[X]$, such as

$$|b_j| \leq \binom{d-1}{j} \left(\sum_{i=0}^{n} a_i^2 \right)^{1/2} + \binom{d-1}{j-1} |a_n|,$$

have been given in 1974 by Mignotte.

The 1982 application of lattice reduction techniques to the factorisation of polynomials in $\mathbf{Q}[X]$ is due to the Dutch brothers Arjen and Hendrik Lenstra and the Hungarian Laszlo Lovász.

- H. Cohen, *A course in computational algebraic number theory*, Springer GTM, 1993.
- M. Mignotte, *An inequality about factors of polynomials*, Math. Comp. **28**, 1153–1157 (1974).
- A. K. Lenstra, H. W. Lenstra, Jr., L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261**, 515–534 (1982).

10. The remark that the rings $\mathbf{Z}/p^k\mathbf{Z}$ acquire better behavior for large k is explained by the 'limit case' $k = \infty$, which gives rise to the ring $\mathbf{Z}_p$ of p-adic numbers. As this is a *complete* ring under a natural metric, it bears more similarity to $\mathbf{R}$ than to $\mathbf{Z}$.

- F. Q. Gouvêa, *p-adic Numbers*, Springer Universitext, 1993.
- N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta-functions*, Springer GTM, 1977; second edition 1984.

11. The n-dimensional projective space $\mathbf{P}^n(K)$ over a field K is created by adding 'points at infinity' to the affine space K^n. Coxeter's book has a classical approach in the spirit of Euclid, with a lot of attention for the case of the projective plane – in which any two lines have non-empty intersection – and the history of the concept. The simple definition as $\mathbf{P}^n(K)$ as a set of lines through the origin can be found in many books on linear algebra and (algebraic) geometry. Chapter 7 in Stillwell's book prvides a nice historic introduction.

- H. S. M. Coxeter, *Projective geometry*, 1964. Second edition, Springer, 1987.
- J. Stillwell, *Mathematics and its history*, Springer UTM, 1989.

12. A proof of the basic theorems of algebraic geometry generalizing 15.4 requires more commutative algebra than is found in these notes. Fulton's book, which restricts itself to the case of curves, and the undergraduate text of Reid are more elementary. Harris's book, which contains many 'truly geometric examples', calls itself an introduction to the graduate text of

Hartshorne, which is technically quite demanding. Hartshorne's text became, right after it appeared in 1977, the 'standard text' for the modern algebraic geometer. It is no longer the only one.

- M. Reid, *Undergraduate algebraic geometry*, Cambridge University Press, 1988
- W. Fulton, *Algebraic curves*, Addison-Wesley, 1969.
- J. Harris, *Algebraic geometry*, Springer GTM 133, 1992.
- R. Hartshorne, *Algebraic geometry*, Springer GTM 52, 1977.

13. Arithmetic algebraic geometry studies number theoretic problems in a geometric language. Apart from the direct relation between solving equations and finding points – such as in the sentence preceding 12.21 – there is also an immediate alegbraic analogy between number rings such as $\mathbf{Z}$ of $\mathbf{Z}[i]$ and the coordinate rings of algebraic curves in affine space.

- D. Lorenzini, *An invitation to arithmetic geometry*, American Mathematical Society, 1996.

14. The *dimension theory* of commutative rings, which is part of commutative algebra, is found in the relevant chapters in the books mentioned below. On page 35 of Matsumura's book you find the identity $\dim(K[X_1, X_2, \ldots, X_n]) = n$.

- M. F. Atiyah, I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley, 1969.
- H. Matsumura, *Commutative ring theory*, Cambridge, 1986.

15. The concise Introduction of Atiyah and Macdonald mentioned above has two proofs of Hilbert's Nullstellensatzs, and another one is found in Reid's book. Eisenbud has an exposition on the Nullstellensatz in Section 1.6, and no less than five proofs.

- M. Reid, *Undergraduate commutative algebra*, Cambridge University Press, 1995.
- D. Eisenbud, *Commutative algebra with a view towards algebraic geometry*, Springer GTM 150, 1995.

16. The founding father of intuitionism is the Dutch mathematician L. E. J. Brouwer (1881–1966), who is also famous for work in topology such as *Brouwer's fixed point theorem*. His *Grundlagenstreit* with the famous German mathematician Hilbert attracted attention around 1930.

- A. Heyting, *Intuitionism, an introduction*, Amsterdam, 1965.

17. Zorn's Lemma is a convenient formulation of the *axiom of choice* in set theory, which says that a product of non-empty sets is non-empty. Lang's *Algebra* treats it in an appendix on set theory. The role in set theory of the axiom of choice in its various formulations is discussed in Devlin's book.

- K. Devlin, *The joy of sets*, 1979. Second edition, Springer UTM, 1993.

18. Reid's booklet **12** on algebraic geometry has a concluding section on 'history and sociology' mentioning the role of Grothendieck and his school. Hartshorne's scheme theoretic algebraic geomtry textbook is no longer the only standard textbook treating schemes.

- D. Eisenbud, J. Harris, *The geometry of schemes*, Springer GTM 197, 2000.

19. Free ultrafilters can be used to set up what is called *nonstandard analysis*, in which 'infinitely small' and 'infinitely large' objects obtain a formal meaning. This sometimes provides short proofs of certain statements in analysis that are classically proved using ε-δ-arguments.

- M. Davis, Applied nonstandard analysis, Wiley, 1977.

20. More information is found in Eisenbud's text listed under **15** and in Atiyah & Macdonald's Introduction listed under **14**.

21. Recommended texts:

- W. Fulton & J. Harris, *Representation Theory (A First Course)*, Springer GTM 129, 1991.
- J.-P. Serre, *Linear Representations of Finite Groups*, Springer GTM 42, 1977.

Index